Extraire les identifiants des gènes. Si votre liste de gènes est au format suivant:

```
sp|B8D9I2|AROQ_BUCA5
tr|B8D8E7|B8D8E7_BUCA5
tr|B8D8E8|B8D8E8_BUCA5
tr|B8D8N2|B8D8N2_BUCA5
tr|B8D8Y1|B8D8Y1_BUCA5
```

Vous pouvez extraire les identifiants Uniprot avec la commande awk, en changeant le séparateur de champ par défaut (des espaces) pour des "|". Ceci se fait avec l'option "Field Separator": FS=

```
awk '{ FS = "|" } ; { print $2 }' [fichier bbh, firstltine ou autre]
```

- Connect to David web site: http://david.abcc.ncifcrf.gov/
- Select "start analysis"
- Paste gene list, select "gene list" + select "submit list"

If you don't know the ID type:

- Select "Submit to conversion tool"

We see our IDs were Uniprot IDs.

- Convert all.

(note you can visualize full gene names, and download the gene list)

Submit converted list as Genlist. Give it a name that you can remember!
The list is now in the main David window.

- Check the “background” tag: it should be your species.

This means your gene list will be compared to the whole species genome for function enrichment.

Select “Functional Annotation Tool”
Check which db are already selected

- GO: Gene Ontology (Biological Process, Cellular Component, Molecular function)
- Pathway databases (KEGG)
- Domain databases (Interpro)

Click "chart" to observe term bias in one of the default databases
Select "Functional annotation clustering" to group results from all databases at once.
(pretty redundant but more exhaustive).

Click on bar to get gene names under term.