KEGG pathway enrichment

Orlando Camargo

KEGG pathway enrichment in non-model organisms using *cluserProfiler*

This document contains steps required to perform enrichment analysis using the clusterprofiler package.

How is structured:

```
1) Install required packages
2) Load packages
3) Set the working and output directory
4) Create a table with IDs and ko:Knumber from input file 1
5) Perform KEGG pathway enrichment analysis
6) Plot results
```

For more details on using *clusterprofiler* see http://yulab-smu.top/biomedical-knowledge-mining-book/index.html.

Install and load packages Some packages can be installed from repository CRAN, others such *clusterprofiler* via Biocoductor. These lines are critical; if the enrichmente analysis shows an error, it is necessary to rerun it.

Install and load **clusterprofiler** package and install **R.utilis**:

```
library (clusterProfiler)
getOption("clusterProfiler.download.method")
```

```
## [1] "libcurl"
```

```
#install.packages("R.utils") #Uncomment to install and reinstall (in case of error).
R.utils::setOption("clusterProfiler.download.method","auto")
```

Also install and load the next packages

```
library("topGO")
library("dplyr")
library("R.utils")
library("ggplot2")
library("enrichplot")
library("RColorBrewer")
```

Set working directory and output file directory The working directory is where the input files are: Input file 1 (IDs-Knumber) and Input file 2 (IDs to be analyzed).

```
setwd("C:/Users/Orlando Camargo/Desktop/cp_enrichment/input_files/")
outpathcount = "C:/Users/Orlando Camargo/Desktop/cp_enrichment/results/"
```

Create a table with IDs and ko:Knumber from input file 1 It is important to generate a organized table where each Knumber has its corresponding ID.

```
KEGG_db<- readMappings("Tatroviride_protein_KEGG_ko.txt",sep="\t", IDsep = ",")

KEGGpath<-data.frame()
for(i in names(KEGG_db)){
  y =KEGG_db[[i]]
  ln<-length(KEGG_db[[i]])
  x=data.frame(ID=rep(i,ln),y)
  KEGGpath<-rbind(KEGGpath,x)
}
KEGGpath$y<-sub("ko:","",KEGGpath$y)
colnames(KEGGpath)<-c("ID","KEGG")

KEGGpath[60:80,]
```

```
##                 ID   KEGG
## 60 Tatro_000067-T1      -
## 61 Tatro_000069-T1      -
## 62 Tatro_000070-T1      -
## 63 Tatro_000071-T1      -
## 64 Tatro_000074-T1      -
## 65 Tatro_000075-T1 K02919
## 66 Tatro_000076-T1 K20794
## 67 Tatro_000077-T1 K00088
## 68 Tatro_000078-T1 K20178
## 69 Tatro_000079-T1 K03861
## 70 Tatro_000080-T1 K14779
## 71 Tatro_000081-T1 K04567
## 72 Tatro_000082-T1 K11340
## 73 Tatro_000082-T1 K11400
## 74 Tatro_000082-T1 K11652
## 75 Tatro_000083-T1      -
## 76 Tatro_000084-T1 K17732
## 77 Tatro_000085-T1 K03133
## 78 Tatro_000086-T1 K14152
## 79 Tatro_000087-T1 K15728
## 80 Tatro_000087-T2 K15728
```

Read input file 2, this can be the differential expression result table or a ID list in .txt.

```
data = read.table("input_file_2.txt", header = TRUE, sep="\t",
                  quote= "", row.names=1, comment.char="")
head(data)
```

```
##                     logFC   logCPM        F       PValue          FDR
## Tatro_002208-T1 -1.843287 5.400751 97.03844 1.756688e-09 5.520017e-06
## Tatro_011096-T1 -1.203280 5.524546 95.27822 2.069822e-09 5.520017e-06
## Tatro_008592-T1 -1.540310 6.419002 93.47258 2.455524e-09 5.520017e-06
## Tatro_011193-T1 -1.776928 5.592757 87.86123 4.251203e-09 7.645363e-06
## Tatro_007063-T1 -1.390135 6.720424 70.98038 2.679955e-08 1.668623e-05
## Tatro_003659-T1 -1.955451 4.865080 70.74980 2.754515e-08 1.668623e-05
```

```
id<-rownames(data)
id_ko<-KEGGpath[which(KEGGpath$ID%in%id),]
head(id_ko, n = 10)
```

```
##                  ID   KEGG
## 2   Tatro_000001-T1      -
## 3   Tatro_000002-T1      -
## 21  Tatro_000022-T1      -
## 85  Tatro_000092-T1 K00972
## 203 Tatro_000219-T1      -
## 214 Tatro_000233-T1 K07300
## 225 Tatro_000243-T1 K13621
## 268 Tatro_000287-T1 K16055
## 279 Tatro_000299-T1      -
## 287 Tatro_000311-T1      -
```

```
length(id_ko$ID)
```

```
## [1] 360
```

```
ko<-id_ko[id_ko$KEGG!="-",] # Delete those IDs that have non-Knumber associated.
head(ko)
```

```
##                   ID   KEGG
## 85  Tatro_000092-T1 K00972
## 214 Tatro_000233-T1 K07300
## 225 Tatro_000243-T1 K13621
## 268 Tatro_000287-T1 K16055
## 321 Tatro_000347-T1 K08197
## 328 Tatro_000355-T1 K03381
```

Perform the KEGG pathway enrichment analysis Use Knumber from data frame **ko**. In this step, you might get an error; if it happens, go back to **Install and load packages**.

```
knum<-ko$KEGG

enKEGG<-enrichKEGG(knum, organism = 'ko', minGSSize = 1,keyType = "kegg",
                   pvalueCutoff = 0.05,pAdjustMethod = "BH", qvalueCutoff = 1)
```

```
## Reading KEGG annotation online:
##
## Reading KEGG annotation online:
```

```
write.table(enKEGG, file=paste(outpathcount,"KEGG.txt", sep = ""),
            row.names=TRUE, col.names=NA, quote=FALSE, sep="\t")
#export only significant results
#use enKEGG@result to get all results (significant and not significant results)

enKEGG
```

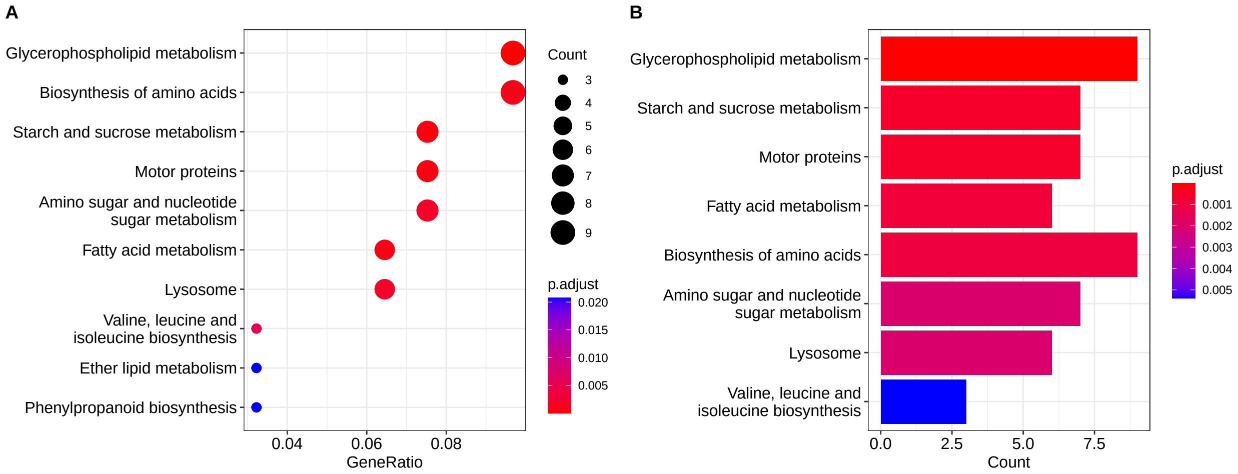
```
## #
## # over-representation test
## #
## #...@organism     ko
## #...@ontology     KEGG
## #...@keytype      kegg
## #...@gene     chr [1:141] "K00972" "K07300" "K13621" "K16055" "K08197" "K03381" "K00480" ...
## #...pvalues adjusted by 'BH' with cutoff <0.05
## #...16 enriched terms found
## 'data.frame':    16 obs. of  9 variables:
##  $ ID         : chr  "map00564" "map00500" "map04814" "map01212" ...
##  $ Description: chr  "Glycerophospholipid metabolism" "Starch and sucrose metabolism" "Motor proteins
##  $ GeneRatio  : chr  "9/93" "7/93" "7/93" "6/93" ...
##  $ BgRatio    : chr  "115/13850" "106/13850" "113/13850" "84/13850" ...
##  $ pvalue     : num  7.34e-08 6.95e-06 1.06e-05 2.08e-05 3.08e-05 ...
##  $ p.adjust   : num  0.000012 0.00057 0.00058 0.000852 0.001011 ...
##  $ qvalue     : num  1.01e-05 4.79e-04 4.88e-04 7.16e-04 8.50e-04 ...
##  $ geneID     : chr  "K13621/K01126/K13507/K18696/K16369/K00006/K00111/K01114/K01115" "K16055/K05349
##  $ Count      : int  9 7 7 6 9 7 6 3 3 3 ...
## #...Citation
##  T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo, and G Yu
##  clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
##  The Innovation. 2021, 2(3):100141
```

```
head(enKEGG@result)
```

```
##                 ID                             Description GeneRatio
## map00564 map00564             Glycerophospholipid metabolism      9/93
## map00500 map00500               Starch and sucrose metabolism      7/93
## map04814 map04814                              Motor proteins      7/93
## map01212 map01212                       Fatty acid metabolism      6/93
## map01230 map01230                 Biosynthesis of amino acids      9/93
## map00520 map00520 Amino sugar and nucleotide sugar metabolism      7/93
##             BgRatio       pvalue     p.adjust       qvalue
## map00564 115/13850 7.341551e-08 1.204014e-05 1.012361e-05
## map00500 106/13850 6.954338e-06 5.702557e-04 4.794833e-04
## map04814 113/13850 1.060942e-05 5.799814e-04 4.876609e-04
## map01212  84/13850 2.078043e-05 8.519977e-04 7.163780e-04
## map01230 238/13850 3.082017e-05 1.010902e-03 8.499878e-04
## map00520 156/13850 8.464802e-05 2.014599e-03 1.693918e-03
##                                                                   geneID Count
## map00564 K13621/K01126/K13507/K18696/K16369/K00006/K00111/K01114/K01115     9
## map00500               K16055/K05349/K01835/K01194/K01196/K00963/K01179     7
## map04814               K10426/K07374/K10413/K10396/K10401/K10392/K10357     7
## map01212                      K00626/K00667/K00668/K00507/K11262/K10256     6
## map01230 K01754/K13830/K01702/K01850/K00838/K00549/K00789/K00826/K03785     9
## map00520               K00972/K20844/K01183/K12373/K01835/K00963/K00820     7
```

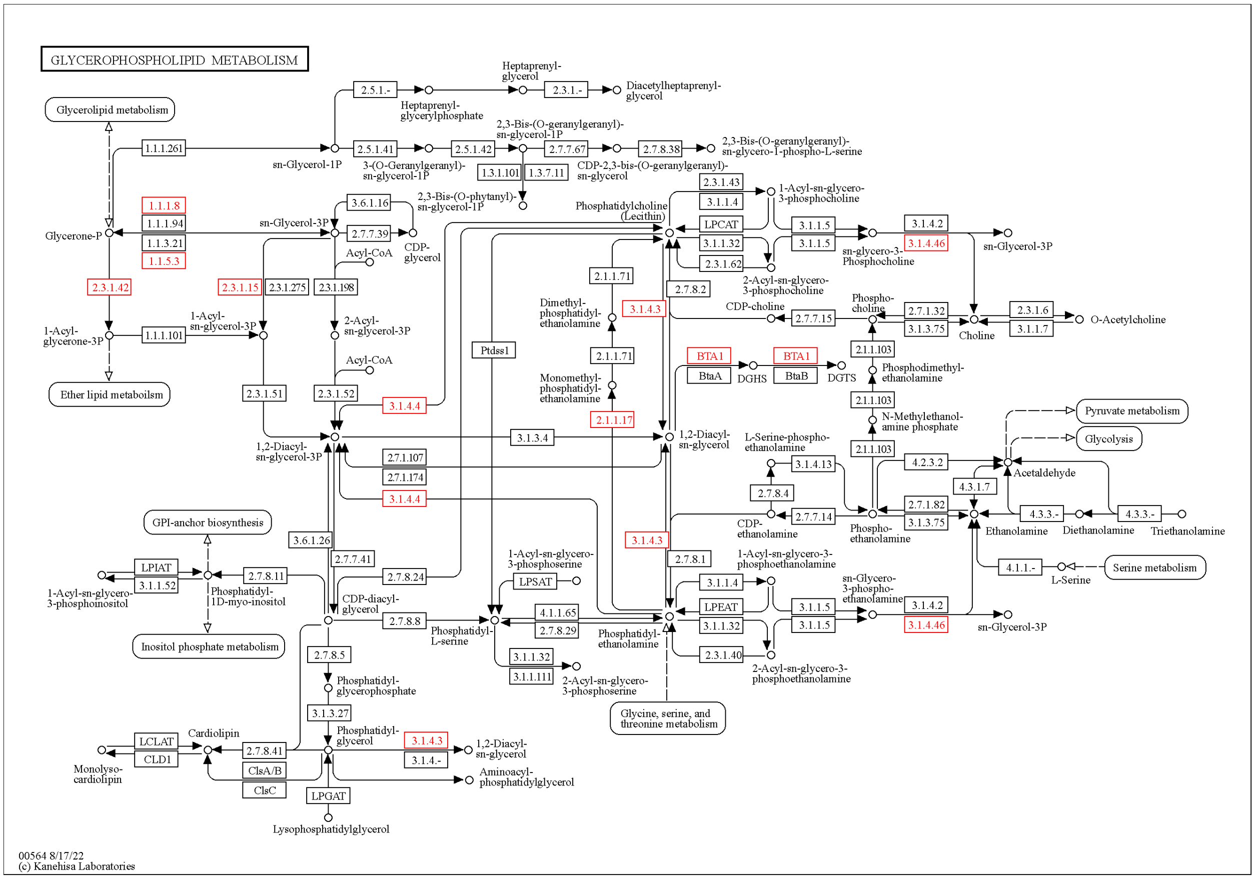
Plot results With *enrichplot*, it is possible to generate distinct plots to represent enrichment results, either using all the significant results or some. Also, with *clusterprofiler*, the results can be visualized in a KEGG pathway with a web browser.

```
p1<-dotplot(enKEGG)
p2<-barplot(enKEGG)
cowplot::plot_grid(p1, p2,ncol=2, labels=LETTERS[1:2], rel_widths=c(2, 2))
```



Visualize in a web browser:

```
browseKEGG(enKEGG, map00564)
```



map00564