Rice gene functional data

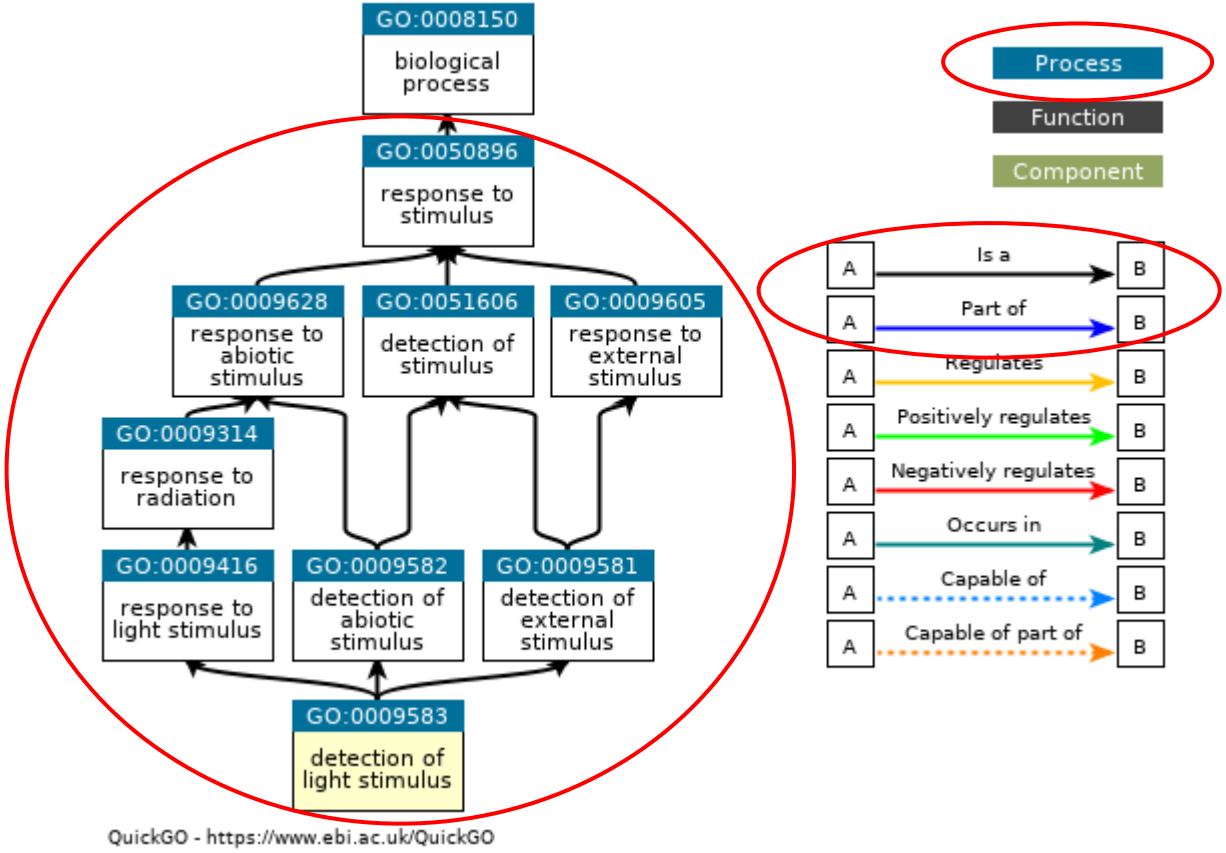
Miguel Romero
Ómicas P5
Sep 3rd, 2020

Gene functions

GO:0009583 detection of light stimulus

Ancestor Chart

Function and all its ancestors



Biological processes

Gene functions

Hypothesis

If a gene is associated to a function (biological process) then it is associated to all ancestors of the function.

Total data

Genes	19.663
Co-expression interactions	550.813
Gene-function associations	220.598
Genes with function	9.294
Functions	3.743
Ancestral relations	7.186

Taken from: RAPDB, Oryzabase, Gene2GO, QuickGO, and Gene Ontology.

Term hierarchies

Filtered terms (5<x<=300)	1.797
Terms to predict (including parents)	1.938
Components of DAG	27
Subhierarchies	20
Average number of terms	340
Min-Max number of terms	2-1996

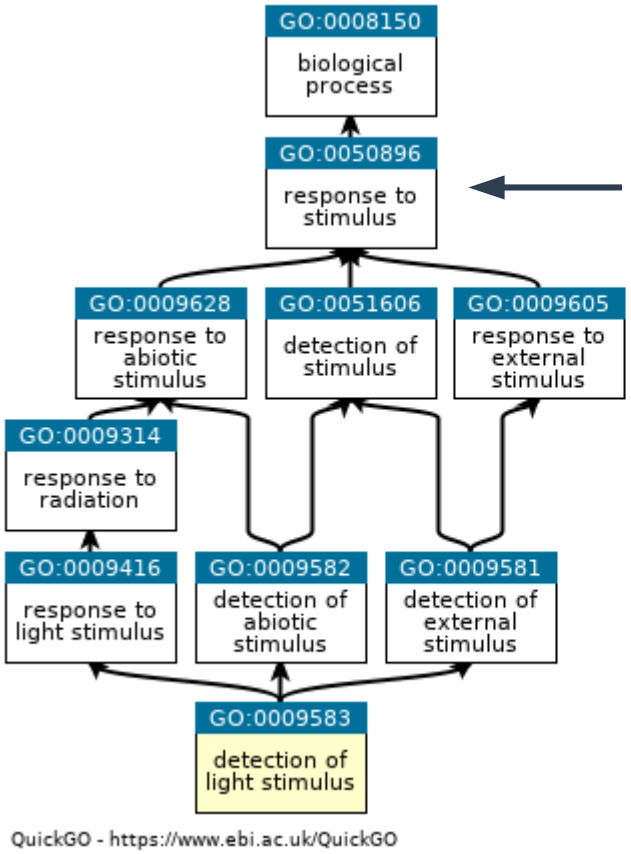
Term hierarchies

Root	Terms	Pred	Genes	Desc
GO:0044085	121	50	377	cellular component biogenesis
GO:0000003	146	72	648	reproduction
GO:0006796	209	118	1270	phosphate-containing compound metabolic process
GO:0032501	250	120	1043	multicellular organismal process
GO:0032502	290	149	1063	developmental process
GO:0016043	298	140	661	cellular component organization
GO:0051179	325	164	1350	localization
GO:0050896	470	261	3319	response to stimulus
GO:0065007	1027	485	2224	biological regulation
GO:0008152	1463	779	5862	metabolic process
GO:0009987	1996	1025	5900	cellular process

Prediction of gene functions

GO:0009583 detection of light stimulus

Ancestor Chart



Root of hierarchy

Genes associated to root term are use for prediction (hypothesis)

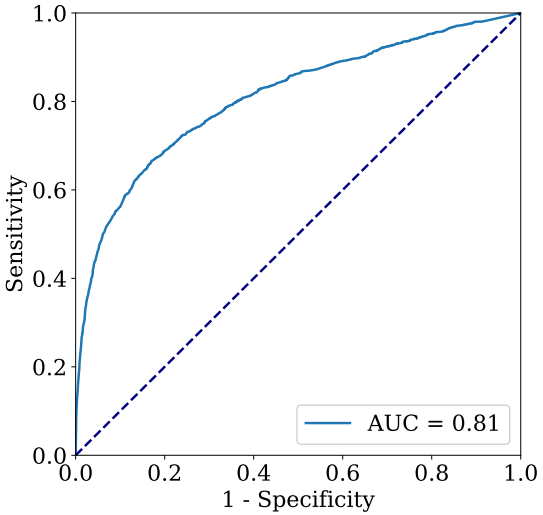
Top-down strategy. Term by term, based on the ancestor.

Term hierarchies

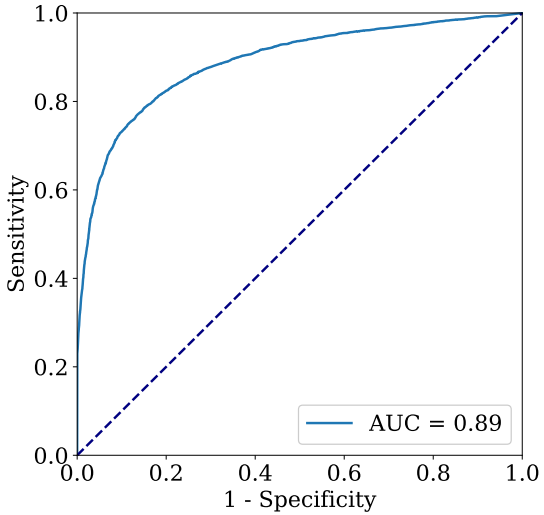
Root	Terms	Genes	Desc	AUC
GO:0044085	50	377	cellular component biogenesis	0.84
GO:0000003	72	648	reproduction	0.88
GO:0006796	118	1270	phosphate-containing compound metabolic process	0.87
GO:0032501	120	1043	multicellular organismal process	0.87
GO:0032502	149	1063	developmental process	0.87
GO:0016043	140	661	cellular component organization	0.81
GO:0051179	164	1350	localization	0.89
GO:0050896	261	3319	response to stimulus	0.89
GO:0065007	485	2224	biological regulation	0.90
GO:0008152	779	5862	metabolic process	--
GO:0009987	1025	5900	cellular process	--

HBN model

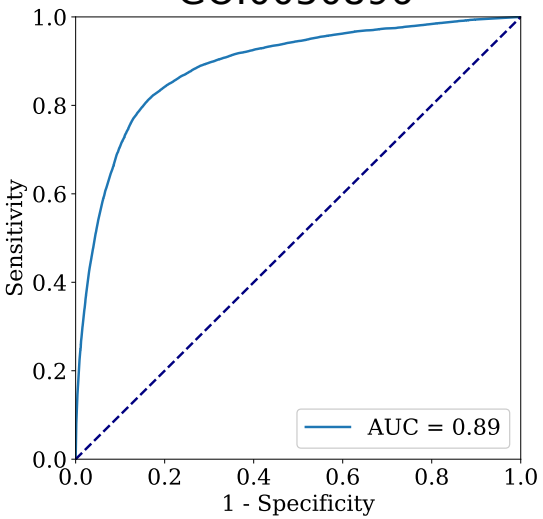
GO:0016043



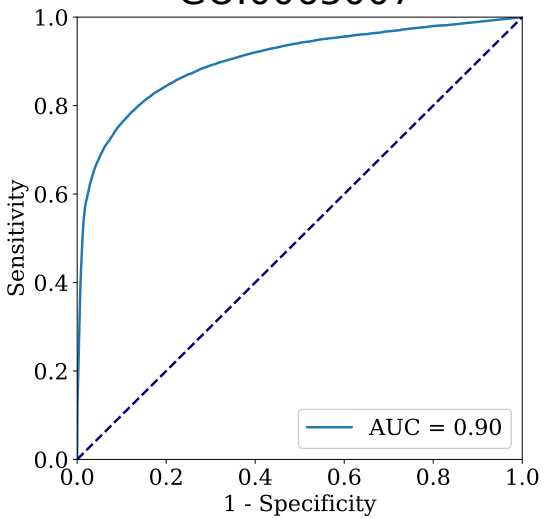
GO:0051179



GO:0050896



GO:0065007



To do

- Experiments with HBN model on Hydra (ongoing).
- Experiments with ML model on Hydra (next).
  - Define attributes and datasets (ongoing).
  - Using graph embeddings *node2vec* and *GCN* (ongoing).
  - Run experiments for every dataset.
- Compare performance of the models.
- Write a paper! (hopefully soon).