

Package ‘funfea’

April 23, 2025

Type Package

Title FunFEA: an R Package for Fungal Functional Enrichment Analysis

Version 1.2.0

Author Julien Charest, Paul Loebenstein

Maintainer Julien Charest <julien.charest@tuwien.ac.at>

Description Performs enrichment analysis for functional annotations (KOG, GO, and EC) specifically tailored to fungal species.

This package allows users to identify overrepresented functional categories in their datasets and provides tools for visualization and interpretation.

It is designed for fungal genomics studies, facilitating insights into gene function, metabolic pathways, and regulatory mechanisms.

Compatible with standard annotation formats and customizable for diverse experimental data.

License GPL-2

Encoding UTF-8

LazyData true

Depends R (>= 2.10)

RoxygenNote 7.3.2

Imports ggplot2,
tidyr,
igraph,
dplyr

Contents

create_go_model	2
create_go_model_eggnog	3
create_kegg_model	3
create_kegg_model_eggnog	4
create_kegg_model_from_ec	4
create_kog_model	5
create_kog_model_eggnog	5
create_transcript2protein_id_df	6
fetch_models	6
gene2protein_id	7
generate_go_plot	8
generate_kegg_plot	9

generate_kog_plot	10
go_enrichment	10
kegg_enrichment	11
kog_enrichment	12
load_ec_annotation	13
load_eggnog_annotation	14
load_go_annotation	14
load_gtf_annotation	15
load_kegg_annotation	15
load_kog_annotation	16
transcript2protein_id	16
Index	17

create_go_model	<i>Create GO Model</i>
-----------------	------------------------

Description

This function generates a background frequency model for GO term enrichment analysis from a GO annotation.

Usage

```
create_go_model(go_annotation)
```

Arguments

`go_annotation` A dataframe with GO annotation information (see `load_go_annotation`)

Value

A list of dataframes with GO terms background frequencies by `gotermType` (`biological_processes`, `cellular_component`, `molecular_function`)

Examples

```
go_model_df <- create_go_model(go_annotation_df)
go_model_biological_processes_df <- go_model_df$biological_processes
go_model_cellular_component_df <- go_model_df$cellular_component
go_model_molecular_function_df <- go_model_df$molecular_function
```

`create_go_model_eggnog`*Create GO Model from eggNOG-mapper Annotation*

Description

This function generates a background frequency model for GO term enrichment analysis based on eggNOG-mapper annotation.

Usage

```
create_go_model_eggnog(eggnog_annotation)
```

Arguments

`eggnog_annotation`

A dataframe with eggNOG-mapper annotation information (see `load_eggnog_annotation`)

Value

A list of dataframes with GO terms background frequencies by `gotermType` (`biological_processes`, `cellular_component`, `molecular_function`)

Examples

```
go_model_df <- create_go_model_eggnog(eggnog_annotation_df)
go_model_biological_processes_df <- go_model_df$biological_processes
go_model_cellular_component_df <- go_model_df$cellular_component
go_model_molecular_function_df <- go_model_df$molecular_function
```

`create_kegg_model`*Create KEGG Pathway Model*

Description

This function generates a background frequency model for KEGG pathway enrichment analysis.

Usage

```
create_kegg_model(kegg_annotation)
```

Arguments

`kegg_annotation`

A dataframe with KEGG annotation information (see `load_kegg_annotation`)

Value

A list of dataframes with KEGG pathway background frequencies per KEGG category (`pathway_type`, `pathway_class`, `pathway_name`, `definition`, `complete`)

Examples

```
kegg_model_df <- create_kegg_model(kegg_annotation_df)
kegg_model_type_df <- kegg_model_df$pathway_type
kegg_model_class_df <- kegg_model_df$pathway_class
kegg_model_name_df <- kegg_model_df$pathway_name
kegg_model_definition_df <- kegg_model_df$definition
kegg_model_complete_df <- kegg_model_df$complete
```

```
create_kegg_model_eggnog
```

Create KEGG Pathway Model from eggNOG-mapper Annotation

Description

This function generates a background frequency model for KEGG pathway enrichment analysis based on eggNOG-mapper annotation.

Usage

```
create_kegg_model_eggnog(eggnog_annotation)
```

Arguments

eggnog_annotation

A dataframe with eggNOG-mapper annotation information (see load_eggnog_annotation)

Value

A list of dataframes with KEGG pathway background frequencies per KEGG category (pathway_type, pathway_class, pathway_name, definition, complete).

Examples

```
kegg_model_df <- create_kegg_model_eggnog(eggnog_annotation_df)
```

```
create_kegg_model_from_ec
```

Create KEGG Pathway Model from EC Annotation

Description

This function generates a background frequency model for KEGG pathway enrichment analysis from an EC (Enzyme Commission) Annotation.

Usage

```
create_kegg_model_from_ec(ec_annotation_df)
```

Arguments

ec_annotation A dataframe with EC annotation information (see load_ec_annotation)

Value

A list of dataframes with KEGG pathway background frequencies per KEGG category (pathway_type, pathway_class, pathway_name, definition, complete)

Examples

```
kegg_model_df <- create_kegg_model_from_ec(ec_annotation_df, filter = "All")
```

create_kog_model	Create COG/KOG Model
------------------	----------------------

Description

This function generates a background frequency model for COG/KOG categories enrichment analysis from a COG/KOG annotation.

Usage

```
create_kog_model(kog_annotation)
```

Arguments

`kog_annotation` A dataframe with COG/KOG annotation information (see `load_kog_annotation`)

Value

A dataframe with COG/KOG categories background frequencies

Examples

```
kog_model_df <- create_kog_model(kog_annotation_df)
```

create_kog_model_eggnog	Create COG/KOG Model from eggNOG-mapper Annotation
-------------------------	--

Description

This function generates a background frequency model for COG/KOG categories enrichment analysis based on eggNOG-mapper annotation.

Usage

```
create_kog_model_eggnog(eggnog_annotation)
```

Arguments

`eggnog_annotation`
A dataframe with eggNOG-mapper annotation information (see `load_eggnog_annotation`)

Value

A dataframe with COG/KOG categories background frequencies

Examples

```
kog_model_df <- create_kog_model_eggnog(eggnog_annotation_df)
```

```
create_transcript2protein_id_df
```

Create Transcript ID to Protein ID Conversion Dataframe

Description

This function parses a GTF/GFF/GFF3 annotation dataframe to generate a transcript ID to protein ID conversion dataframe.

Usage

```
create_transcript2protein_id_df(gtf_df)
```

Arguments

gtf_df A dataframe with GTF/GFF/GFF3 annotation information (see load_gtf_annotation)

Value

A dataframe with gene ID, gene name, transcript ID and protein ID information

Examples

```
transcript2protein_id_df <- create_transcript2protein_id_df(gtf_df)
```

```
fetch_models
```

Fetch Models

Description

This function loads precomputed KOG/GO/KEGG models for an organism and print infos on data origin

Usage

```
fetch_models(strain)
```

Arguments

strain Strain identifier used to query the database for available precomputed models (see available_models\$strain).

Value

A list with model information, gtf annotation and precomputed KOG/GO/KEGG models (info, gtf, transcript2protein_id, kog, go, kegg)

Examples

```
qm6a_models <- fetch_models("Trichoderma reesei (QM6a)")
qm6a_gtf <- qm6a_models$gtf
qm6a_transcript2protein_id <- qm6a_models$transcript2protein_id
qm6a_kog <- qm6a_models$kog
qm6a_go <- qm6a_models$go
qm6a_kegg <- qm6a_models$kegg
```

gene2protein_id	<i>Convert a Vector of Gene IDs to Protein IDs</i>
-----------------	--

Description

This function converts a vector of gene IDs and/or gene names to protein IDs, using a transcript2protein_id dataframe constructed from GTF/GFF/GFF3 annotation.

Usage

```
gene2protein_id(gene_ids, transcript2protein_id_df)
```

Arguments

gene_ids	A vector of gene IDs and/or gene names
transcript2protein_id_df	A dataframe with gene ID, gene name, transcript ID and protein ID information (see create_transcript2protein_id_df)

Value

A vector of protein IDs

Examples

```
protein_ids <- gene2protein_id(gene_ids, transcript2protein_id_df)
```

generate_go_plot	<i>Generate GO Terms Enrichment Plot</i>
------------------	--

Description

This function generates a bar plot for visualization of GO terms enrichment statistics.

Usage

```
generate_go_plot(  
  enrichment_df,  
  gotermType = "biological_process",  
  n = 10,  
  significant = "#E58601",  
  plot_title = NA,  
  plot_type = "bar"  
)
```

Arguments

enrichment_df	A list of dataframes with GO terms enrichment statistics per GO term type (see <code>go_enrichment</code>)
gotermType	A GO term type for plotting (default = "biological_process"; options = c("biological_process", "molecular_function", "cellular_component"))
n	A number of top-ranking GO terms to include in the plot (default = 10)
significant	A color assignment for statistically significant GO terms (default = "#E58601")
plot_title	A plot title (default = NA)
plot_type	A plot type option (default = "bar"; options = c("bar", "lollipop"))

Value

A ggplot2 bar plot object

Examples

```
generate_go_plot(go_enrichment_df)  
generate_go_plot(go_enrichment_df, gotermType = "biological_process", n = 15, plot_title = "GO Enrichment: Bio  
generate_go_plot(go_enrichment_df, gotermType = "molecular_function")  
generate_go_plot(go_enrichment_df, gotermType = "cellular_component")
```

generate_kegg_plot	<i>Generate KEGG Pathway Enrichment Plot</i>
--------------------	--

Description

This function generates a bar plot for visualization of KEGG pathway enrichment statistics.

Usage

```
generate_kegg_plot(  
  enrichment_df,  
  model = "pathway_class",  
  n = 10,  
  significant = "#9986A5",  
  plot_title = NA,  
  plot_type = "bar"  
)
```

Arguments

enrichment_df	A list of dataframes with KEGG pathway enrichment statistics per KEGG category (pathway_type, pathway_class, pathway_name, definition, complete) (see kegg_enrichment)
n	A number of top-ranking KEGG pathways to include in the plot (default = 10)
significant	A color assignment for statistically significant GO terms (default = "#9986A5")
plot_title	A plot title (default = NA)
plot_type	A plot type option (default = "bar"; options = c("bar", "lollipop"))
pathwayType	A KEGG category for plotting (default = "pathway_class"; options = c("definition", "pathway_name", "pathway_class", "pathway_type"))

Value

A ggplot2 plot object

Examples

```
generate_kegg_plot(kegg_enrichment_df)  
generate_kegg_plot(kegg_enrichment_df, model = "pathway_class", plot_title = "KEGG Pathway Enrichment: Class")  
generate_kegg_plot(kegg_enrichment_df, model = "pathway_name", significant = "blue")  
generate_kegg_plot(kegg_enrichment_df, model = "pathway_name", n = 5, plot_type = "lollipop")
```

generate_kog_plot	<i>Generate COG/KOG Enrichment Plot</i>
-------------------	---

Description

This function generates a bar plot for visualization of COG/KOG categories enrichment statistics.

Usage

```
generate_kog_plot(
  enrichment_df,
  significant = "#0B775E",
  plot_title = NA,
  plot_type = "bar"
)
```

Arguments

enrichment_df	A dataframe with COG/KOG enrichment statistics (see kog_enrichment)
significant	A color assignment for statistically significant COG/KOG categories (default = "#0B775E")
plot_title	A plot title (default = NA)
plot_type	A plot type option (default = "bar"; options = c("bar", "lollipop"))

Value

A ggplot2 bar plot object

Examples

```
generate_kog_plot(kog_enrichment_df)
generate_kog_plot(kog_enrichment_df, plot_type = "lollipop")
generate_kog_plot(kog_enrichment_df, plot_title = "KOG Enrichment", significant = "blue")
```

go_enrichment	<i>Calculate GO Enrichment</i>
---------------	--------------------------------

Description

This function generates a dataframe containing background/sample frequencies and enrichment statistics per GO term from a GO model and a vector of protein IDs. Protein IDs must match the Protein IDs from the annotation used to build the model.

Usage

```
go_enrichment(
  go_model,
  protein_ids,
  test = "fisher",
  p.adjust.method = "BH",
  reduce_redundancy_by_ancestor = FALSE,
  max_ancestor_depth = 5
)
```

Arguments

<code>go_model</code>	A list of dataframes with GO terms background frequencies per GO term type (see <code>create_go_model</code>)
<code>protein_ids</code>	A vector of protein IDs. Must match the Protein IDs from the annotation used to build the model.
<code>test</code>	A statistical method for testing the null of independence of rows and columns in a contingency table (default = "fisher"; options = c("fisher", "chisq"))
<code>p.adjust.method</code>	A statistical method to adjust p-values for multiple comparisons (default = "BH"; options = c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"))(see <code>p.adjust.methods</code>)
<code>reduce_redundancy_by_ancestor</code>	A method to remove enriched GO terms that are ancestors of other enriched terms to reduce redundancy and retain the most specific, informative terms. (default = FALSE; options = c(TRUE, FALSE))
<code>max_ancestor_depth</code>	Maximum distance (in terms of graph depth) allowed between a GO term and its descendant when applying redundancy reduction; higher values allow removal of more distant ancestor terms. (default = 5)

Value

A list of dataframes with GO terms enrichment statistics per GO term type

Examples

```
go_enrichment_df <- go_enrichment(go_model_df, protein_ids)
go_enrichment_df <- go_enrichment(go_model_df, protein_ids, test = "fisher", p.adjust.method = "BH")
go_enrichment_df <- go_enrichment(go_model_df, protein_ids, test = "chisq", p.adjust.method = "bonferroni")
go_enrichment_df <- go_enrichment(go_model_df, protein_ids, test = "fisher", p.adjust.method = "BH", reduce_re
```

kegg_enrichment

Calculate KEGG Pathway Enrichment

Description

This function generates a list of dataframes containing background/sample frequencies and enrichment statistics per KEGG category.

Usage

```
kegg_enrichment(
  kegg_model,
  protein_ids,
  test = "fisher",
  p.adjust.method = "BH"
)
```

Arguments

kegg_model	A list of dataframes with KEGG pathway background frequencies per KEGG category (pathway_type, pathway_class, pathway_name, definition, complete) (see create_kegg_model)
protein_ids	A vector of protein IDs. Must match the Protein IDs from the annotation used to build the model.
test	A statistical method for testing the null of independence of rows and columns in a contingency table (default = "fisher"; options = c("fisher", "chisq"))
p.adjust.method	A statistical method to adjust p-values for multiple comparisons (default = "BH"; options = c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"))(see p.adjust.methods))

Value

A list of dataframes with KEGG pathways enrichment statistics per KEGG category (pathway_type, pathway_class, pathway_name, definition, complete)

Examples

```
kegg_enrichment_df <- kegg_enrichment(kegg_model_df, protein_ids)
kegg_enrichment_df <- kegg_enrichment(kegg_model_df, protein_ids, test = "fisher", p.adjust.method = "BH")
kegg_enrichment_df <- kegg_enrichment(kegg_model_df, protein_ids, test = "chisq", p.adjust.method = "bonferroni")
kegg_enrichment_type_df <- kegg_enrichment_df$pathway_type
kegg_enrichment_class_df <- kegg_enrichment_df$pathway_class
kegg_enrichment_name_df <- kegg_enrichment_df$pathway_name
kegg_enrichment_definition_df <- kegg_enrichment_df$definition
kegg_enrichment_complete_df <- kegg_enrichment_df$complete
```

kog_enrichment

Calculate COG/KOG Enrichment

Description

This function generates a dataframe containing background/sample frequencies and enrichment statistics per COG/KOG category from a COG/KOG model and a vector of protein IDs. Protein IDs must match the Protein IDs from the annotation used to build the model.

Usage

```
kog_enrichment(kog_model, protein_ids, test = "fisher", p.adjust.method = "BH")
```

Arguments

kog_model	A dataframe with COG/KOG categories background frequencies (see create_kog_model)
protein_ids	A vector of protein IDs. Must match the Protein IDs from the annotation used to build the model.
test	A statistical method for testing the null of independence of rows and columns in a contingency table (default = "fisher"; options = c("fisher", "chisq"))
p.adjust.method	A statistical method to adjust p-values for multiple comparisons (default = "BH"; options = c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"))(see p.adjust.methods)

Value

A dataframe with COG/KOG categories enrichment statistics

Examples

```
kog_enrichment_df <- kog_enrichment(kog_model_df, protein_ids)
kog_enrichment_df <- kog_enrichment(kog_model_df, protein_ids, test = "fisher", p.adjust.method = "BH")
kog_enrichment_df <- kog_enrichment(kog_model_df, protein_ids, test = "chisq", p.adjust.method = "bonferroni")
```

load_ec_annotation	<i>Load EC Annotation</i>
--------------------	---------------------------

Description

This function loads a EC (Enzyme Commission) based KEGG annotation into a dataframe.

Usage

```
load_ec_annotation(path)
```

Arguments

path	Path to EC annotation
------	-----------------------

Value

A dataframe with EC annotation information

Examples

```
ec_annotation_df <- load_ec_annotation(path/to/ec/annotation)
```

`load_eggnog_annotation`*Load EggNOG-Mapper Annotation*

Description

This functions loads a eggNOG-mapper annotation into a dataframe.

Usage

```
load_eggnog_annotation(path)
```

Arguments

path Path to eggNOG-mapper annotation file

Value

A dataframe with eggNOG-mapper annotation information

Examples

```
eggnog_annotation_df <- load_eggnog_annotation(path/to/eggnog-mapper/annotation)
```

`load_go_annotation`*Load GO Annotation*

Description

This function loads a GO (Gene Ontology) annotation into a dataframe.

Usage

```
load_go_annotation(path)
```

Arguments

path Path to GO annotation

Value

A dataframe with GO annotation information

Examples

```
go_annotation_df <- load_go_annotation(path/to/go/annotation)
```

load_gtf_annotation	<i>Load GTF/GFF Annotation</i>
---------------------	--------------------------------

Description

This function loads a GTF/GFF/GFF3 annotation into a dataframe.

Usage

```
load_gtf_annotation(path)
```

Arguments

path	Path to GTF/GFF/GFF3 annotation
------	---------------------------------

Value

A dataframe with GTF/GFF/GFF3 annotation information

Examples

```
gtf_df <- load_gtf_annotation(path/to/gtf/annotation)
```

load_kegg_annotation	<i>Load KEGG Annotation</i>
----------------------	-----------------------------

Description

This function loads a KEGG annotation into a dataframe.

Usage

```
load_kegg_annotation(path)
```

Arguments

path	Path to KEGG annotation
------	-------------------------

Value

A dataframe with KEGG annotation information

Examples

```
kegg_annotation_df <- load_kegg_annotation(path/to/kegg/annotation)
```

load_kog_annotation	<i>Load COG/KOG Annotation</i>
---------------------	--------------------------------

Description

This function loads a COG/KOG (Clusters of Orthologous Genes) annotation file into a dataframe.

Usage

```
load_kog_annotation(path)
```

Arguments

path	Path to COG/KOG annotation
------	----------------------------

Value

A dataframe with COG/KOG annotation information

Examples

```
kog_annotation_df <- load_kog_annotation(path/to/kog/annotation)
```

transcript2protein_id	<i>Convert a Vector of Transcript IDs to Protein IDs</i>
-----------------------	--

Description

This function converts a vector of transcript IDs to protein IDs, using a transcript2protein_id dataframe constructed from GTF/GFF/GFF3 annotation.

Usage

```
transcript2protein_id(transcript_ids, transcript2protein_id_df)
```

Arguments

transcript_ids	A vector of transcript IDs
transcript2protein_id_df	A dataframe with gene ID, gene name, transcript ID and protein ID information (see create_transcript2protein_id_df)

Value

A vector of protein IDs

Examples

```
protein_ids <- transcript2protein_id(transcript_ids, transcript2protein_id_df)
```


Index

`create_go_model`, [2](#)
`create_go_model_eggnog`, [3](#)
`create_kegg_model`, [3](#)
`create_kegg_model_eggnog`, [4](#)
`create_kegg_model_from_ec`, [4](#)
`create_kog_model`, [5](#)
`create_kog_model_eggnog`, [5](#)
`create_transcript2protein_id_df`, [6](#)

`fetch_models`, [6](#)

`gene2protein_id`, [7](#)
`generate_go_plot`, [8](#)
`generate_kegg_plot`, [9](#)
`generate_kog_plot`, [10](#)
`go_enrichment`, [10](#)

`kegg_enrichment`, [11](#)
`kog_enrichment`, [12](#)

`load_ec_annotation`, [13](#)
`load_eggnog_annotation`, [14](#)
`load_go_annotation`, [14](#)
`load_gtf_annotation`, [15](#)
`load_kegg_annotation`, [15](#)
`load_kog_annotation`, [16](#)

`transcript2protein_id`, [16](#)