

# Package ‘keggseq’

April 27, 2017

**Title** KEGG analysis and visualizing of differentially expressed genes

**Version** 0.1

**Author** person(``Walid", ``Korani", email = ``korani@uga.edu",  
role = c(``aut", ``cre"))

**Maintainer** Walid Korani <korani@uga.edu>

**Description** The package do a run time KEGG analysis from the original database,  
different functions are included to find, grep and combine the interested  
datasets. It produces the plain text results in addition to graphs which is ready  
to be published.

**Depends** R (>= 3.2.2)

**License** LICENSE

**LazyData** true

**Imports** ggplot2, qvalue

**URL** <https://github.com/w-korani/keggseq>

## R topics documented:

|                        |   |
|------------------------|---|
| calculate_kegg         | 1 |
| corrected_datasets_ids | 2 |
| create_filtered_plot   | 3 |
| create_kegg_map        | 3 |
| create_plot            | 4 |
| grep_datasets          | 5 |
| kegg_map_enzymes       | 5 |
| kegg_summary           | 6 |
| list_species           | 7 |

---

|                |                       |
|----------------|-----------------------|
| calculate_kegg | <i>calculate_kegg</i> |
|----------------|-----------------------|

---

## Description

calculate\_kegg calculates hypergeometric test pvalues of kegg pathways and their adjusted\_pvalue  
using Bonferroni method.

It takes the input of grep\_datasets function or corrected\_datasets\_ids function as an input

**Usage**

```
calculate_kegg(dataset, deg_list)
```

**Arguments**

`dataset`            A output dataset of `grep_datasets` or `corrected_datasets_ids`.  
`deg_list`            A list containing the differentially expressed genes.

**Value**

A data frame containing 6 columns as following:  
column1: `map_numbers`, the map numbers of the KEGG pathway.  
column2: `background_gene_numbers`, number of all genes included in the pathway.  
column3: `deg_gene_numbers`, number of differentially expressed genes in the pathway.  
column4: `pvalue`, p-value of the hypergeometric test for the pathway.  
column5: `adjusted_pvalue`, the Bonferroni method adjustment of p-values.  
column6: `gene_list`, a list of the differentially expressed genes included in the pathway

**Examples**

```
a <- list_species("arachis")
b <- grep_datasets(c("adu", "aip"))
deg_list <- as.matrix(b[b$kegg_pathways=='00061',1])[,1]
c <- calculate_kegg(b, deg_list)
```

---

```
corrected_datasets_ids
      corrected_datasets_ids
```

---

**Description**

`corrected_datasets_ids` corrects `genes_id` to a different id set

**Usage**

```
corrected_datasets_ids(dataset_original, dataset_alternate_file)
```

**Arguments**

`dataset_original`            A dataset containing kegg `genes_id`, the output of `grep_datasets` function.  
`dataset_alternate_file`        A string for the name, and path, of a csv file containing the new `genes_id` in the same format of the dataset, the first one is the new gene ids and the second one is the map numbers.

**Value**

A dataset of the new `genes_id` in the same format of the input dataset

**Examples**

```
a <- grep_datasets(c("adu", "aip"))
path <- system.file(package = "keggseq")
file <- paste0(path, "/extdata/new_ids.csv")
b <- corrected_datasets_ids(a, file)
```

---

```
create_filtered_plot
      create_filtered_plot
```

---

**Description**

create\_filtered\_plot generates a scatter plot of the output of the calculate\_kegg function with passing a specific filter

**Usage**

```
create_filtered_plot(dataset, filter_name, filter_value)
```

**Arguments**

|              |  |
|--------------|--|
| dataset      | The output dataset of the calculate_kegg function  |
| filter_name  | a string of one of the following filters can be used:<br>background_gene_numbers, all genes included in the pathways.<br>deg_gene_numbers, the differentially expressed genes included int the pathway.<br>pvalue, the p-values.<br>adjusted_pvalue, the adjusted p-values.<br>qvalue, the q-values. |
| filter_value | A string starts by <,>==,<= or >= then the values for filtering  |

**Examples**

```
a <- list_species("arachis")
b <- grep_datasets(c("adu", "aip"))
deg_list <- as.matrix(b[b$kegg_pathways=='00061',1])[,1]
c <- calculate_kegg(b,deg_list)
create_filtered_plot(c,"deg_gene_numbers", ">0")
```

---

```
create_kegg_map      create_kegg_map
```

---

## Description

create\_kegg\_map generates kegg\_map with highlighted DEGs for an interested pathway. the function creates three objects that saved directly to the working directory:  
 "map\_number".png: it has the pathway map graph with highlighted differentially expressed genes, the differentially expressed gene are highlighted by blue color,  
 the differentially expressed genes that translated in a unique, specific, enzyme for the interested pathway are highlighted by red color.  
 "map\_number"\_enzymes.csv: a two column csv file contains the enzyme names and corresponding gene ids included in the interested pathway  
 "map\_number"\_de\_enzymes.csv: a two column csv file contains the enzyme names and corresponding differentially expressed gene ids included in the interested pathway

## Usage

```
create_kegg_map(species, dataset, deg_list, map_number, correction_file_name)
```

## Arguments

|                      |   |
|----------------------|---|
| species              | A list of abbreviations of the interested species.  |
| dataset              | The output dataset of the calculate_kegg function.  |
| deg_list             | A list containing the differentially expressed genes.   |
| map_number           | A string of the interested map number.  |
| correction_file_name | An optional filed, a string to file name, and path, of the dataset_altername_file in case of using corrected_datasets_ids function. |

## Examples

```
a <- list_species("arachis")
b <- grep_datasets(c("adu", "aip"))
deg_list <- as.matrix(b[b$kegg_pathways=='00061',1])[,1]
c <- calculate_kegg(b, deg_list)
create_kegg_map(c("adu", "aip"), c, deg_list, "00480")
```

---

create\_plot

*create\_plot*

---

## Description

create\_plot generates a scatter plot of the output of the calculate\_kegg function

## Usage

```
create_plot(dataset)
```

## Arguments

|         |  |
|---------|--|
| dataset | The output dataset of the calculate_kegg function. |
|---------|--|

**Examples**

```
a <- list_species("arachis")
b <- grep_datasets(c("adu", "aip"))
deg_list <- as.matrix(b[b$kegg_pathways=='00061',1])[,1]
c <- calculate_kegg(b,deg_list)
create_plot(c)
```

---

|               |                      |
|---------------|----------------------|
| grep_datasets | <i>grep_datasets</i> |
|---------------|----------------------|

---

**Description**

grep\_datasets function takes a list of species as an input and produces a datasets of map numbers of the pathways and gene IDs

**Usage**

```
grep_datasets(species)
```

**Arguments**

species            A list of abbreviations of the interested species.

**Value**

A data frame of two columns, the first one is KEGG gene ids and the second one is the map numbers

**Examples**

```
a <- list_species("arachis")
b <- grep_datasets(c("adu", "aip"))
```

---

|                  |                         |
|------------------|-------------------------|
| kegg_map_enzymes | <i>kegg_map_enzymes</i> |
|------------------|-------------------------|

---

**Description**

kegg\_map\_enzymes returns a dataset of enzymes included in an interested pathway.

**Usage**

```
kegg_map_enzymes(species, dataset, deg_list, map_number, correction_file_name, deg)
```

**Arguments**

|                                   |  |
|-----------------------------------|--|
| <code>species</code>              | A list of abbreviations of the interested species.   |
| <code>dataset</code>              | The output dataset of the <code>calculate_kegg</code> function   |
| <code>deg_list</code>             | A list containing the differentially expressed genes.  |
| <code>map_number</code>           | A string of the interested map number.   |
| <code>correction_file_name</code> | An optional file, a string to file name, and path, of the <code>dataset_alterate_file</code> in case of using <code>corrected_datasets_ids</code> function.        |
| <code>deg</code>                  | A logical value can be set to TRUE in case of returning only enzymes of differentially expressed genes, and FALSE to return all enzymes in the interested pathway. |

**Value**

A data set with three colnames, the first is the enzyme id, the second is the corresponding gene id and the third is the description of the enzyme.

**Examples**

```
a <- list_species("arachis")
b <- grep_datasets(c("adu", "aip"))
deg_list <- as.matrix(b[b$kegg_pathways=="00061",1])[,1]
c <- calculate_kegg(b,deg_list)
enz = kegg_map_enzymes(c("adu", "aip"), c, deg_list, "00061", , TRUE)
```

---

|                           |                     |
|---------------------------|---------------------|
| <code>kegg_summary</code> | <i>kegg_summary</i> |
|---------------------------|---------------------|

---

**Description**

`kegg_summary` produces a summary table of the output of `calculate_kegg` function with including the qvalues.

**Usage**

```
kegg_summary(dataset)
```

**Arguments**

|                      |   |
|----------------------|---|
| <code>dataset</code> | A output dataset of <code>calculate_kegg</code> . |
|----------------------|---|

**Value**

A data frame containing 7 columns as following:

- column1: `map_numbers`, the map numbers of the KEGG pathway.
- column2: `pathway_description`, the description of the pathway.
- column3: `background_gene_numbers`, number of all genes included in the pathway.
- column4: `deg_gene_numbers`, number of differentially expressed genes in the pathway.
- column5: `pvalue`, p-values of the hypergeometric test for the pathway.
- column6: `adjusted_pvalue`, the Bonferroni method adjustment of p-values.
- column7: `qvalue`, q-values of the p-values of column5.

**Examples**

```
a <- list_species("arachis")
b <- grep_datasets(c("adu", "aip"))
deg_list <- as.matrix(b[b$kegg_pathways=='00061',1])[,1]
c <- calculate_kegg(b, deg_list)
d <- kegg_summary(c)
```

---

|              |                     |
|--------------|---------------------|
| list_species | <i>list_species</i> |
|--------------|---------------------|

---

**Description**

The function returns the all currently available datasets in the KEGG database. In addition, a search term can be passed to the function to find a particular organism.

The searching term is not case sensitive.

**Usage**

```
list_species(key_word)
```

**Arguments**

key\_word      A not case sensitive string.

**Value**

A data frame of two columns, the first one is the dataset abbreviation and the second is the scientific name of the species.

**Examples**

```
a <- list_species("arachis")
a <- list_species()
```