

# Package ‘KOGMWU’

November 17, 2016

**Type** Package

**Title** Functional Summary and Meta-Analysis of Gene Expression Data

**Version** 1.1

**Date** 2016-11-13

**Author** Mikhail V. Matz

**Maintainer** Mikhail V. Matz <matz@utexas.edu>

## Description

Rank-based tests for enrichment of KOG (euKaryotic Orthologous Groups) classes with up- or down-regulated genes based on a continuous measure. The meta-analysis is based on correlation of KOG delta-ranks across datasets (delta-rank is the difference between mean rank of genes belonging to a KOG class and mean rank of all other genes). With binary measure (1 or 0 to indicate significant and non-significant genes), one-tailed Fisher's exact test for over-representation of each KOG class among significant genes will be performed.

**License** GPL-3

**Depends** R (>= 2.10), pheatmap

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2016-11-17 08:27:29

## R topics documented:

KOGMWU-package	2
adults.3dHeat.logFoldChange	3
corrPlot	4
gene2kog	4
kog.ft	5
kog.mwu	5
kog.mwut	7
larvae.longTerm	8
larvae.shortTerm	8
makeDeltaRanksTable	9
panel.cor	9
panel.cor.pval	10

---

KOGMWU-package

*Functional summary and meta-analysis of gene expression data*

---

## Description

Rank-based tests for enrichment of KOG (euKaryotic Orthologous Groups) classes with up- or down-regulated genes based on a continuous measure. The meta-analysis is based on correlation of KOG delta-ranks across datasets (delta-rank is the difference between mean rank of genes belonging to a KOG class and mean rank of all other genes). With binary measure (1 or 0 to indicate significant and non-significant genes), one-tailed Fisher's exact test for over-representation of each KOG class among significant genes will be performed.

## Details

Package: KOGMWU  
Type: Package  
Version: 1.1  
Date: 2016-11-13  
License: GPL-3

The most important function is `kog.mwu`, which performs a series of Mann-Whitney U tests when given two data tables: one, containing measures of interest for each gene (for example, log fold-change), and another, listing the association of each gene with a KOG class. The KOG class annotations for a collection of genes can be obtained using Weizhong Li's lab KOG BLAST server.

## Author(s)

Mikhail V. Matz

Maintainer: Mikhail V. Matz <matz@utexas.edu>

## References

Dixon GB, Davies SW, Aglyamova GA, Meyer E, Bay LK and Matz MV (2015) Genomic determinants of coral heat tolerance across latitudes. Weizhong Li's KOG BLAST server: <http://weizhonglab.ucsd.edu/metagenomic-analysis/server/kog/>

## Examples

```
## Not run:  
data(adults.3dHeat.logFoldChange)  
data(larvae.longTerm)  
data(larvae.shortTerm)  
data(gene2kog)  
  
# Analyzing adult coral response to 3-day heat stress:
```

```

alfc.lth=kog.mwu(adults.3dHeat.logFoldChange, gene2kog)
alfc.lth

# coral larvae response to 5-day heat stress:
l.lth=kog.mwu(larvae.longTerm, gene2kog)
l.lth

# coral larvae response to 4-hour heat stress
l.sth=kog.mwu(larvae.shortTerm, gene2kog)
l.sth

# compiling a table of delta-ranks to compare these results:
ktable=makeDeltaRanksTable(list("adults.long"=alfc.lth, "larvae.long"=l.lth, "larvae.short"=l.sth))

# Making a heatmap with hierarchical clustering trees:
pheatmap(as.matrix(ktable), clustering_distance_cols="correlation")

# exploring correlations between datasets
pairs(ktable, lower.panel = panel.smooth, upper.panel = panel.cor)
# p-values of these correlations in the upper panel:
pairs(ktable, lower.panel = panel.smooth, upper.panel = panel.cor.pval)

# plotting individual delta-rank correlations:
corrPlot(x="adults.long", y="larvae.long", ktable)
corrPlot(x="larvae.short", y="larvae.long", ktable)

## End(Not run)

```

---

adults.3dHeat.logFoldChange

*Heat stress response of adult coral*

---

## Description

Acropora millepora (adult) response to three days of heat stress (31.5oC) log-fold-changes inferred using DESeq package from tag-based RNA-seq data.

## Usage

```
data("adults.3dHeat.logFoldChange")
```

## Format

A data frame with 44363 observations on the following 2 variables.

gene gene id, a factor with 44363 levels

lfc log fold-change, a numeric vector

---

corrPlot	<i>Plots a pairwise correlation with linear regression line</i>
----------	---

---

**Description**

Plots Pearson's correlation between two columns in a dataframe, identified by column names. Also plots linear regression line and lists the correlation coefficient (r) and cor.test p-value.

**Usage**

```
corrPlot(x, y, data, ...)
```

**Arguments**

x	Name of the column to form X axis
y	Name of the column to form Y axis
data	The dataframe containing the two columns
...	Additional options for plot()

**References**

Dixon GB, Davies SW, Aglyamova GA, Meyer E, Bay LK and Matz MV (2015) Genomic determinants of coral heat tolerance across latitudes.

---

gene2kog	<i>KOG class annotations</i>
----------	------------------------------

---

**Description**

KOG class annotations for Acropora millerpoa transcriptome.

**Usage**

```
data("gene2kog")
```

**Format**

A data frame with 16175 observations on the following 2 variables.

V1 a factor with 16175 levels

V2 a factor with 23 levels

**Source**

[https://dl.dropboxusercontent.com/u/37523721/amillepora\\_transcriptome\\_july2014.zip](https://dl.dropboxusercontent.com/u/37523721/amillepora_transcriptome_july2014.zip)

**References**

Transcriptome assembly: Moya et al (2012), Mol Ecol 21:2440-2454. Transcriptome annotation: Dixon GB, Davies SW, Aglyamova GA, Meyer E, Bay LK and Matz MV (2015) Genomic determinants of coral heat tolerance across latitudes.

---

`kog.ft`*One-tailed Fisher's exact test for KOG enrichment.*

---

**Description**

Accessory function to kog.mwu()

**Usage**

```
kog.ft(gos)
```

**Arguments**

`gos` A dataframe with three columns, 'seq' (gene id), 'term' (KOG class) and 'value' (either 0 or 1, indicating significance).

**Value**

A dataframe with three columns: 'term', 'nseqs', 'pval' and 'padj'

**Author(s)**

Mikhail V. Matz

**References**

Dixon GB, Davies SW, Aglyamova GA, Meyer E, Bay LK and Matz MV (2015) Genomic determinants of coral heat tolerance across latitudes.

---

`kog.mwu`*Tests for KOG class enrichment.*

---

**Description**

Determines whether some KOG classes are significantly enriched with up- or down-regulated genes (Mann-Whitney U test for continuous measure), or whether some KOG classes are significantly over-represented among "significant" genes (one-tailed Fisher's exact test for binary measure, 0 or 1).

**Usage**

```
kog.mwu(data, gene2kog, Alternative = "t")
```

**Arguments**

data	Two-column dataframe: gene id, measure of significance.
gene2kog	Two-column dataframe of gene annotations: gene id, KOG class. The gene list can be longer or shorter than the first column in the 'data' item.
Alternative	Tailedness of the Mann-Whitney U test: two-tailed ("t"), greater ("g"), or less ("l")

**Details**

The measure can be continuous (such as log fold change), in which case Mann-Whitney U test will be performed, or binary (1 or 0: significant or not), in which case Fisher's exact test will be performed. The KOG class annotations for a collection of genes can be obtained using Weizhong Li's lab KOG BLAST server.

**Value**

For continuous measure, a dataframe with three columns: term : KOG class nseqs : Number of genes in this class delta.rank : Difference between the mean rank of genes belonging to this KOG class and all other genes pval : p-value of the Mann-Whitney U test padj : p-value adjusted using Benjamini-Hochberg 1995 "fdr" procedure

For binary measure, the output is similar but does not contain the delta.rank column.

**Author(s)**

Mikhail V. Matz <matz@utexas.edu>

**References**

Dixon GB, Davies SW, Aglyamova GA, Meyer E, Bay LK and Matz MV (2015) Genomic determinants of coral heat tolerance across latitudes. Weizhong Li's KOG BLAST server: <http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/kog/>

**Examples**

```
## Not run:
data(adults.3dHeat.logFoldChange)
data(larvae.longTerm)
data(larvae.shortTerm)
data(gene2kog)

# Analyzing adult coral response to 3-day heat stress:
alfc.lth=kog.mwu(adults.3dHeat.logFoldChange,gene2kog)
alfc.lth

# coral larvae response to 5-day heat stress:
```

```

l.lth=kog.mwu(larvae.longTerm, gene2kog)
l.lth

# coral larvae response to 4-hour heat stress
l.sth=kog.mwu(larvae.shortTerm, gene2kog)
l.sth

# compiling a table of delta-ranks to compare these results:
ktable=makeDeltaRanksTable(list("adults.long"=alfc.lth, "larvae.long"=l.lth, "larvae.short"=l.sth))

# Making a heatmap with hierarchical clustering trees:
pheatmap(as.matrix(ktable), clustering_distance_cols="correlation")

# exploring correlations between datasets
pairs(ktable, lower.panel = panel.smooth, upper.panel = panel.cor)
# p-values of these correlations in the upper panel:
pairs(ktable, lower.panel = panel.smooth, upper.panel = panel.cor.pval)

# plotting individual delta-rank correlations:
corrPlot(x="adults.long", y="larvae.long", ktable)
corrPlot(x="larvae.short", y="larvae.long", ktable)

## End(Not run)

```

---

kog.mwut

*Mann-Whitney U test for KOG enrichment.*


---

## Description

Accessory function to kog.mwu()

## Usage

```
kog.mwut(gos, Alternative = "t")
```

## Arguments

gos	A dataframe with three columns, 'seq' (gene id), 'term' (KOG class) and 'value' (continuous measure, such as log fold-change).
Alternative	Tailedness of the MWU test: two-tailed ("t"), greater-than ("g"), or less-than ("l")

## Value

A dataframe with three columns: 'term', 'nseqs', 'delta.rank', 'pval' and 'padj'

## Author(s)

Mikhail V. Matz

**References**

Dixon GB, Davies SW, Aglyamova GA, Meyer E, Bay LK and Matz MV (2015) Genomic determinants of coral heat tolerance across latitudes.

---

larvae.longTerm	<i>Long-term heat stress response of coral larvae</i>
-----------------	---

---

**Description**

Acropora millepora (larvae) response to five days of heat stress (31.5oC) log-fold-changes inferred using DESeq package from tag-based RNA-seq data from Meyer et al Mol Ecol 2011,17:3599-3616

**Usage**

```
data("larvae.longTerm")
```

**Format**

A data frame with 31844 observations on the following 2 variables.

gene gene id, a factor with 31844 levels

lfc log fold-change, a numeric vector

---

larvae.shortTerm	<i>Short-term heat stress response of coral larvae</i>
------------------	--

---

**Description**

Acropora millepora (larvae) response to four hours of heat stress (31.5oC) log-fold-changes inferred using DESeq package from tag-based RNA-seq data from Meyer et al Mol Ecol 2011,17:3599-3616.

**Usage**

```
data("larvae.shortTerm")
```

**Format**

A data frame with 32307 observations on the following 2 variables.

gene gene id, a factor with 32307 levels

lfc log fold-change, a numeric vector



---

makeDeltaRanksTable     *Make a combined delta-ranks table from several kog.mwu() results.*

---

**Description**

Extracts delta ranks from several kog.mwu() result tables and combines them into a single dataframe for heat map plotting and correlation analysis.

**Usage**

```
makeDeltaRanksTable(l1)
```

**Arguments**

l1                      A list of dataframes output by kog.mwu() function.

**Value**

A dataframe of delta-ranks (rows - KOG classes, columns - delta-ranks in different datasets).

**Author(s)**

Mikhail V. Matz

**References**

Dixon GB, Davies SW, Aglyamova GA, Meyer E, Bay LK and Matz MV (2015) Genomic determinants of coral heat tolerance across latitudes.

---

panel.cor                      *accessory function for pairs() to display Pearson correlations*

---

**Description**

works as upper.panel or lower.panel argument of pairs() (package graphics).

**Usage**

```
panel.cor(x, y, digits=2, cex.cor)
```

**Arguments**

x                      x element of the pairs() matrix  
y                      y element of the pairs() matrix  
digits                number of non-zero digits to leave at the end  
cex.cor                scaling factor for displayed text

**References**

cannibalized from an example in ?pairs (package graphics)

---

panel.cor.pval	<i>accessory function for pairs() to display pvalue of the Pearson correlation</i>
----------------	--

---

**Description**

works as upper.panel or lower.panel argument of pairs() (package graphics). Displays pvalues better than 0.1.

**Usage**

```
panel.cor.pval(x, y, digits = 2, cex.cor, p.cut=0.1)
```

**Arguments**

x	x element of the pairs() matrix
y	y element of the pairs() matrix
digits	number of non-zero digits to leave at the end
cex.cor	scaling factor for displayed text
p.cut	p-value cutoff

**References**

cannibalized from an example in ?pairs (package graphics)

# Index

## \*Topic **package**

KOGMWU-package, [2](#)

adults.3dHeat.logFoldChange, [3](#)

corrPlot, [4](#)

gene2kog, [4](#)

kog.ft, [5](#)

kog.mwu, [5](#)

kog.mwut, [7](#)

KOGMWU (KOGMWU-package), [2](#)

KOGMWU-package, [2](#)

larvae.longTerm, [8](#)

larvae.shortTerm, [8](#)

makeDeltaRanksTable, [9](#)

panel.cor, [9](#)

panel.cor.pval, [10](#)