PROGRAMME GOLIAT

Gene Ontology LInk Analysis Tool

Aurélien Birer - François Bonnardel - Samaneh Kamyabiazar - Yoann Pageaud
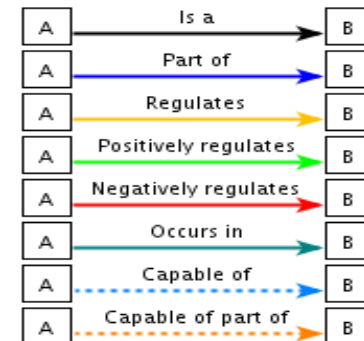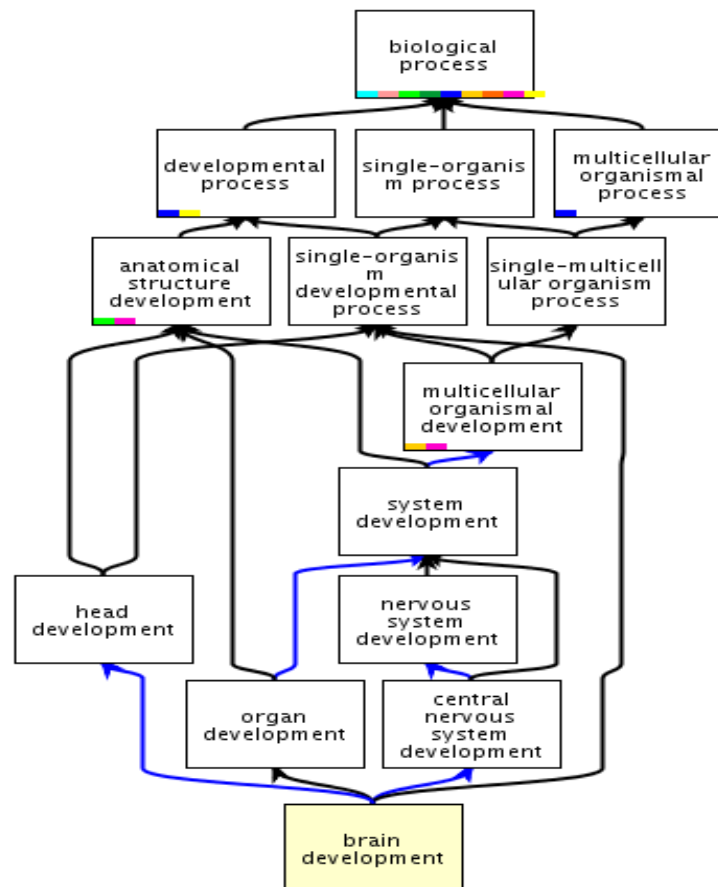
Master 1 – GENIOMHE – 2015/2016

INTRODUCTION

- **Les données : Liste de Gènes + GO ID**

- **La question :** quelles sont les fonctions biologiques majoritaires ?

- **La méthode :** Etude de surreprésentation de termes GO.

- **Comment ?** Compter le nombre d'occurrence des termes GO pour

  notre sélection de gènes, en utilisant les GOslims

1

GENE ONTOLOGY

QuickGO – http://www.ebi.ac.uk/QuickGO

DATASET

- ## Fichier go.obo:

    [Term]
    id: GO:0071979
    name: cytoskeleton-mediated cell swimming
    namespace: biological_process
    def: "Cell motility in which contractile cytoskeletal elements alter cell shape, resulting in the smooth movement of a cell through a liquid medium." [PMID:18461074]
    subset: gosubset_prok
    synonym: "cytoskeleton-mediated swimming motility" EXACT [GOC:mah]
    is_a: GO:0071975 ! cell swimming

- ## Input :

AT1G27320	GO:0006355
AT1G27390	GO:0006626
AT1G27390	GO:0006626
AT1G28380	GO:0045087
AT1G29400	GO:0016558
AT1G31340	GO:0005886

WORKFLOW DE GOLIAT
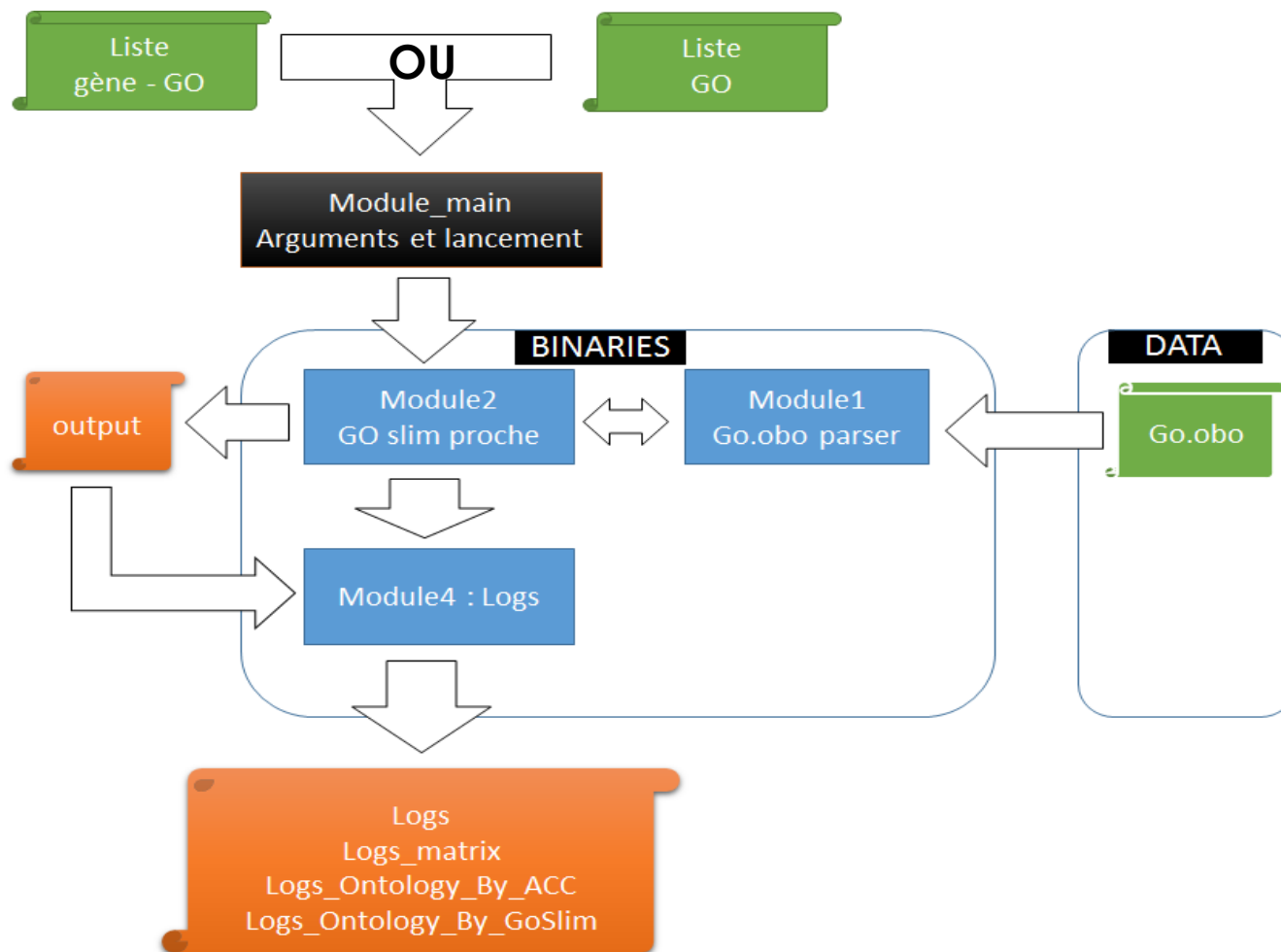
OutPut_(relation).csv

Accession Number	GO	close GOSlim	close GOSlim name	path GOSlims			
AT1G01120	GO:0016020	GO:0016020	membrane	GO:0016020 GO:0005575			
AT1G02730	GO:0042546	GO:0042546	cell wall biogenesis	GO:0008150 GO:0042546 GO:0009987 GO:0044085 GO:0071554
AT1G26830	GO:0005515	GO:0005515	protein binding	GO:0003674 GO:0005488 GO:0005515			
AT1G27320	GO:0006355	GO:0006355	regulation of transcription, D	GO:0097659 GO:0043170 GO:0006351 GO:0009987 GO:0006139
AT1G27390	GO:0006626	GO:0006605	protein targeting	GO:0006886 GO:0045184 GO:0051234 GO:0006605 GO:0051649
AT1G27390	GO:0006626	GO:0006605	protein targeting	GO:0006886 GO:0045184 GO:0051234 GO:0006605 GO:0051649
AT1G28380	GO:0045087	GO:0006952	defense response	GO:0008150 GO:0050896 GO:0006952 GO:0006950 GO:0002376
AT1G29400	GO:0016558	GO:0006605	protein targeting	GO:0006886 GO:0045184 GO:0051234 GO:0006605 GO:0007031

Logs_(relation).csv

Initial Number Gene:	243	
Compute Number Gene:	242	
Number of you own GOSlims:	2	
Number of unique close GOSlim:	119	

GOSlim	List_gene_present	List_gene_unpresent
GO:0016310	AT1G53700 AT1G5370(	AT1G01120 AT1G02730 AT1G26830 AT1G27320 AT1G27390 AT1G2
GO:0080134	AT2G41940 AT2G4194(	AT1G01120 AT1G02730 AT1G26830 AT1G27320 AT1G27390 AT1G2
GO:0009617	AT4G18780 AT4G1878(	AT1G01120 AT1G02730 AT1G26830 AT1G27320 AT1G27390 AT1G2
GO:0034641	AT1G27320 AT1G5449(	AT1G01120 AT1G02730 AT1G26830 AT1G27390 AT1G27390 AT1G2
GO:0005886	AT1G31340 AT1G3421(	AT1G01120 AT1G02730 AT1G26830 AT1G27320 AT1G27390 AT1G2
GO:0042623	AT3G10270	AT1G01120 AT1G02730 AT1G26830 AT1G27320 AT1G27390 AT1G2
GO:0050794	AT1G27320 AT2G2254(	AT1G01120 AT1G02730 AT1G26830 AT1G27390 AT1G27390 AT1G2
GO:0005834	AT3G15354	AT1G01120 AT1G02730 AT1G26830 AT1G27320 AT1G27390 AT1G2
GO:0006812	AT1G05580	AT1G01120 AT1G02730 AT1G26830 AT1G27320 AT1G27390 AT1G2

Logs_matrix_(relation).csv

GOSlim	AT1G01120	AT1G02730	AT1G26830	AT1G27320	AT1G27390	AT1G27390	AT1G28380	AT1G29400	AT1G31340	AT1G32450	AT1G0273
GO:00163100	0	0	0	0	0	0	0	0	0	0	
GO:00801340	0	0	0	0	0	0	0	0	0	0	
GO:00096170	0	0	0	0	0	0	0	0	0	0	
GO:00346410	0	0	1	0	0	0	0	0	0	0	
GO:00058860	0	0	0	0	0	0	0	1	0	0	
GO:00426230	0	0	0	0	0	0	0	0	0	0	
GO:00507940	0	0	1	0	0	0	0	0	0	0	
GO:00058340	0	0	0	0	0	0	0	0	0	0	
GO:00068120	0	0	0	0	0	0	0	0	0	0	
GO:00094140	0	0	0	0	0	0	0	0	0	0	
GO:00713100	0	0	0	0	0	0	0	0	0	0	

Logs_Ontology_By_GoSlim_(relation).csv

Number of Gene with a close MF GOSlim	59	
Number of Gene with a close BP GOSlim	139	
Number of Gene with a close CC GOSlim	44	
Compute Number Gene :	242	

GOSlim	Ontology	count of gene
GO:0016310	biological_process	7
GO:0080134	biological_process	3
GO:0009617	biological_process	6
GO:0034641	biological_process	21
GO:0005886	cellular_component	8

Logs_Ontology_By_ACC_(relation).csv

| Accession Number	Ontology	GOSlims				
AT1G01120	cellular_component	GO:0016020 GO:0005575				
AT1G02730	biological_process	GO:0008150 GO:0042546 GO:0009987 GO:0044085 GO:0071554
AT1G26830	molecular_function	GO:0003674 GO:0005488 GO:0005515				
AT1G27320	biological_process	GO:0097659 GO:0043170 GO:0006351 GO:0009987 GO:0006139
AT1G27390	biological_process	GO:0006886 GO:0045184 GO:0051234 GO:0006605 GO:0051649
AT1G27390	biological_process	GO:0006886 GO:0045184 GO:0051234 GO:0006605 GO:0051649

Launch_ChildGO
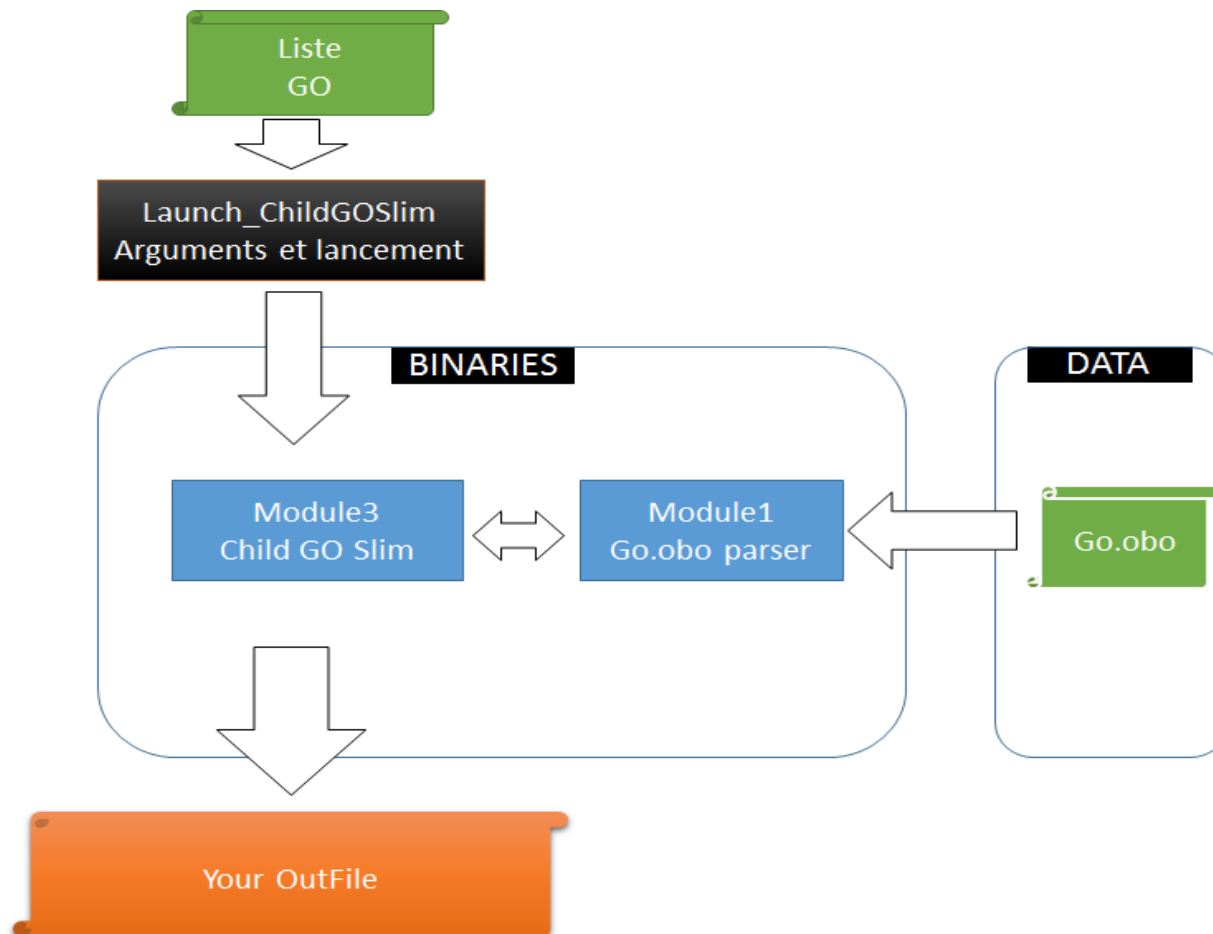
ChildrenOutput.csv

| Numero d'accesion	Numero GO	goslim_chemin_children
AT1G02730	GO:0042546	GO:0044085 GO:0071554
AT1G26830	GO:0005515	GO:0005488
AT1G27320	GO:0006355	GO:2000112 GO:1903506 GO:0010468
AT1G27390	GO:0006626	GO:0006605 GO:0016482 GO:0072655 GO:1902580 GO:0070585 GO:0006839
AT1G27390	GO:0006626	GO:0006605 GO:0016482 GO:0072655 GO:1902580 GO:0070585 GO:0006839
AT1G28380	GO:0045087	GO:0006952 GO:0006955
AT1G29400	GO:0016558	GO:0006625 GO:0044743
AT1G31340	GO:0005886	GO:0044464 GO:0016020
AT1G32450	GO:0010167	GO:0001101 GO:1901698 GO:1901700 GO:0010035
AT1G02730	GO:0042546	GO:0044085 GO:0071554
AT1G34210	GO:0005886	GO:0044464 GO:0016020

Launch_cytoscape :
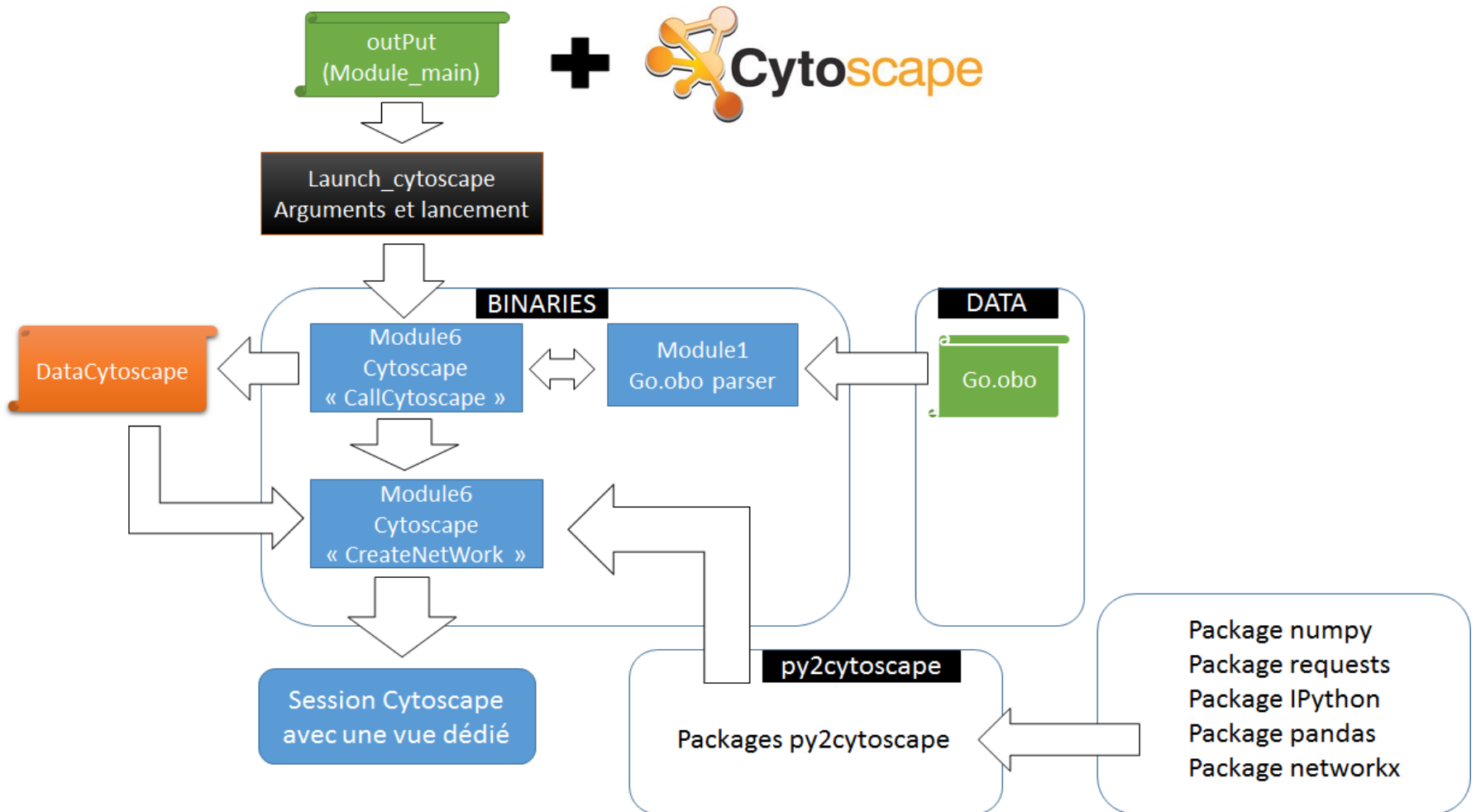
DataCytoscape_(relation).csv

Source	Interaction	Target	Ontology
AT1G15510	gene-go	GO:0009451	no_ontology
AT5G44340	gene-go	GO:0007017	no_ontology
AT3G54340	gene-go	GO:0000003	no_ontology
AT4G24280	gene-go	GO:0006457	no_ontology
AT4G18780	gene-go	GO:0042742	no_ontology
AT1G29400	gene-go	GO:0006605	no_ontology
AT2G29130	gene-go	GO:0009414	no_ontology
AT3G05630	gene-go	GO:0045892	no_ontology
AT1G52520	gene-go	GO:0005634	no_ontology

VISUALISATION

CYTOSCAPE …

12
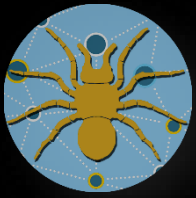
CONCLUSION

Ce que fait GOLIAT	Ce que pourrait faire GOLIAT
Prise en compte des relations : is_a, part_of, regulates, positevely_regulates et negatively regulates	Prise en compte des autres relations présentes dans le fichier go.obo
Liste de GO ID pour lesquels sont récupérés les GOslims ancestraux	Sélection de subset en paramètre
Liste de GO ID pou lesquels sont récupérés les Gos fils	Extraction d'informations du fichier de sortie
Création de fichier tabulé pour R	Enrichissement d'informations dans les fichiers de sorties
Visualisation Cytoscape	Amélioration de la visualisation

MERCI DE VOTRE ATTENTION !