

GO-ing beyond Term Enrichment

Maximizing *A. thaliana* GO
annotations

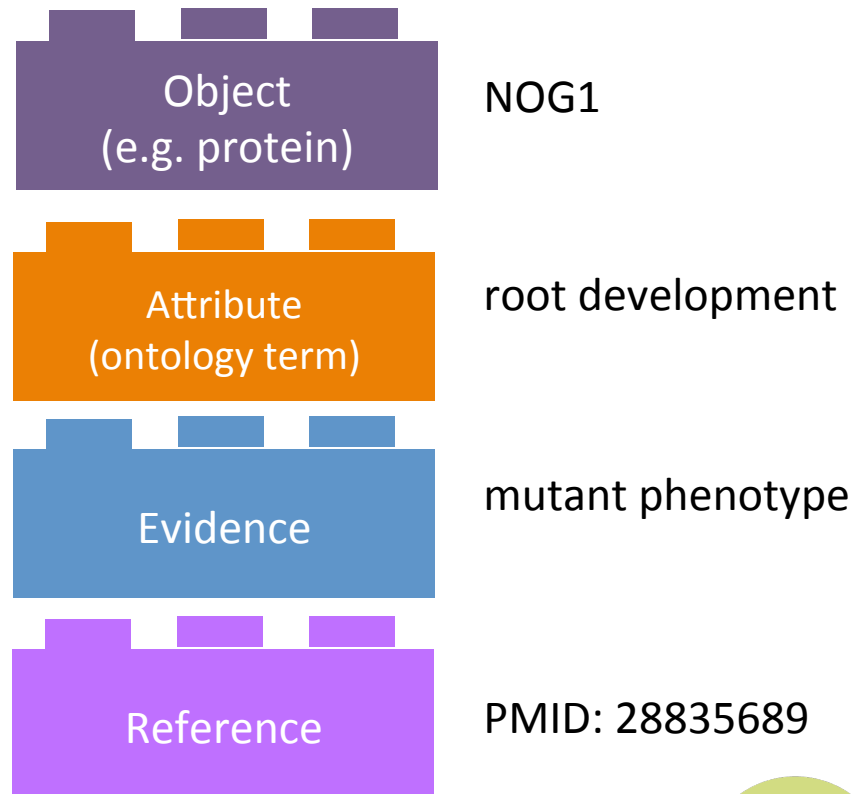
Tanya Berardini,

The Arabidopsis Information Resource



Ontology Annotations

- ... a statement that a gene product:
 1. has a particular **molecular function** *or* is involved in a particular **biological process** *or* is located within a certain **cellular component** *or* is expressed in a particular **anatomical structure** *or* is expressed during a particular **developmental stage**
 2. as determined by a particular **method**
 3. as described in a particular **reference**



Example

yeast-two hybrid system. Yeast two-hybrid analysis revealed that CCA1, but not the other proteins in the photoperiod pathway, interact with MLK4 (Figure 4A). These results were confirmed using a protein pull-down assay; MLK4 fused to GST was observed to bind beads containing His fused to CCA1, but not His beads alone

8

Locus/Gene Model	Gene Symbol/Full Name	Relationship Type	Keyword	Keyword Category	Evidence Code [?] : Evidence Description [?] : Evidence With: Reference [?]	Annotated By/ Date Last Modified
AT2G46830	CCA1/ CIRCADIAN CLOCK ASSOCIATED 1	functions in	protein binding	molecular function	<i>inferred from physical interaction:</i> Co-immunoprecipitation: AT3G13670: Su, et al. (2017)	The Arabidopsis Information Resource/ 2018-01-06



GO Evidence Codes



Inferred from

- Direct Assay (IDA)
- Physical Interaction (IPI)
- Mutant Phenotype (IMP)
- Genetic Interaction (IGI)
- Expression Pattern (IEP)



Inferred from

- Sequence or structural Similarity (ISS)
- Electronic Annotation (IEA)
- Others...



- Traceable Author Statement (TAS)
- Non-traceable Author Statement (NAS)
- Inferred by Curator (IC)
- No Biological Data available (ND)



Who makes experimental GO annotations?

Annotations per Assigned By

	%	Count
TAIR	70.55	49,025
UniProt	17.81	12,375
IntAct	10.05	6,984
CAFA	1.06	738
ComplexPortal	0.33	227
CACAO	0.20	136
GO_Central	0.00	3
AgBase	0.00	2

Gene products per Assigned By

	%	Count
TAIR	90.51	11,745
UniProt	24.89	3,230
IntAct	12.37	1,605
CAFA	2.46	319
ComplexPortal	0.76	99
CACAO	0.81	105
GO_Central	0.02	2
AgBase	0.02	2



Things to do with GO annotations:

Small scale

- Compare one gene's annotations to another
- Find an GO term of interest, find all genes in a single species annotated to this term
- Find a GO term of interest, find all genes in all annotated species annotated to this term



Things to do with GO annotations:

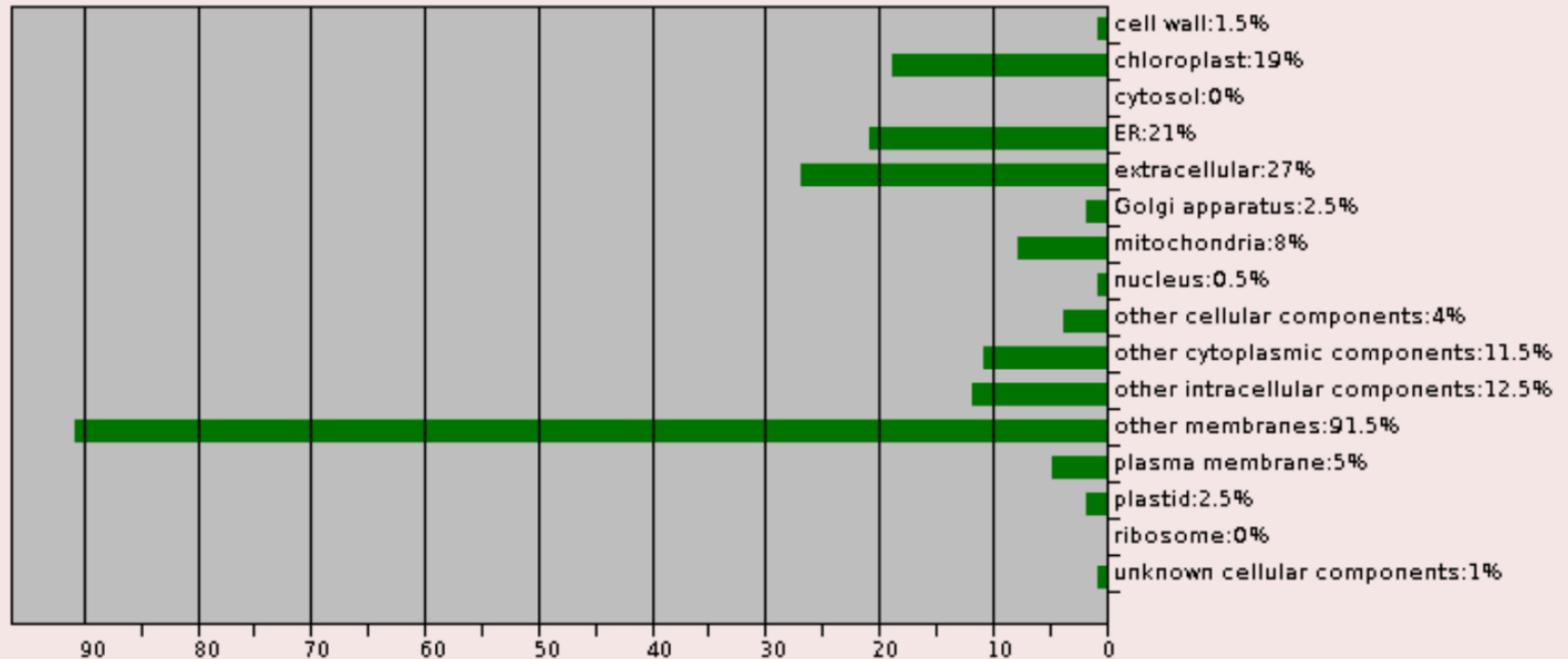
Large scale

- Project from Arabidopsis experimental annotations to other species
- **Functional Categorization:** Take a list of annotated genes and see how their annotations group in broad categories
- **Term Enrichment:** A list of genes/proteins is compared to a reference list (usually the whole genome) to statistically determine **over- or under-representation** of the selected ontology categories among genes/proteins in the list.



Functional Categorization @ TAIR

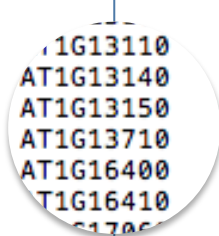
Functional Categorization by loci for : GO Cellular Component
N = 200



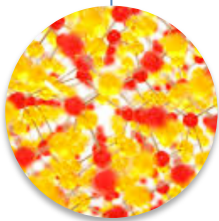
New Tool: Overrepresented processes in my set of genes?



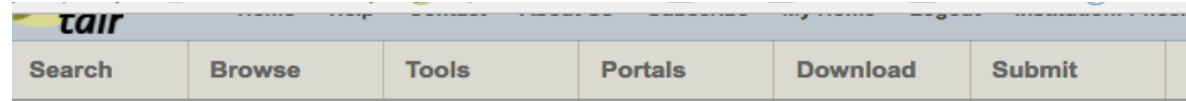
Differential expression analysis :
normal vs. salty conditions



Genes with relative increased expression



Similarities/
Differences



GO Term Enrichment for Plants

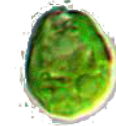
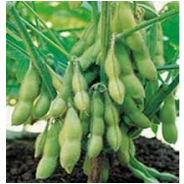
Statistical Over/Under Representation (powered by PANTHER)

Use this tool to identify Gene Ontology terms that are over or under-represented in a set of genes (for example from co-expression or RNAseq data). The data are sent to the PANTHER Classification System which contains up to date GO annotation data for Arabidopsis and other plant species. Choose the advanced setting if you want to change parameters or explore PANTHER's other tools for analyzing sets of genes. [Help]

Enter a list of valid identifiers, separated by newline.

[Try a sample gene list](#)

```
Solyc04g011550.2.1
Solyc06g068130.2.1
Solyc06g036170.1.1
Solyc01g087750.2.1
Solyc01g112260.2.1
Solyc12g098890.1.1
Solyc03g005090.2.1
Solyc05g005550.2.1
Solyc03g098020.2.1
Solyc02g067690.2.1
```



Choose Organism

- Arabidopsis thaliana
- False Brome (Brachypodium distachyon)
- Chlamydomonas reinhardtii
- Soybean (Glycine max)
- Rice (Oryza sativa)
- Moss (Physcomitrella patens)
- Poplar (Populus trichocarpa)
- Tomato (Solanum lycopersicum)
- Sorghum (Sorghum bicolor)
- Grape (Vitis vinifera)

http://www.arabidopsis.org/tools/go_term_enrichment.jsp

Data Source: Nguyen, et al., Tomato Golden2-like transcription factors reveal molecular gradients that function during fruit development and ripening. (<https://doi.org/10.1105/tpc.113.118794>)



GO Term Enrichment Results

	Arabidopsis thaliana (REF)		upload_1 (▼ Hierarchy NEW! ?)			
	b #	c #	expected	Fold Enrichment	+/-	P value
GO biological process complete						
thylakoid membrane organization	29	6	d .26	e 22.73	+	7.51E-04
↳ plastid membrane organization	30	6	.27	21.97	+	9.13E-04
↳ membrane organization	124	8	1.13	7.09	+	4.69E-02
↳ cellular process	8843	122	80.50	1.52	+	7.68E-05
translational elongation	58	6	.53	11.36	+	3.86E-02

Powered by PANTHER.



Term Enrichment tools:

Pay Attention to Data Dates

- GO Term Enrichment (powered by Panther)
 - [Gene Ontology](#) (from GO database version 1.2, released 2018-06-01)
- AgriGO
 - Last maintain: 2017-04-04
- DAVID
 - GENE ONTOLOGY: Apr 2016, May 2016



What else?

AT2G46830	CCA1/ CIRCADIAN CLOCK ASSOCIATED 1	functions in	protein binding	molecular function	<i>inferred from physical interaction:</i> Co-immunoprecipitation:	The Arabidopsis Information Resource/ 2018-01-06
AT1G73030	VPS46.2	involved in	seedling	biological	<i>inferred from genetic interaction:</i>	
AT5G14750	MYB66/ MYB DOMAIN PROTEIN 66	functions in	DNA binding	molecular function	<i>inferred from physical interaction:</i> yeast one-hybrid assay: AT2G46410: Ryu, et al. (2005)	

- Mine “Evidence with”
 - Protein binding partners (IPI)
 - DNA binding targets (IPI)
 - Genetic interactions (IGI)
- Mine Experimental methods



Sample: Molecular Function Annotation Evidence Descriptions

Description	# of annotations
Enzyme assays	1996
Yeast two-hybrid assay	1316
Co-purification	933
Co-immunoprecipitation	594
In vitro binding assay	422
Yeast one-hybrid assay	416
Functional complementation in heterologous system	385

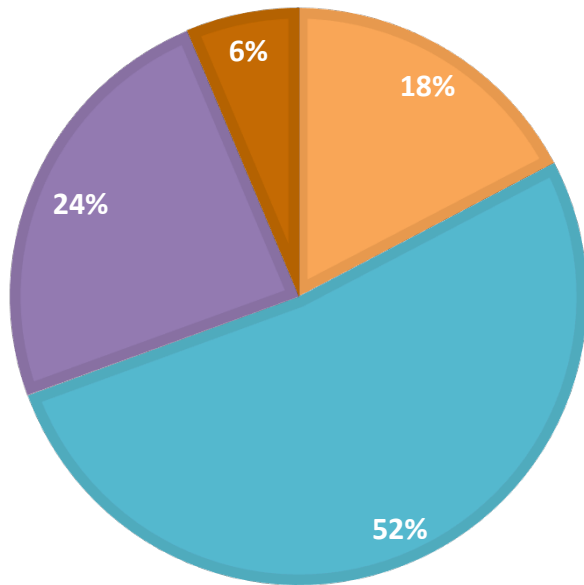


HOW CAN YOU HELP?

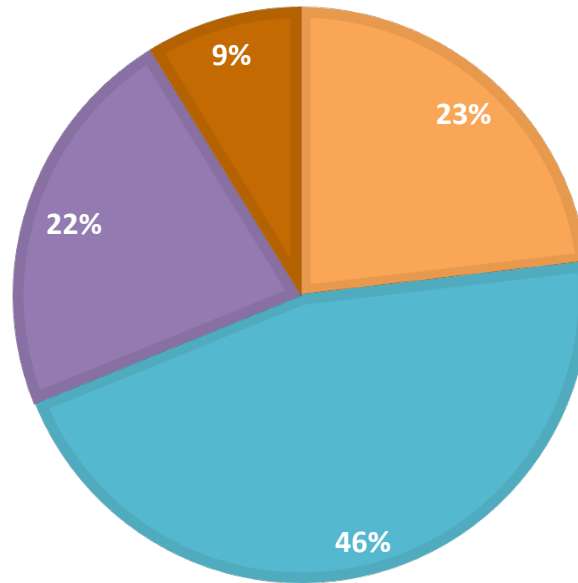


Arabidopsis thaliana Functional Annotation (Araport 11, as of June 2018)

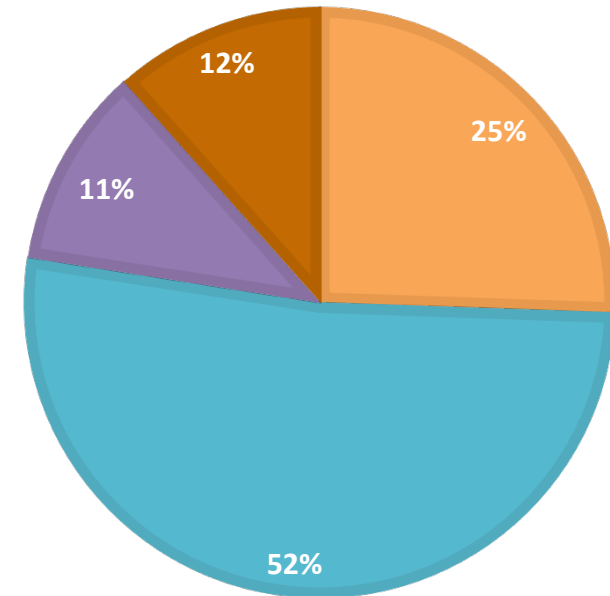
GO MOLECULAR FUNCTION



GO BIOLOGICAL PROCESS



GO CELLULAR COMPONENT



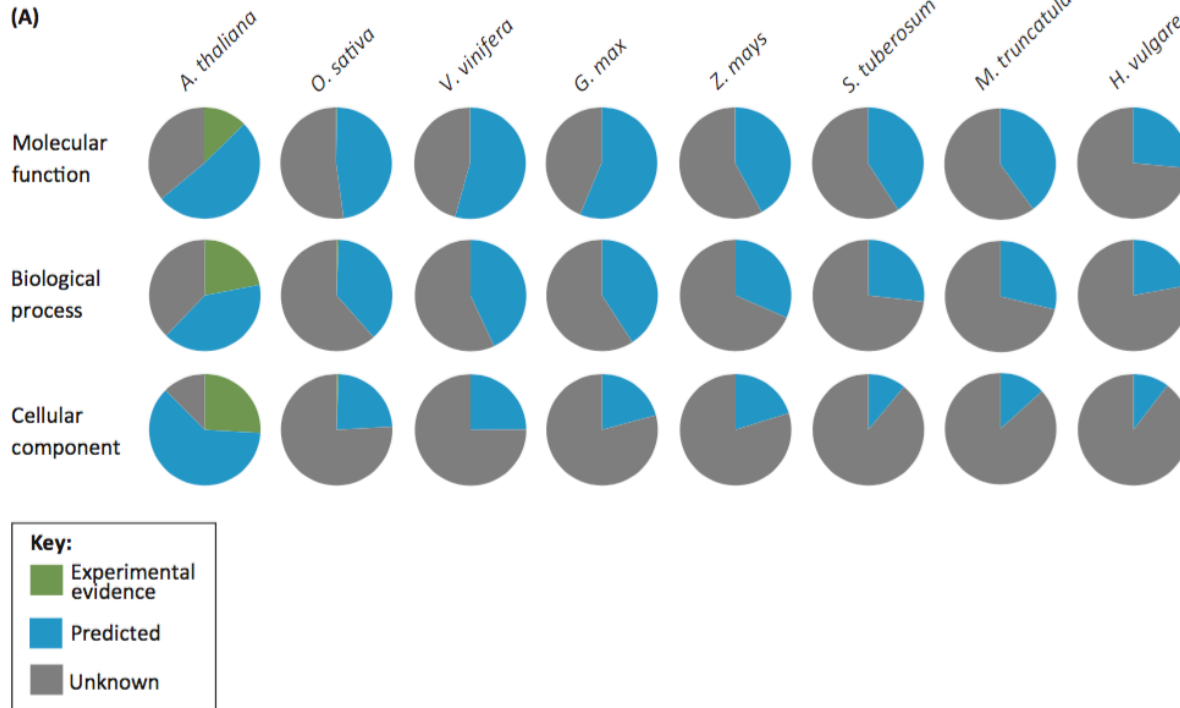
- Term Enrichment
- Functional classification
- Structural/Homology based assignments



Other Plants

Review

Trends in Plant Science April 2014, Vol. 19, No. 4



Rhee and Mutwil. 2014. TIPS. Vol. 9 No. 4 212-221



Harness the power of the community's knowledge

- Gene functional information (function, expression, binding interactions, comments)
- Community curation tool developed by Phoenix and RIT SE Seniors
- Applicable to **any** gene product (protein or ncRNA) from **any** organism



New Annotation Submission

1. Publication

Enter a PubMed ID or a DOI.

Publication ID 16223361 ✓

Title Characterization of an Arabidopsis inositol 1,3,4,5,6-pentakisphosphate

Author Sweetman D

2. Genes

Enter genes with a UniProt ID, AGI locus ID, or RNA Central ID. Optionally enter a gene symbol and full name.

Locus 1 AT5G42810 ✓ ✕

Gene Symbol IPK1

Full Gene Name inositol 1,3,4,5,6-pentakisphosphate 2-kinase 1

[+ Add Another Gene](#)

3. Annotations

Select an annotation format and a gene. All fields are required.

Annotation 1 Molecular Function (GO Function) ⇅ ✕

Gene AT5G42810 ⇅

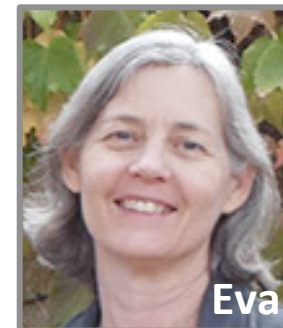
Molecular Function (GO Function) inositol pentakisphosphate 2-kinase activity

Method enzyme assay evidence used in manual assertion

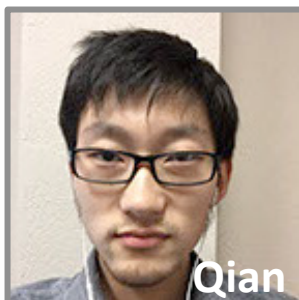
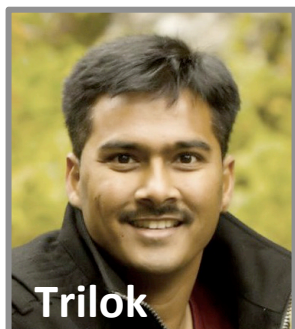
[+ Add Another Annotation](#)[Reset Form](#)[Review Submission >](#)

Phoenix Bioinformatics

Director

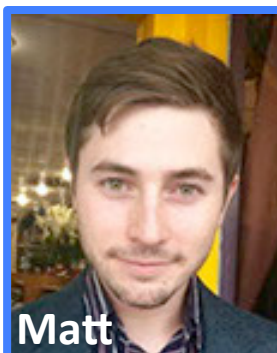
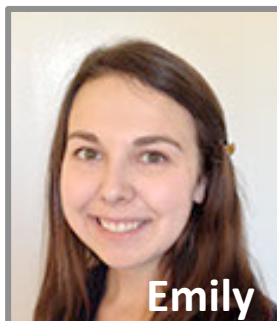


Tech Team

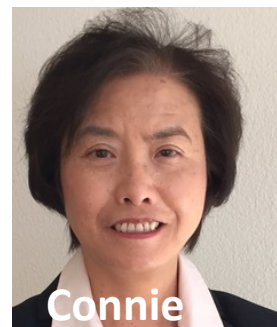


Swapnil

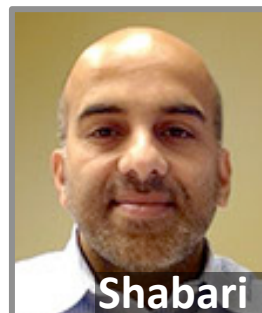
Sales Team



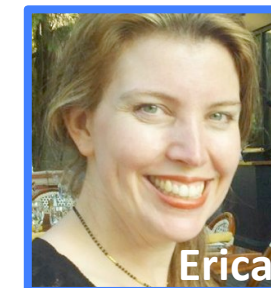
Finance



Business Development



Science Team



- curator@arabidopsis.org
- AgBioData Consortium Booth in Exhibit Hall

