斑马鱼基因组中在多物种中相对保守的基因的功能分析

张纳安

2019/03/21

多物种基因组的直系同源

Scientific_name	Abbreviation
Gallus_gallus	ggallus
Mus_musculus	mmusculus
Homo_sapiens	hsapiens
Xenopus_tropicalis	xtropicalis
Danio_rerio	drerio
Oryzias_latipes	olatipes
Ictalurus_punctatus	ipunctatus
Oreochromis_niloticus	oniloticus
Cynoglossus_semilaevis	csemilaevis

```
rm(list=ls())
#source("https://bioconductor.org/biocLite.R")
#biocLite("biomaRt")
setwd("/home/zhangna/Fish_genome/")
library("biomaRt")
library("dplyr", lib.loc="/usr/local/lib/R/site-library")
listMarts()
##get species
ensembl95=useMart(host='asia.ensembl.org',biomart='ENSEMBL_MART_ENSEMBL' )
table1=listDatasets(ensembl95)
sp <- read.table("/home/zhangna/Fish_genome/SP_4W5F/SP_4W5F.csv",sep=",",header=T, comment.char="",quote="\"",colClasses = "character")


## useMart
for (i in 1:length(sp$Scientific_name)){
  assign(sp$Scientific_name[i],useMart(host='asia.ensembl.org',
                                       biomart='ENSEMBL_MART_ENSEMBL',
                                       dataset=as.character(paste(sp$Abbreviation[i],"gene_ensembl",sep="_"))))
}

### get homolog for each sp
for (i in sp$Abbreviation){
  sapply(sp[sp$Abbreviation!=i,"Scientific_name"],function(a){
    filename=paste("~/Fish_genome/SP_4W5F/homolog/res190320",sub(" ","_",a),i,sep="_")
    homolog=getBM(attributes=c("ensembl_gene_id",
                               as.character(sapply(c("ensembl_gene",
                                                     "subtype",
                                                     "orthology_type",
                                                     "goc_score",
                                                     "orthology_confidence"),function(x){paste(i,"homolog",x,sep="_")}))),
                  mart = get(a))
    colnames(homolog)[1]=paste(a,"ensembl_gene_ID",sep="_")
    homolog=homolog[homolog[,2]!="",]
    write.csv(homolog,filename,
              quote = F,row.names = F)
  })
}
```

多物种基因组的直系同源

```
### get the gene info
sapply(sp[,"Scientific_name"],function(a){
  filename=paste("~/Fish_genome/SP_4W5F/gene_info/res190320",sub(" ","_",a),sep="_")
  homolog=getBM(attributes=c("ensembl_gene_id",
                             "external_gene_name",
                             "chromosome_name",
                             "start_position",
                             "end_position",
                             "family"),
                mart = get(a))
  colnames(homolog)[1]=paste(a,"ensembl_gene_ID",sep="_")
  write.csv(homolog,filename,
            quote = F,row.names = F)
})
```

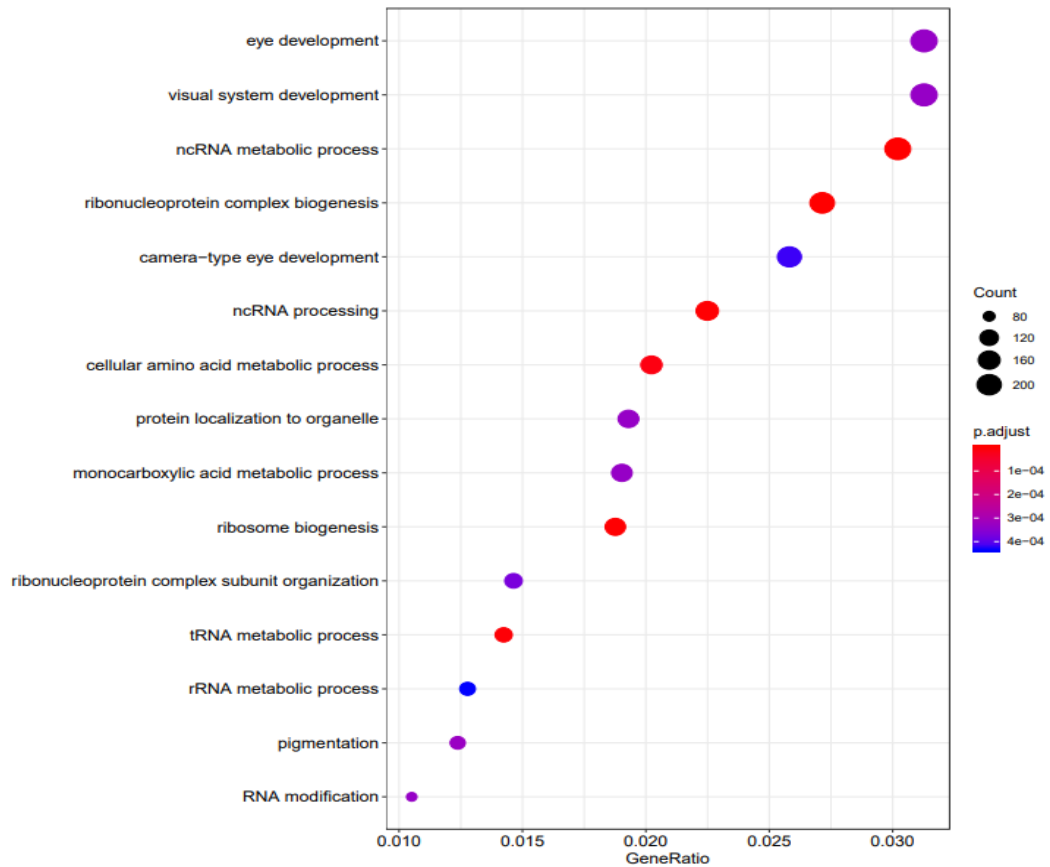
Danio_rerio_ensembl_gene_ID	ggallus_homolog_ensembl_gene	ggallus_homolog_subtype	ggallus_homolog_orthology_type	ggallus_homolog_goc_score	ggallus_homolog_orthology_confidence
ENSDARG00000063344	ENSGALG00000007378	Euteleostomi	ortholog_one2many	75	1
ENSDARG00000063344	ENSGALG00000005964	Euteleostomi	ortholog_one2many	0	1
ENSDARG00000069301	ENSGALG00000011593	Euteleostomi	ortholog_one2one	75	1
ENSDARG00000104901	ENSGALG00000010526	Euteleostomi	ortholog_one2one	0	1
ENSDARG00000031836	ENSGALG00000038541	Euteleostomi	ortholog_one2one	0	0
ENSDARG00000103391	ENSGALG00000004918	Euteleostomi	ortholog_one2one	25	1

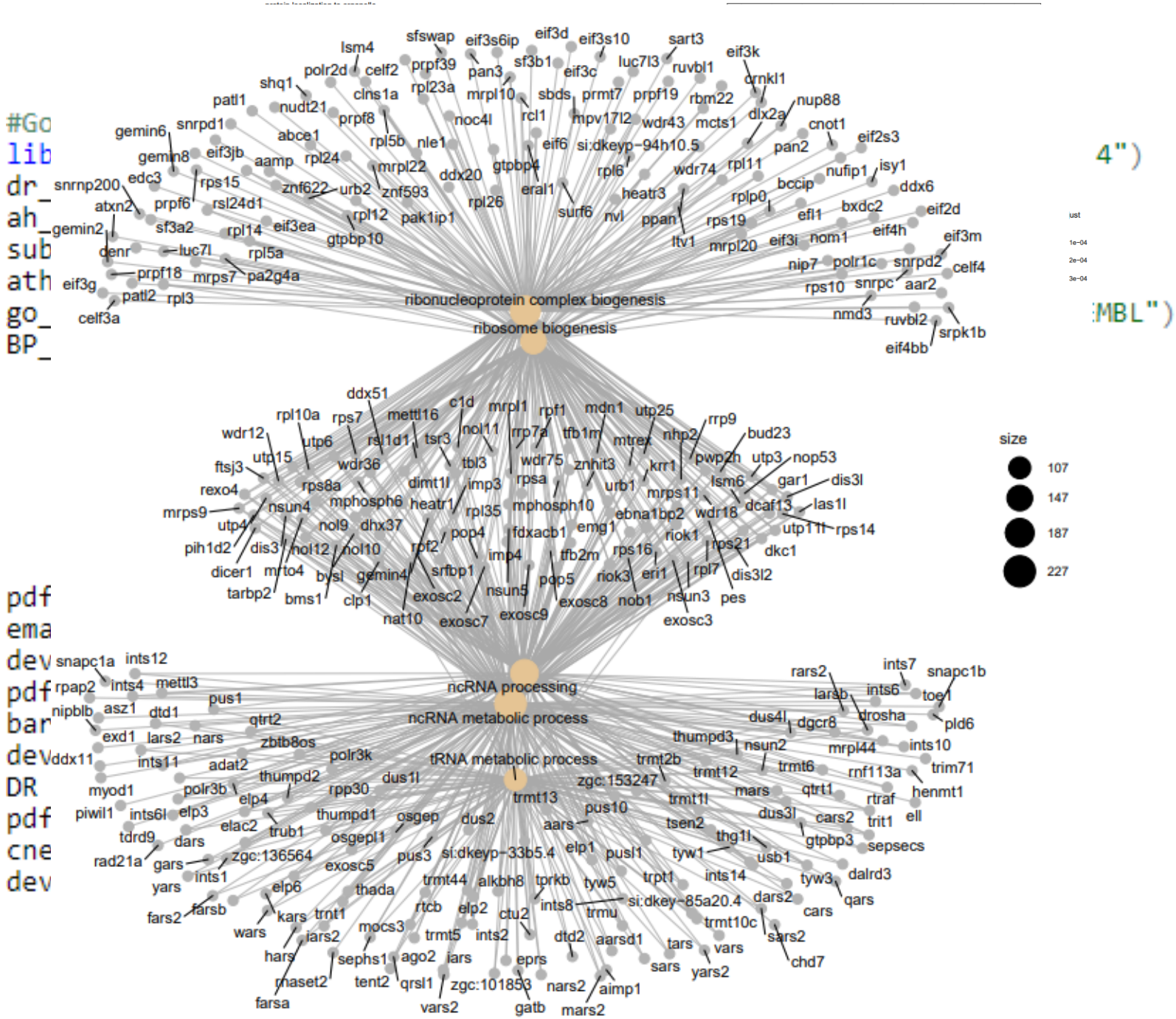
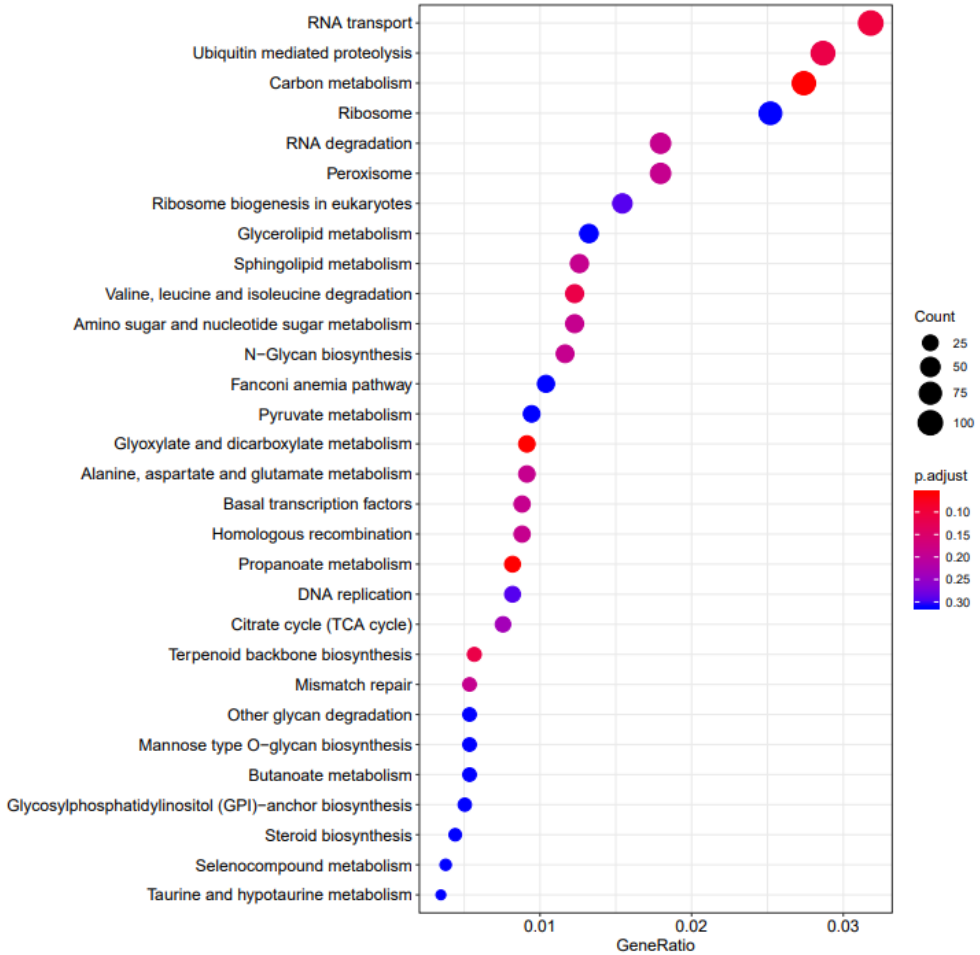
多物种基因组的直系同源

```
### get the Orthologous_gene in all sp
#### all sp
addlines <- function(tables){
  dataline = rep("0", nrow(Ontology))
  names(dataline) = Ontology[,1]
  dataline[names(tables)] = tables
  dataline = dataline[Ontology[,1]]
  return(dataline)
}

for (i in sp$Scientific_name){
  Ontology = read.table(paste("~/Fish_genome/SP_4W5F/gene_info/res190320",i,sep="_"),
                        sep=",",header=T, comment.char="",quote="\"",colClasses = "character")
  resultname=paste("~/Fish_genome/SP_4W5F/result/res190320",i,sep="_")
  result_sa = sapply(sp[sp$Scientific_name!=i,"Abbreviation"],function(a){
    filename=paste("~/Fish_genome/SP_4W5F/homolog/res190320",i,sub(" ","_",a),sep="_")
    OG <- read.table(filename,sep=",",header=T, comment.char="",quote="\"",colClasses = "character")
    OG=OG[OG[,2]!="",1]
    assign(sp[sp$Abbreviation==a,"Scientific_name"],table(OG))
  })
  for ( j in names(result_sa)){
    Ontology[,j] = addlines(result_sa[[j]])
  }
  Orthologous_DF=as.data.frame(Ontology)
  write.csv(Orthologous_DF, file = paste0(resultname,".csv"), row.names = FALSE,
            quote = FALSE)
}
```

定义斑马鱼基因组中在多物种中相对保守的基因集

```
library("clusterProfiler", lib.loc="/usr/local/lib/R/site-library")
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("org.Dr.eg.db", version = "3.8")
Danio = read.csv("~/Fish_genome/SP_4W5F/result/res190320_Danio_rerio.csv",sep=",",header=T, comment.char="",quote="\"",colClasses = "character")
degs = Danio[Danio[,11]!="0" & Danio[,12]!="0" & Danio[,13]!="0" & Danio[,14]!="0" & Danio[,10]!="0" & Danio[,9]!="0" & Danio[,8]!="0" & Danio[,7]!="0",1]

ego <- enrichGO(gene          = degs,
                keyType       = "ENSEMBL",
                OrgDb         = "org.Dr.eg.db",
                ont           = "ALL",
                pAdjustMethod = "BH",
                pvalueCutoff  = 0.1,
                qvalueCutoff  = 0.1)
dim(ego)
#[1] 508  10
dim(ego[ego$ONTOLOGY=='BP',])
#[1] 343  10
dim(ego[ego$ONTOLOGY=='CC',])
#[1] 87 10
dim(ego[ego$ONTOLOGY=='MF',])
#[1] 78 10

pdf("Danio_20190321_1.pdf", width=10, height=10)
dotplot(ego, showCategory=15)
dev.off()
```




ribonucleoprotein complex biogenesis
ribosome biogenesis
ncRNA processing
ncRNA metabolic process
tRNA metabolic process
size
107
147
187
227

```r
#KEGG analysis
dr_ENTREZID_list <- bitr(dr_list, fromType="ENSEMBL", toType=c("ENTREZID"), OrgDb="org.Dr.eg.db")
dr_ENTREZID = dr_ENTREZID_list[,2]
kegg_dr <- enrichKEGG(dr_ENTREZID,
                      organism = 'dre',
                      keyType = 'ncbi-geneid',
                      pvalueCutoff = 0.5,
                      pAdjustMethod = 'BH',
                      minGSSize = 10,
                      maxGSSize = 500,
                      qvalueCutoff = 0.5,
                      use_internal_data = FALSE)
pdf("Danio_20190321_6.pdf", width=10, height=10)
dotplot(kegg_dr, showCategory=30)
dev.off()
```