# Package 'STRINGdb'

April 10, 2015

**Type** Package

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

**Version** 1.5.5

**Date** 2014-10-15

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

**Maintainer** Andrea Franceschini <andrea.franceschini@isb-sib.ch>, Alexander Roth <alexander.roth@gmail.com>, Christian Von Mering <mering@imls.uzh.ch>, Michael Kuhn <michael.kuhn@biotec.tu-dresden.de>, Lars J Jensen <lars.juhl.jensen@gmail.com>

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

**License** GPL-2

**Depends** R (>= 2.14.0), png, sqldf, plyr, igraph, RCurl, plotrix, methods, RColorBrewer, gplots, hash

**Suggests** RUnit, BiocGenerics

**biocViews** Network

## R topics documented:

STRINGdb-package *STRINGdb (an R interface to http://string-db.org)*

## 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 informationd 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/

## Details

| | |
|---|---|
| Package: | STRINGdb |
| Type: | Package |
| Version: | 1.0 |
| Date: | 2013-05-29 |
| License: | What license is it under? |
| Depends: | methods |

## 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

```
library(STRINGdb)
data(diff_exp_example1)

# create a new STRING_db object
```

```
string_db <- STRINGdb$new()

# map to STRING
example1_mapped = string_db$map( diff_exp_example1, "gene", removeUnmappedRows = TRUE )

# get the best 200 hits
hits = example1_mapped$STRING_id[1:200]

# plot the STRING network png
string_db$plot_network( hits )

# plot a protein-protein enrichment graph of the best 1000 hits in order to see how the ppi signal is distributed alo
string_db$plot_ppi_enrichment( example1_mapped$STRING_id[1:1000] )


######## use the "halo" mechanism in STRING to distinguish between down and up-regulated genes ########

# filter by p-value and add a color column (i.e. green down-regulated gened and red for up-regulated genes)
example1_mapped_pval05 = string_db$add_diff_exp_color( subset(example1_mapped, pvalue<0.05), logFcColStr="logFC"

# post payload information to the STRING server
payload_id = string_db$post_payload( example1_mapped_pval05$STRING_id, colors=example1_mapped_pval05$color )

# display a STRING network png with the "halo"
string_db$plot_network( hits, payload_id=payload_id )


######## compute enrichment in GO annotations  ########
enrichmentGO = string_db$get_enrichment( hits, category = "Process", methodMT = "fdr", iea = TRUE )
enrichmentKEGG = string_db$get_enrichment( hits, category = "KEGG", methodMT = "fdr", iea = TRUE )
head(enrichmentGO, n=7)
head(enrichmentKEGG, n=7)



########  Use STRING in order to get homologous proteins in other species ########

# get the reciprocal best hits of the following protein in all the STRING species
string_db$get_homologs_besthits(c("9606.ENSP00000365757"), symbets = TRUE)

# get the homologs of the following two proteins in the mouse (i.e. species_id=10090)
string_db$get_homologs(c("9606.ENSP00000365757", "9606.ENSP00000352336"), target_species_id=10090, bitscore_thr



########  Retrieve protein-protein interactions ########

# get the neighbors (in the ppi graph) of a given protein(s)
string_db$get_neighbors(c("9606.ENSP00000365757", "9606.ENSP00000352336"))

# get the interactions in between the input proteins
string_db$get_interactions(c("9606.ENSP00000365757", "9606.ENSP00000352336"))
```

```
######## Find clusters of interactions ########

# get clusters
clustersList = string_db$get_clusters(example1_mapped$STRING_id[1:600])

# plot first 4 clusters
par(mfrow=c(2,2))
for(i in seq(1:4)){
  string_db$plot_network(clustersList[[i]])
}
```

---

`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 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` *get_bioc_graph*

---

### 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 *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

```
## 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 = "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 *get_graph*

---

### 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

---

`get_link` *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

```
## 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

```
## 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 *get_png*

## 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, windowExtendedReferenceThresh
```

### 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 *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_pubmed *get_pubmed*

## 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

get_pubmed_interaction *get_pubmed_interaction*

## 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="9_05", 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

---

`get_summary` *get_summary*

---

### 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

---

`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

---

## load *load*

---

### 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 *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. |

## 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=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=0.0000000001, title="", quiet
```

### Arguments

| | |
|---|---|
| `string_ids` | a vector of STRING identifiers |
| `file` | 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

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, windowExtendedReferen
```

## Arguments

| | |
|---|---|
| proteins | a vector of protein identifiers |
| ppi_network | an igraph object containing the protein-protein interactions' graph. |
| file | 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 *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

ppie.compLambda *ppie.compLambda*

## Description

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

## Usage

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

## Arguments

| | |
|---|---|
| degrees | vector containing the degrees of the nodes of a set of proteins |
| edgeNum | 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

| | |
|---|---|
| degreesI | vector containing the degrees of the nodes present in nodeSet I |
| degreesJ | vector containing the degrees of the nodes present in nodeSet J |
| edgeNum | 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

| | |
|---|---|
| `degI` | degree of protein I |
| `degJ` | degree of protein J |
| `edgeNum` | 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). windowExtendedReferenceThrehold 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 length(hitList)/sliceWindow ).

enrichmentWindow
: vector containing the enrichments in protein-protein interactions of the input list of genes (the length of the vector is length(hitList)/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 length(hitList)/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)

---

# 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.

### Usage

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

### Arguments

`background_vector`
vector of STRING protein identifiers

### Author(s)

Andrea Franceschini

---

`STRINGdb` *STRINGdb (an R interface to http://string-db.org)*

---

### 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 informationd 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/

## Usage

```
STRINGdb(...)
```

## Arguments

| | |
|---|---|
| ... | description |

## Source

http://string-db.org

## 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'.

## Examples

```
library(STRINGdb)
data(diff_exp_example1)

# create a new STRING_db object
string_db <- STRINGdb$new()

# map to STRING
example1_mapped = string_db$map( diff_exp_example1, "gene", removeUnmappedRows = TRUE )

# get the best 200 hits
hits = example1_mapped$STRING_id[1:200]

# plot the STRING network png
string_db$plot_network( hits )

# plot a protein-protein enrichment graph of the best 1000 hits in order to see how the ppi signal is distributed alo
string_db$plot_ppi_enrichment( example1_mapped$STRING_id[1:1000] )


######## use the "halo" mechanism in STRING to distinguish between down and up-regulated genes ########

# filter by p-value and add a color column (i.e. green down-regulated gened and red for up-regulated genes)
example1_mapped_pval05 = string_db$add_diff_exp_color( subset(example1_mapped, pvalue<0.05), logFcColStr="logFC"

# post payload information to the STRING server
payload_id = string_db$post_payload( example1_mapped_pval05$STRING_id, colors=example1_mapped_pval05$color )

# display a STRING network png with the "halo"
string_db$plot_network( hits, payload_id=payload_id )


######## compute enrichment in GO annotations  ########
enrichmentGO = string_db$get_enrichment( hits, category = "Process", methodMT = "fdr", iea = TRUE )
```

```
enrichmentKEGG = string_db$get_enrichment( hits, category = "KEGG", methodMT = "fdr", iea = TRUE )
head(enrichmentGO, n=7)
head(enrichmentKEGG, n=7)



########  Use STRING in order to get homologous proteins in other species ########

# get the reciprocal best hits of the following protein in all the STRING species
string_db$get_homologs_besthits(c("9606.ENSP00000365757"), symbets = TRUE)

# get the homologs of the following two proteins in the mouse (i.e. species_id=10090)
string_db$get_homologs(c("9606.ENSP00000365757", "9606.ENSP00000352336"), target_species_id=10090, bitscore_thr



########  Retrieve protein-protein interactions ########

# get the neighbors (in the ppi graph) of a given protein(s)
string_db$get_neighbors(c("9606.ENSP00000365757", "9606.ENSP00000352336"))

# get the interactions in between the input proteins
string_db$get_interactions(c("9606.ENSP00000365757", "9606.ENSP00000352336"))



######## Find clusters of interactions ########

# get clusters
clustersList = string_db$get_clusters(example1_mapped$STRING_id[1:600])

# plot first 4 clusters
par(mfrow=c(2,2))
for(i in seq(1:4)){
  string_db$plot_network(clustersList[[i]])
}
```

---

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 informationd 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