

Tarea Dominios

Francisco J. Falcon Chavez

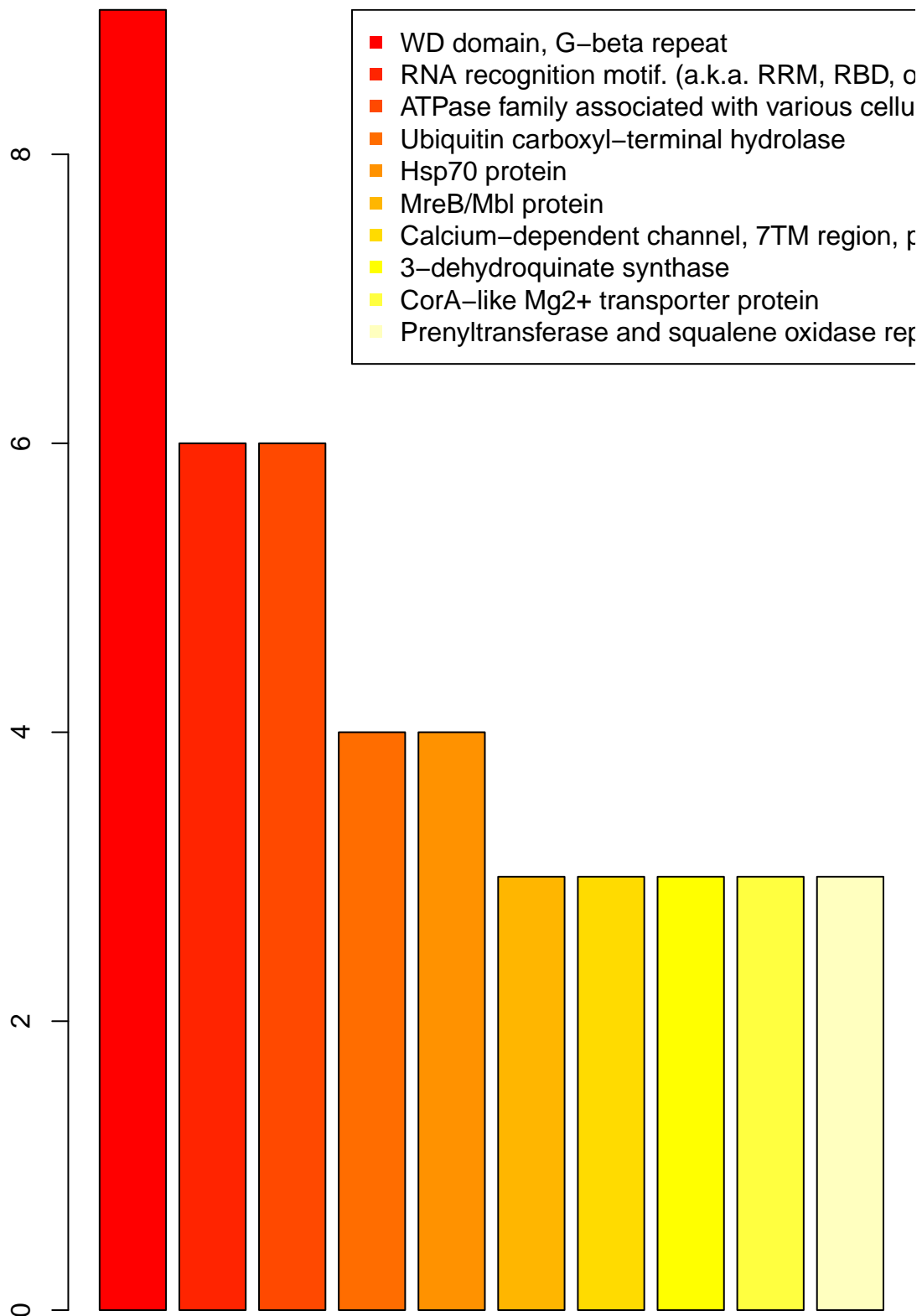
April 11, 2016

Primero obtuve dos listas: una con los identificadores de la base de datos de Pfam, y otra con las funciones de los dominios

```
cut -f8 *xls|sed '1d' | sed '/^.$/d'|sed 's/`/\n/g'| sed 's/\`.*//> Pfam.txt
cut -f8 *xls|sed '1d' | sed '/^.$/d'|sed 's/`/\n/g'| sed 's/\`/\t/g'| cut -f3 > Names.txt
```

You can also embed plots, for example:

```
a <- read.table('Pfam.txt',sep='\t')
b <- read.csv('Names.txt',sep='\t',header=F)
barplot(head(rev(table(a)[order(table(a))]),10),col=heat.colors(10),xaxt="n")
tableTotal <- cbind(a,b)
labs=c()
for (i in names(head(rev(table(a)[order(table(a))]),10))) {
  labs <- cbind(labs,as.character(tableTotal[which(tableTotal[,1]==i)[1],2]))
}
legend(4,9,labs,pch=15,col=heat.colors(10))
```



Al conocer los dominios, podemos saber cuales funciones son importantes para la celula. Por ejemplo, en la grafica una de las proteinas que tienen dominios sobrerrepresentados es la ATPasa que es esencial para la obtencion de energia, o las proteinas involucradas en la sintesis de prote+inas (HSP70) o de su degradacion (Ubiquitin carboxyl-terminal hydrolase).