

# TAIR Workshop

## Part II: Tips for Accessing TAIR Gene Function Data

Donghui Li

ICAR, June 24, 2011, Madison WI

download this talk from <http://min.us/u/Tair>

# TAIR Workshop

## Part II: Tips for Accessing TAIR Gene Function Data

Donghui Li

ICAR, June 24, 2011, Madison WI

download this talk from <http://min.us/u/Tair>

# Sources of TAIR gene function information

- Primary literature
- Computational analyses of the genome sequence

- 1) InterPro scan (predicts function based on protein domain signature)
- 2) TargetP (predicts subcellular localization based on N-terminal presequences)

*Both done on a whole-genome scale with each TAIR genome release*

- Contributions from the research community
  - author submission
  - NSF Arabidopsis 2010 program
- Additional sources:

Gene Ontology Consortium, Gene Ontology Annotation (GOA) at European Bioinformatics Institute (EBI)

TAIR provides comprehensive up-to-date  
gene function information

gene names and synonyms

short summary

controlled vocabulary (GO, PO)

annotations (function and expression)

polymorphism and germplasm information

phenotype

publication

# Outline

- How gene function data are organized on the TAIR locus page
- How to correctly interpret Gene Ontology (GO) annotations
- Selected examples of data access and analysis tools

# Access function information from TAIR locus page

or At1g09570

PHYA Gene

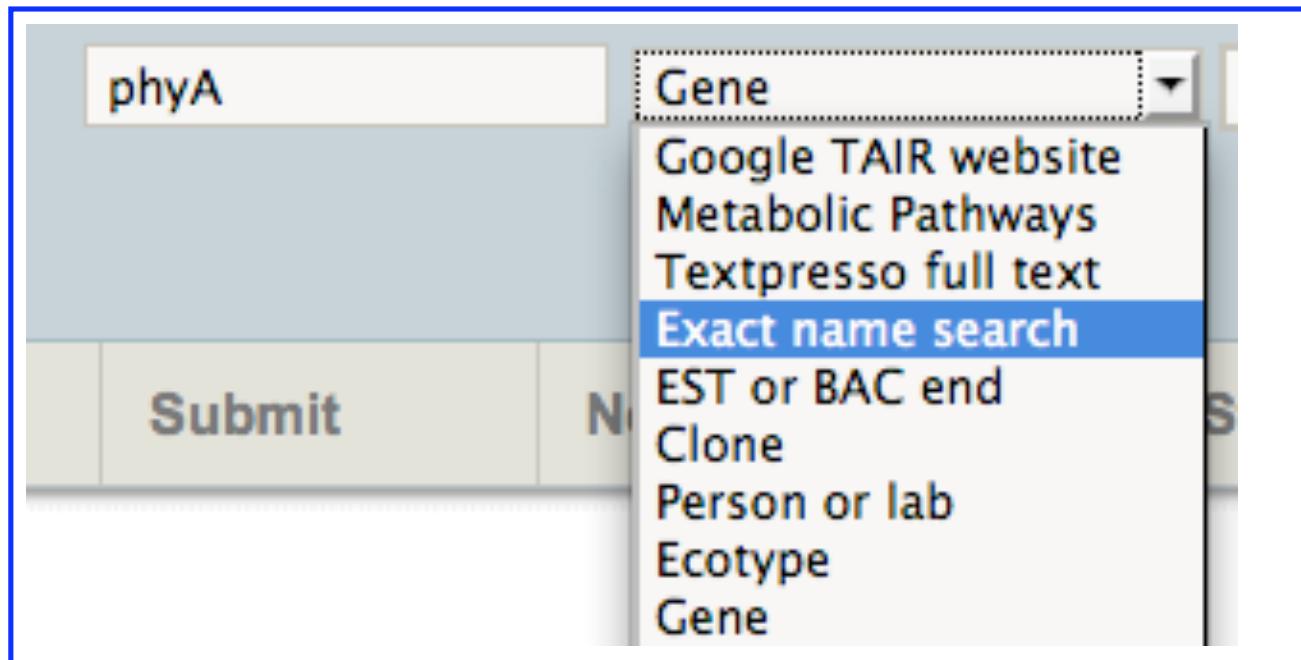
 tair Home Help Contact About Us Login/Register

Your query for genes where gene name, description, phenotype, locus name, uniprot id or GenBank accession contains the term **phyA** resulted in **21 loci** matches with **30** distinct gene models.

Displaying 1 - 21.

To see ESTs associated with your gene of interest, click on the Locus link.

	<a href="#">Check All</a>	<a href="#">Uncheck All</a>			
	Locus	Description	Gene Model(s)	Other Names	Keywords
1	<a href="#">AT1G09570</a>	Light-labile cytoplasmic red/far-red light photoreceptor involved in the regulation of photomorphogenesis. It exists in two i....	AT1G09570.1	ELONGATED HYPOCOTYL 8 F14J9.23 F14J9_23 FAR RED ELONGATED 1 <b>FAR RED ELONGATED</b> <b>HYPOCOTYL 2</b> PHY2 FRE1 HY8 <b>PHYA</b> PHYTOCHROME A	4 anthesis, 4 leaf senescence stage, C globular stage, D bilateral stage, DNA-dependent, E expanded cotyledon stage, F mature embryo stage, G-protein coupled photoreceptor activity, LP.02 two leaves visible, LP.04 four leaves visible, LP.06 six leaves visible, LP.08 eight leaves visible, LP.10 ten leaves visible, LP.12 twelve leaves visible, carpel, caulin leaf, cotyledon, cytoplasm, embryo, flower, gravitropism, hypocotyl, inflorescence meristem, leaf, leaf apex, leaf lamina base, leaf whorl, male gametophyte, nuclear body, nucleus, pedicel, petal, petal differentiation and expansion stage, petiole, phototropism, protein histidine kinase activity, red light signaling pathway, red or far-red light photoreceptor activity, regulation of transcription, regulation of transcription, DNA-dependent, response to arsenic, response to continuous far red light stimulus by the high-irradiance response system, response to far red light, response to very low fluence red light stimulus, root, seed, sepal, shoot, shoot apex, signal transducer activity, stamen, stem



A screenshot of the TAIR website showing search results. At the top, there is a navigation bar with links for Home, Help, Contact, About Us, and Login/Register. On the far right of the navigation bar is a search bar containing "PHYA", a dropdown menu set to "Gene", and a "Search" button. Below the navigation bar is a horizontal menu with links for Search, Browse, Tools, Stocks, Portals, Download, Submit, and News. A blue box surrounds the search bar and its associated dropdown menu. A blue arrow points from the "Exact name search" option in the dropdown menu above to the search bar on this page.

Your query '**PHYA**' matched records in multiple types. Select from the list below to see selected records of that type.

3 genetic marker(s) were found

2 polymorphism(s) were found

1 locus/loci were found

2 protein(s) were found

# Access function information from TAIR locus page

<b>Locus:</b> AT1G09570	<b>locus identifier</b>
Date last modified	2003-05-02
TAIR Accession	Locus:2012300
Representative Gene Model	<a href="#">AT1G09570.1</a>
Gene Model Type	protein_coding
Other names:	ELONGATED HYPOCOTYL 8, F14J9.23, F14J9_23, FAR RED ELONGATED 1, FAR RED ELONGATED HYPOCOTYL 2, FHY2, FRE1, HY8, PHYA, PHYTOCHROME A
Description	Light-labile cytoplasmic red/far-red light photoreceptor involved in the regulation of photomorphogenesis. It exists in two inter-convertible forms: Pr and Pfr (active) and functions as a dimer. The N terminus carries a single tetrapyrrole chromophore, and the C terminus is involved in dimerization. It is the sole photoreceptor mediating the FR high irradiance response (HIR). Major regulator in red-light induction of phototropic enhancement. Involved in the regulation of de-etiolation. Involved in gravitropism and phototropism. Requires FHY1 for nuclear accumulation.

## names

## description

# Access function information from TAIR locus page

## curated description

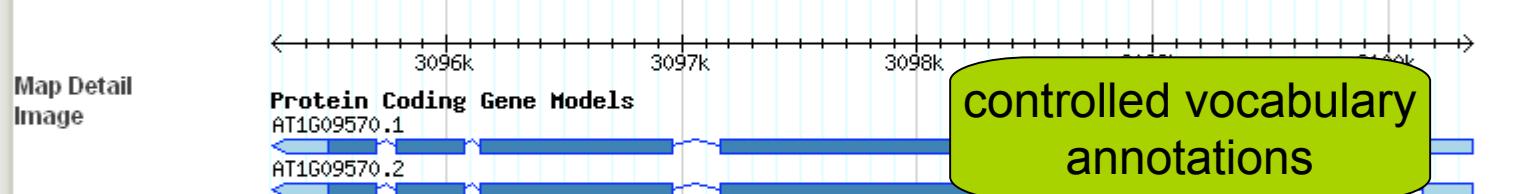
Description Light-labile cytoplasmic red/far-red light photoreceptor involved in the regulation of photomorphogenesis. It exists in two inter-convertible forms: Pr and Pfr (active) and functions as a dimer. The N terminus carries a single tetrapyrrole chromophore, and the C terminus is involved in dimerization. It is the sole photoreceptor mediating the FR high irradiance response (HIR). Major regulator in red-light induction of phototropic enhancement. Involved in the regulation of de-etiolation. Involved in gravitropism and phototropism. Requires FHY1 for nuclear accumulation.

## computational description

Description C2H2 and C2HC zinc fingers superfamily protein;  
FUNCTIONS IN: sequence-specific DNA binding transcription factor activity, zinc ion binding, nucleic acid binding; INVOLVED IN: regulation of transcription;  
LOCATED IN: intracellular; CONTAINS InterPro DOMAIN/s: Zinc finger, C2H2-like (InterPro:IPR015880), Zinc finger, C2H2-type (InterPro:IPR007087); BEST Arabidopsis thaliana protein match is: C2H2-like zinc finger protein (TAIR:AT2G15740.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLINK).

## Locus: AT1G09570

Date last modified	2003-05-02
TAIR Accession	Locus:2012300
Representative Gene Model	<a href="#">AT1G09570.1</a>
Gene Model Type	protein_coding
Other names:	ELONGATED HYPOCOTYL 8, F14J9.23, F14J9_23, FAR RED ELONGATED 1, FAR RED ELONGATED HYPOCOTYL 2, FHY2, FRE1, HY8, PHYA, PHYTOCHROME A
Description	Light-labile cytoplasmic red/far-red light photoreceptor involved in the regulation of photomorphogenesis. It exists in two inter-convertible forms: Pr and Pfr (active) and functions as a dimer. The N terminus carries a single tetrapyrrole chromophore, and the C terminus is involved in dimerization. It is the sole photoreceptor mediating the FR high irradiance response (HIR). Major regulator in red-light induction of phototropic enhancement. Involved in the regulation of de-etiolation. Involved in gravitropism and phototropism. Requires FHY1 for nuclear accumulation.
Other Gene Models	<a href="#">AT1G09570.2</a> (splice variant)



controlled vocabulary annotations

Annotations	Category	Relationship Type	Keyword
	GO Biological Process	involved in	gravitropism, phototropism, regulation of transcription, DNA-dependent, response to arsenic, red light signaling pathway, response to very low fluence red light stimulus, response to continuous far red light stimulus by the high-irradiance response system
			response to far red light
	GO Cellular Component	related to	cytoplasm, nuclear body, nucleus
	GO Molecular Function	has	G-protein coupled photoreceptor activity, protein histidine kinase activity, signal transducer activity, red or far-red light photoreceptor activity
	Growth and Developmental Stages	expressed during	petal differentiation and expansion stage, 4 anthesis, 4 leaf senescence stage, C globular stage, E expanded cotyledon stage, F mature embryo stage, LP.02 two leaves visible, LP.04 four leaves visible, LP.06 six leaves visible, LP.08 eight leaves visible, LP.10 ten leaves visible, LP.12 twelve leaves visible
			D bilateral stage
	Plant structure	expressed in	male gametophyte, seed, embryo, cotyledon, hypocotyl, root, shoot, flower, stamen, carpel, sepal, petal, inflorescence meristem, pedicel, leaf, leaf apex, petiole, stem, cauline leaf, shoot apex, leaf lamina base, leaf whorl

## Gene Ontology annotations

## Plant Ontology annotations

Gene Ontology (GO) has become the standard for gene function annotation

**molecular function:** catalytic / binding activities  
kinase activity, DNA binding activity

**biological process:** biological goal or objective  
signal transduction  
mitosis, purine metabolism

**cellular component:** location or complex  
nucleus  
ribosome, proteasome

More information: <http://www.geneontology.org/>

We use Plant Ontology (PO) to annotate expression

**Plant structure:** morphological and anatomical structures

stamen, petal, guard cell

**Growth and developmental stages:** whole plant growth stages and plant structure developmental stages

seedling growth, rosette growth, leaf development stages

More information: <http://www.plantontology.org/>

# Always check the Annotation Detail page

## TAIR locus page

Map Detail Image

Protein Coding Gene Models  
AT1G09570.1 (HY8,FHY2,FRE1,PHYA)  
Confidence : \*\*\*\* (T2)  
AT1G09570.2 (HY8,FHY2,FRE1,PHYA)  
Confidence : \*\* (T6)

controlled vocabulary annotations

Annotations

category	relationship type	keyword
GO Biological Process	involved in	G-protein coupled receptor protein signaling pathway, detection of visible light, gravitropism, phototropism, protein phosphorylation, red light signaling pathway, red, far-red light phototransduction, regulation of transcription, DNA-dependent, response to arsenic-containing substance, response to continuous far red light stimulus by the high-irradiance response system, response to very low fluence red light stimulus, signal transduction
GO Biological Process	related to	response to far red light
GO Cellular Component	located in	cytoplasm, nuclear body, nucleus
GO Molecular Function	has	G-protein coupled photoreceptor activity, identical protein binding, protein binding, protein histidine kinase activity, red or far-red light photoreceptor activity, signal transducer activity
Growth and Developmental Stages	expressed during	4 anthesis, 4 leaf senescence stage, C globular stage, D bilateral stage, E expanded cotyledon stage, F mature embryo stage, LP.02 two leaves visible, LP.04 four leaves visible, LP.06 six leaves visible, LP.08 eight leaves visible, LP.10 ten leaves visible, LP.12 twelve leaves visible, petal differentiation and expansion stage
Plant structure	expressed in	carpel, cauline leaf, collective leaf structure, cotyledon, flower, hypocotyl, inflorescence meristem, leaf apex, leaf lamina base, obsolete microgametophyte, pedicel, petal, petiole, plant embryo, root, seed, sepal, root apex, shoot system, stamen, stem, vascular leaf

Annotation Detail

# Annotation Detail page



tair

Home Help Contact About Us Login/Register

Search Browse Tools Portals Download Submit News ABRC Stocks

**TAIR Annotation Search [Help]**

Your query for annotations based on the locus **AT1G09570** resulted in **68** records.

Displaying 1 - 68 of 68 records on page 1 of 1 pages.

Locus/Gene Model	Gene Symbol/Full Name	Relationship Type	Keyword	Keyword Category	Evidence Code ⓘ Evidence Description ⓘ Evidence With: Reference ⓘ	Annotated By/ Date Last Modified
AT1G09570	PHYA/ PHYTOCHROME A	involved in	gravitropism	biological process	<i>inferred from mutant phenotype:</i> analysis of visible trait: none: Correll MJ, et al. (2005)	The Arabidopsis Information Resource/ 2005-08-29
AT1G09570	PHYA/ PHYTOCHROME A	involved in	phototropism	biological process	<i>inferred from mutant phenotype:</i> analysis of visible trait: none: Correll, et al. (2003)	The Arabidopsis Information Resource/ 2005-08-29
AT1G09570	PHYA/ PHYTOCHROME A	involved in	regulation of transcription, DNA-dependent	biological process	<i>inferred from sequence or structural similarity:</i> manually reviewed TIGR computational analysis: InterPro:IPR001294: TIGR Arabidopsis annotation team (2005-02-17)	The Institute for Genomic Research/ 2003-09-15

# The anatomy of a GO annotation

Locus/Gene Model	Gene Symbol/Full Name	Relationship Type	Keyword	Keyword Category	Evidence Code <small>?</small> :	Evidence Description <small>?</small> :	Evidence With:	Reference <small>?</small>	Annotated By/ Date Last Modified
AT1G09570	PHYA/ PHYTOCHROME A	involved in	gravitropism	biological process	<i>inferred from mutant phenotype:</i> analysis of visible trait: none: Correll MJ, et al. (2005)				The Arabidopsis Information Resource/ 2005-08-29

# Commonly used evidence codes

## Experimental evidence codes (EXP)

		Evidence description
IDA:	Inferred from Direct Assay	(enzyme assays, in situ hybridization)
IMP:	Inferred from Mutant Phenotype	(analysis of visible trait)
IPI:	Inferred from Physical Interaction	(yeast-2-hybrid)
IEP:	Inferred from Expression Pattern	(RT-PCR, Western blot)
IGI:	Inferred from Genetic Interaction	(double mutant analysis)

## Computational Analysis Evidence Codes (non-EXP)

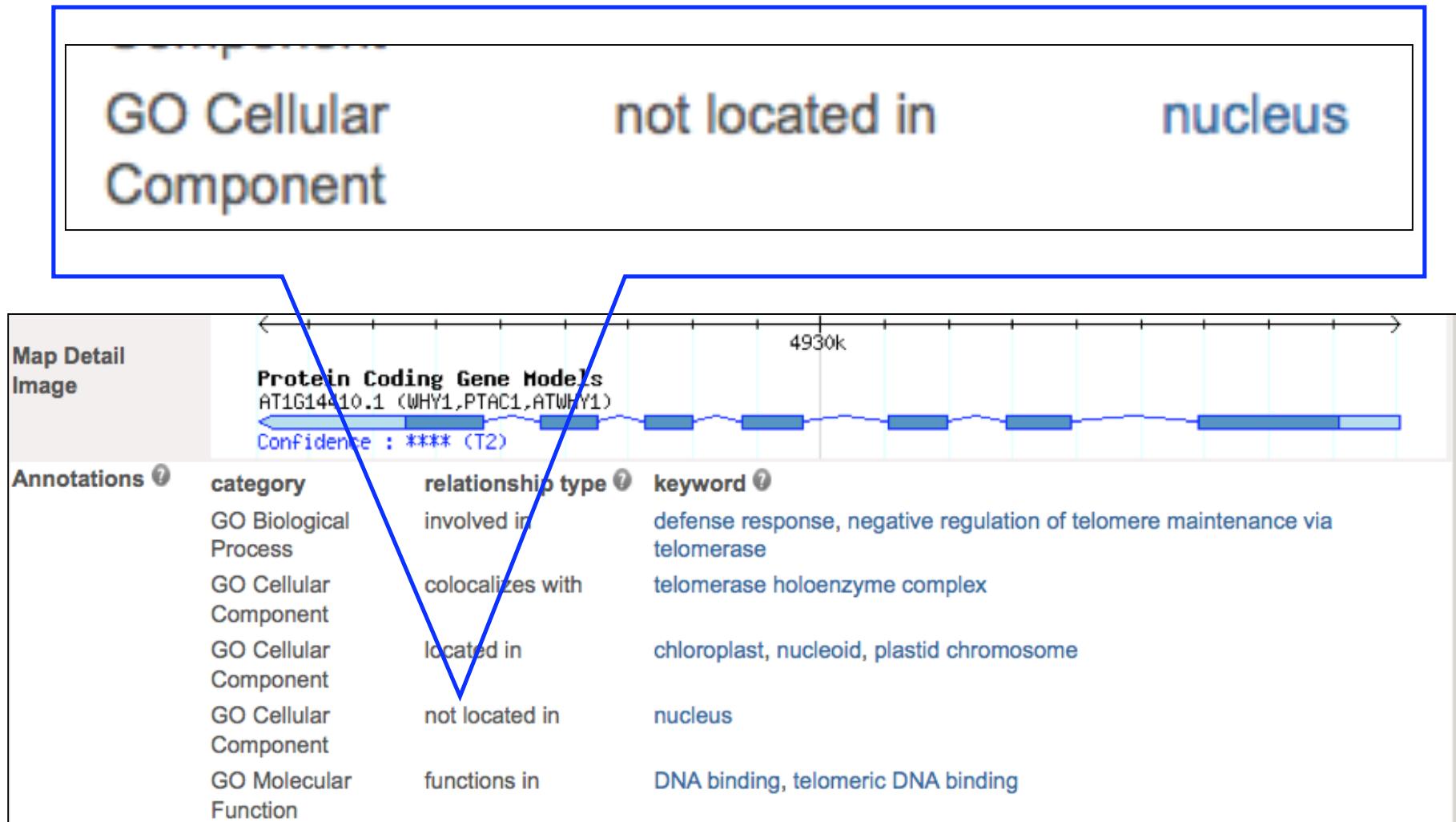
ISS:	Inferred from Sequence or Structural Similarity - based on published sequence alignment
IEA:	Inferred from Electronic Annotation - InterPro scan, TargetP annotations

As of 6/20/2011,  
total number of GO annotations 167,530,  
total number of GO experimental annotations 48,719 (28%)

# Some (but not all) annotations have supporting information in the Evidence with field

ATP binding AT1G14410	molecular function	<i>inferred from electronic annotation:</i> traceable computational prediction: IEA - domain identifier INTERPRO:IPR017442 INTERPRO:IPR000719 INTERPRO:IPR017441 IN analysis reference : InterPro to GO annotation (2010-09-13)
defense response AT4g26090	biological process	<i>inferred from sequence or structural similarity:</i> manually reviewed TIGR computational analysis: NCBI_gi:13661831 NCBI_gi:11357255: ISS - homolog TIGR Arabidopsis annotation team (2005-02-17)
protein binding AT3G25070	molecular function	<i>inferred from physical interaction:</i> Co-immunoprecipitation: AT4G26090: IPI - protein interacting partner Axtell, et al. (2003)
flower development AT2G32700	biological process	<i>inferred from genetic interaction:</i> double mutant analysis: AT4G32551: IGI - other mutated loci in a double, Sitaraman, et al. (2008) triple mutant

# Pay attention to the NOT qualifier in relationship type



# Access function information from TAIR locus page

Polymorphism ?	name ?	type ?	polymorphism site	allele type ?
Showing 15 of 45 entries <a href="#">(see all)</a>	ars4	insertion	unknown	intron
	eid4	substitution	hypermorphic	unknown
	GK-437D07-018192	insertion	unknown	promoter
	GK-483H11-019729	insertion	unknown	intron
	GK-508F06-019540	insertion	unknown	intron
	ossowski_1129689	insertion	unknown	intron
	ossowski_16555	substitution	unknown	intron
	ossowski_16556	substitution	unknown	intron
	ossowski_16557	substitution	unknown	promoter
	ossowski_16558	substitution	unknown	promoter
	PERL0019826	substitution	unknown	5'UTR
	PERL0019827	substitution	unknown	5'UTR
	PERL0019828	substitution	unknown	coding_region
	PERL0019830	substitution	unknown	coding_region
	PERL0019834	substitution	unknown	coding_region

## Access function information from TAIR locus page

Germplasm Showing 10 of 22 entries <a href="#">Show all</a>	Name CS6224	Polymorphisms phyB-5; phyA-201	Background	Stock Name CS6224	Select <input type="checkbox"/>
<a href="#">Images</a>					
<a href="#">Phenotypes</a>					
	<ul style="list-style-type: none"> <li>double mutant; slightly longer hypocotyl than the phy B single mutant in white light; longer hypocotyl than the phyB single mutant in red light and same length as that of the phyA mutant in far-red light; poorly developed cotyledons in red light unlike the single mutants, reduced red-light induction of CAB gene expression and reduced chlorophyll induction. Phenotype curated by ABRC.</li> </ul>				
<a href="#">Images</a>		<a href="#">phyA-205</a>	<a href="#">CS6222</a>	<input type="checkbox"/>	
<a href="#">Phenotypes</a>					
	<ul style="list-style-type: none"> <li>weak allele; medium hypocotyl and unexpanded cotyledons in far-red light; hypocotyl elongation not affected in white or red light; reduced ability to de-etiolate after growth in the dark; mature plants grown under continuous white light have similar phenotype and flowering time to wild type Phenotype curated by ABRC.</li> </ul>				
				<b>Thanks to ABRC!</b>	
				<a href="#">Order seeds?</a>	

# Access function information from TAIR locus page

## External Link section

**External Link ?**

- [AtGDB View](#)
- [MPSS](#)
- [NASCArrays Digital Northern](#)
- [NASCArrays Spot History](#)
- [Genevestigator Gene Atlas](#)
- [Genevestigator Gene Chronologer](#)
- [Genevestigator Response Viewer](#)
- [e-FP Browser](#)
- [AtGenExpress Visualization Tool](#)
- [MIPS View](#)
- [NCBI-Entrez Gene](#)
- [InParanoid Ortholog Groups](#)
- [Plant Proteome Database](#)
- [T-DNA Express](#)
- [Salk SNP Viewer](#)
- [IntAct \(Protein Interaction Database at EBI\)](#)
- [ATTED-II](#)
- [AtProteome](#)
- [AceView](#)



# View gene model-specific data

**Locus: AT1G09570**

Date last modified 2003-05-02

TAIR Accession Locus:2012300

Representative Gene Model **AT1G09570.1**

Gene Model protein\_coding

Type

Other names: ELONGATED HYPOCOTYL 8, F14J9.23, F14J9\_23, FAR RED ELONGATED 1, FAR RED ELONGATED HYPOCOTYL 2, FHY2, FRE1, HY8, PHYA, PHYTOCHROME A

Description Light-labile cytoplasmic red/far-red light photoreceptor involved in the regulation of photomorphogenesis. It exists in two inter-convertible forms: Pr and Pfr (active) and functions as a dimer. The N terminus carries a single tetrapyrrole chromophore, and the C terminus is involved in dimerization. It is the sole photoreceptor mediating the FR high irradiance response (HIR). Major regulator in red-light induction of phototropic enhancement. Involved in the regulation of de-etiolation. Involved in gravitropism and phototropism. Requires FHY1 for nuclear accumulation.

Other Gene Models **AT1G09570.2** (splice variant)

Map Detail Image

Protein Coding Gene Models

AT1G09570.1

AT1G09570.2

Annotations

Category	Relationship Type	Keyword
GO Biological Process	involved in	gravitropism, phototropism, regulation of transcription, DNA-dependent, response to arsenic, red light signaling pathway, response to very low fluence red light stimulus, response to continuous far red light stimulus by the high-irradiance response system

# View gene model-specific data

**Protein: AT1G09570.1**

Date last modified ?  
TAIR Accession ?  
External IDs ?

Properties

Domains

**Sequence**

[Send to WU-BLAST](#)

```
1 MSGSRPTQSS EGSRRSRHSA RIIAQTTVDA KLHADFEESG SSFDYSTS
 51 VTGPVVENQP PRSDKVTITY LHHIQKGKLI QPFGCLLALD EKTFKVIAYS
 101 ENASELLTMA SHAVPSVGEH PVLIGITDIR SLFTAPSASA LQKALGFCDV
 151 SLLNPILVHC RTSAKPFYAI IHRVTGSIII DFEPVKPYEV PMTAAGALQS
 201 YKLAAKAITR LQSLPSGSME RLCDTMVQEY FELTYDRVM AYKFHEDDHG
 251 EVVSEVTKPG LEPYLGHLHYP ATDIPQAARF LFMKMKVRMI VDCNAKHARV
 301 LQDEKLSFDL TLCGSTLRAP HSCHLQYMAN MDSIASLVMA VVVNEEDGE
 351 DAPDATTQHQ KRKRWLGLVV CHNTTPRFVP FPLRYACEFL AQVFAIHVN
 401 EVELDNQMVE KNILRTQTLL CDMLMRDAPL GIVSQSPNIM DLVKCDGAAL
 451 LYKDKIWKLG TTPSEFHQLQE IASWLCEYHM DSTGLSTDLS HDAGFPRALS
 501 LGDSVCGMAA VRISSKDMIF WFRSHTAGEV RWGGAKHDPD DRDDARRMHP
 551 RSSFKAFLEV VKTRSLPWKD YEMDAIHSLQ LILRMAFKDS ETTDVNTKVI
 601 YSKLNDLKID GIQELEAVTS EMVRLIETAT VPILAVDSDG LVNGWNTKIA
 651 ELTGLSVDEA IGKHFLTVE DSSVEIVKRM LENALEGTEE QNVQFEIKTH
 701 LSRADAGPIs LVVNACASRD LHENVVGVCF VAHDLTGQKT VMDKFTRIEG
 751 DYKAIIQNPN PLIPPIFGTD EFGWCTEWNP AMSKITGLKR EEVIDKMLLG
 801 EVFGTQKSCC RLKNQEAFVN LGIVLNNAVT SQDPEKVSFA FFTRGKGKYVE
 851 CLLCVSKKLD REGVVTGVFC FLQLASHELQ QALHVQRLLAE RTAVKRLKAL
 901 AYIKRQIRNP LSGIMFTRKM IEGTELGPSEQ RRILQTSALC QKQLSKILD
 951 SDLESIIEGC LDLEMKEFTL NEVLTASTSQ VMMKSNKGKV RITNETGEEV
1001 MSDTLYGDSI RLQQVLAQFM LMAVNFTPSC GQLTVSASLR KDQLGRSVHL
1051 ANLEIRLHTT GAGIPEFLNN QMFGTEEDVS EEGLSLMVS R KLVKLMNGDV
1101 QYLRQAGKSS FIITAEAAA NK
```

**Associated Genes**      **Gene Model** ?    **Locus** ?    [seqviewer link](#)  
AT1G09570.1    AT1G09570    Sequence Viewer

**External Link**      [The Subcellular Location of Proteins in Arabidopsis Database \(SUBA\)](#)  
[Arabidopsis Protein Phosphorylation Site Database \(PhosPhAt\)](#)

**User Comments** ?  
(shows only the most recent comments by default)

[Add My Comment](#)    [Hide Comments](#)    [Show All Comments](#)

# Use Search tools to access function information

 Exact name search ▾ Search

Home Help Contact About Us Login/Register

Search Browse Tools Portals Download Submit News ABRC Stocks

Search Overview 0

DNA/Clones

Ecotypes

Genes 0

Gene Ontology Annotations

Plant Ontology Annotations

Keywords

Locus History RE1, PHYA, F14J9.23, F14J9\_23, PHYTOCHROME A, FAR RED ELONGATED 1, ELONGATED 8, FAR RED ELONGATED HYPOCOTYL 2

Markers oplasmic red/far-red light photoreceptor involved in the regulation of photomorphogenesis. It exists in two le forms: Pr and Pfr (active) and functions as a dimer.The N terminus carries a single tetrapyrrole and the C terminus is involved in dimerization. It is the sole photoreceptor mediating the FR high irradiance ). Major regulator in red-light induction of phototropic enhancement. Involved in the regulation of involved in gravitropism and phototropism. Requires FHY1 for nuclear accumulation.

Microarray Element

Microarray Experiment

Microarray Expression

People/Labs

Polymorphisms/Alleles )

Proteins

Protocols

Publication 0.1 (HY8,FHY2,FRE1,PHYA)

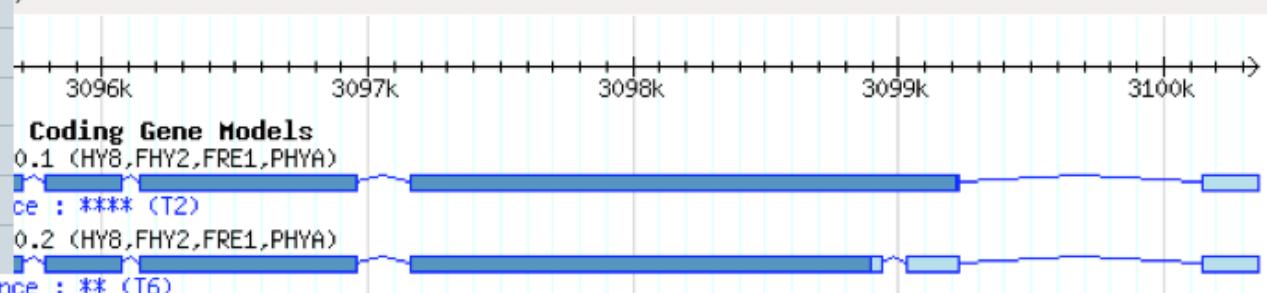
Seed/Germplasm ce : \*\*\*\* (T2)

Textpresso Full Text 0.2 (HY8,FHY2,FRE1,PHYA)

Confidence : \*\* (T6)

Coding Gene Models

Annotations category relationship type keyword



Use TAIR data and tools to drive your research forward

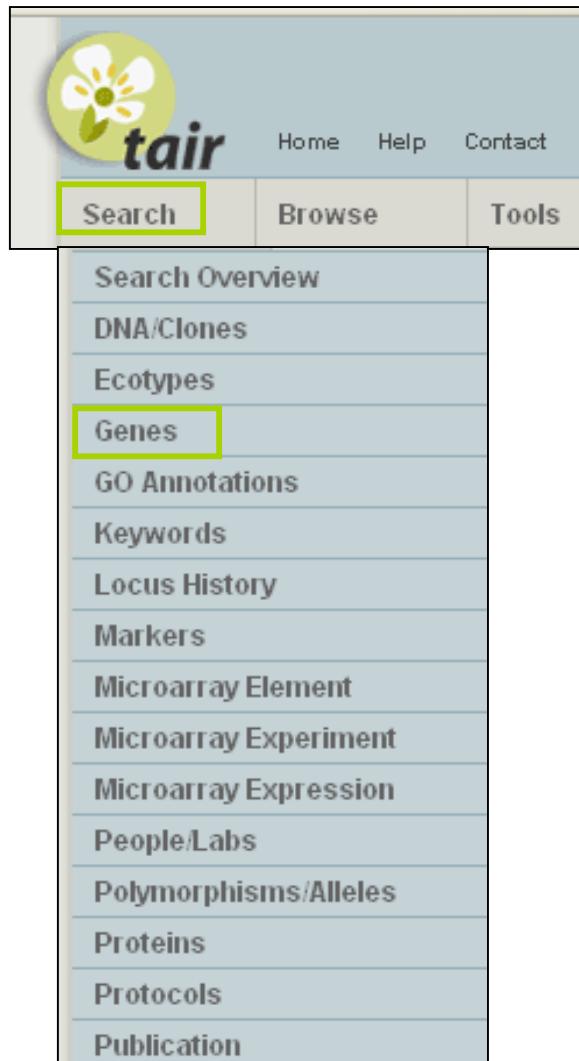
## Example 1

You have mapped a mutation that disrupts flower development to a region of Chromosome 1

Q: What genes in the mapping interval are good candidates?

- Get a list of all the genes in the mapping interval involved in “flower development”

## ■ Get a list of all the genes involved in flower development



The screenshot shows the TAIR (Arabidopsis thaliana Information Resource) website. The top navigation bar includes links for Home, Help, Contact, Search, Browse, and Tools. The left sidebar lists various search categories: Search Overview, DNA/Clones, Ecotypes, Genes (which is highlighted with a yellow box), GO Annotations, Keywords, Locus History, Markers, Microarray Element, Microarray Experiment, Microarray Expression, People/Labs, Polymorphisms/Alleles, Proteins, Protocols, and Publication. The main content area displays a search form for "Search by Name or Phenotype". It features a dropdown menu for "Gene name" and a text input field with a dropdown menu set to "starts with" containing the text "flower development". There is also a checkbox for "Include obsolete genes". Below this is a section for "Search by Associated Keyword" with dropdown menus for "Keyword Term" (set to "contains" with "flower development") and "Keyword Type" (listing GO Molecular Function, GO Biological Process, GO Cellular Component, and Anatomy). The "Evidence" section includes dropdown menus for inferred from expression pattern, inferred from genetic interaction, inferred from mutant phenotype, and inferred from physical interaction. Further down are sections for "Restrict by Features" (listing Gene Model Type like transposable element gene, protein coding, pseudogene, and ribosomal RNA) and "Advanced" filters. A "Time Restriction" section allows searching only within the last 2 months. Finally, there is a "Restrict by Map Locations" section with dropdown menus for Chromosome (set to 1), Map Type (set to AGI), and Range (set to between marker and marker).

Search by Name or Phenotype ?

Gene name starts with flower development  
(leaving the input box blank will retrieve all entries)

Include obsolete genes

Search by Associated Keyword ?

Keyword Term contains flower development

GO/PO ID (exact match only)

Keyword Type

- GO Molecular Function
- GO Biological Process
- GO Cellular Component
- Anatomy

Evidence

- inferred from expression pattern
- inferred from genetic interaction
- inferred from mutant phenotype
- inferred from physical interaction

Restrict by Features ?

Gene Model Type

- transposable element gene
- protein coding
- pseudogene
- ribosomal RNA

Advanced

- gene structure predicted
- has associated literature
- is sequenced
- is not sequenced

Time Restriction ?

only search last 2 months

Restrict by Map Locations ?

Chromosome

Map Type

Range between marker and marker

ment"

## TAIR Gene Search Results

[new search](#)

new gene search

[download all](#)

download all results

[download checked](#)

check the boxes below and download results

Your flowe distin Disp To S8 Ch 1 1	TAIR Accession	Locus	Gene Model	Gene Type	Description	Other Name (Type)	Keywords
	Gene:2007165	AT1G69490	AT1G69490.1	protein_coding	Encodes a member of the NAC transcription fa		
	Gene:2007709	AT1G48270	AT1G48270.1	protein_coding	encodes a protein similar to G-coupled recep		
	Gene:2008864	AT1G11870	AT1G11870.1	protein_coding	Seryl-tRNA synthetase targeted to chloropla		
	Gene:2010365	AT1G12820	AT1G12820.1	protein_coding	AFB3 (AUXIN SIGNALING F-BOX 3); auxin bindin		
	Gene:2011455	AT1G52740	AT1G52740.1	protein_coding	Encodes HTA9, a histone H2A protein. F14G		
	Gene:2012621	AT1G14400	AT1G14400.1	protein_coding	ubiquitin carrier protein ATUBC1; F14L		
	Gene:2015775	AT1G30970	AT1G30970.1	protein_coding	Encodes SUF4 (SUPPRESSOR of FRI 4), a putati		
	Gene:2017661	AT1G51660	AT1G51660.1	protein_coding	Encodes a mitogen-activated map kinase kinas		
	Gene:2017763	AT1G22920	AT1G22920.1	protein_coding	AJH1 encodes a protein similar to JAB1, a sp		
	Gene:2018253	AT1G04400	AT1G04400.1	protein_coding	Blue light receptor mediating blue-light reg		
	Gene:2019503	AT1G64520	AT1G64520.1	protein_coding	RPN12A (REGULATORY PARTICLE NON-ATPASE 12A);		
	Gene:2019942	AT1G10670	AT1G10670.1	protein_coding	One of the three genes encoding subunit A of		
	Gene:2020411	AT1G17110	AT1G17110.1	protein_coding	Encodes a ubiquitin-specific protease, and i		
	Gene:2024597	AT1G63990	AT1G63990.1	protein_coding	Encodes AtSPO11-2, one of the three Arabidop		
	Gene:2026809	AT1G70510	AT1G70510.1	protein_coding	A member of class I knotted1-like homeobox g		
	Gene:2027088	AT1G62360	AT1G62360.1	protein_coding	Class I knotted-like homeodomain protein the		

[Arabidopsis thaliana]

(TAIR:AT2G46870.....)

## Gene search

Function: identify a list of genes (that have certain function) (located in a specific region)

# Use TAIR data and tools to drive your research forward

## Example 2

You work on a transcription factor, you have a mutant of this TF gene, you have obtained a list of genes whose expression is altered in this mutant

Q: What's the cis-element for this TF?

- Get upstream “promoter” sequences for these genes
- Search for over-represented DNA sequences in “promoters”

# Download:Bulk Data Retrieval:Sequences

Search	Browse	Tools	Stocks	Portals	Download	Submit	News
<b>The Arabidopsis Information Resource</b>  The Arabidopsis Information Resource (TAIR) maintains a <a href="#">database</a> of genetic and <a href="#">biology data</a> for the model higher plant <i>Arabidopsis thaliana</i> . Data available from the complete genome sequence along with gene structure, gene product information, metabolism, gene expression, DNA and seed stocks, genome maps, genetic and molecular markers, publications, and information about the Arabidopsis research community. Product function data is updated every two weeks from the latest published research and community data submissions. Gene structures are updated 1-2 times per year using computational and manual methods as well as community submissions of new genes. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.  The <a href="#">Arabidopsis Biological Resource Center</a> at The Ohio State University collects, preserves and distributes seed and DNA resources of <i>Arabidopsis thaliana</i> and related species. Stock information and ordering for the ABRC are fully integrated into TAIR.					<a href="#">Download Overview</a> <a href="#">Genes</a> <a href="#">GO and PO Annotations</a> <a href="#">Maps</a> <a href="#">Pathways</a> <a href="#">Proteins</a> <a href="#">Protocols</a> <a href="#">Microarray Data</a> <a href="#">Sequences</a> <a href="#">Software</a> <a href="#">User Requests</a> <a href="#">Bulk Data Retrieval</a>  Served at ABRC TM from J. Ecker / Salk		

# Download:Bulk Data Retrieval:Sequences

Sequences		AGI transcripts	AGI coding sequences	AGI genomic locus sequences	AGI protein sequences	Upstream Sequences - 500bp	Upstream Sequences - 1000bp
Locus Identifiers or Sequences:	AT4G35790 AT4G11850						
>AT1G31480   chr1:11264861-11265860 FORWARD CCGTCTCGAAGGAGGCTCGGTTAGGCGGAGTTATCTGAAACATCTCTGATACGAGGTCTCATATACAAAGGAAGATGATTGATGTTGACCGATGCTTGC GCAAAAAAGCGTTTAGTTGTTCATCTACCGCTTTGATCTTGTTCAATAAAAGGGATTGGGGTTTGATGTTAATTGTTTATTAAACAG TGTCTGTTAAATGTTTAGGTTAGGTTAGTTGGTATTGATTAGGCAACACTATTGTTGCTTGAAAGTGTAGAAAACGTAACGTCAA AATAAACGTAATGAAAACAAAGAAGTGGTGAGAATTAAAGGAATCTGCACCTAGACTAAGAATCAAGGACCAGTCATATGTTAGACCAA CGAATTCTTATTGGCAAAGCATTAAATTATTATAGTGGTTAAGATTGGTATGGTTAGGAGTAGCACAATGATGAACTCTAAATTCTAA CCATCCAAAAAGAAGTATTGAAAGGTTCAAAGCTACGATTCTTGGTAAGAATCTACGAAGGTGACCAACAAGTAAGTAACATAGAGCCTCATTAGA TCCCAAATTGTAATCCTTGGCTTTGCTATACCAATCGCAAGTATCTTCTAAAAACTACTATCGGTGGAGTATTAGCTAAATGGAGCTGTT AATTATCATTTCATAATTCAAAACACACATGCTTGCTATCTTAAATAACACACTATTGAAACAATTGGTAGATCACACCAACACATT TGACTTTGGCAAATTATCCTTCACTTGTAAACAACAAGACAAATTGGTCTGTTTTCGAAATTATGACTTGGCATCTACGTGGCAGTGGTTAAGTA GATGACAGTACACGCAACCGCTGAAGCAAAGTTCTGGTTCTCTCTCAGAATTCTCAATACCAACAAACCAATTCAATTGATTCCCCA >AT1G67560   chr1:25322562-25323561 FORWARD AAAAGCTTAGATATTCAACGCTTGTAAAGAATCGATTCTGACTTGCAACTTTTACAAGTAGGAACGACTAATTACTACCAACTTGGAGTAACACG TACCGATAAGATTGAAAAGTTACATTAGAGATTGAGGATTATGTACCGTTCAAAGAAATGCCAAGTTAATAATTGTTATTCTGTTAGGTAGG AATACAAACGAGCATGTGCATAATTAAATATAACCTAAACTAACGGCGGTCAATTGCTCCAATGAGTATATGCATGTTATGTTAGGTAGTT TTAATGTCCAACGGTTTATTGACATCAACATGCACTTTTATAGTACAATTAAAGTTGGTGCAGGCGCAGCAAGCAAGGCCACACCGTACATT ACTGCGTACAAACAAGAAGCGGTCAATTTCGAGAAAGACGATCCATACGTTCTGTGTAATAAATAACATACATTGTTGGAAA ACAATTCTTCTATATGAAAACACCGTAGAATGTAAGCCAGAGGAGATTACCAAGGTGGCTCTGAGATATTGAAAGATTGTTCACTTGTGCCTT CTAAAACGTTAGTGTATACGTTACCTCGCTAACACATGTTCATATATAACACACAAAAAGTTATTGTTAGGAGTCTGATTCTCACCGCCTTACACCGT ACGTGACAAGTAAATTCAACACCAACCGTACGGGTGGAAGTGACGACACTCTCACGGTCAAACACATTACCTGTCCGAAAAGTGAACACAA AACGTCACCGTTTCCTGACTCAGTCCTCTAGCACGGTCCACGAGAACGTTCTCGTTGCCTTATTATCATTCTCTATCTCTCTTCAC							

# Tools:Motif Analysis

- Search for over-represented or prevalent DNA sequences in “promoters”
  - Use the Motif Analyzer in TAIR to identify common **6-mer**

The screenshot shows the TAIR Motif Analyzer results page. The left sidebar has a logo and links for 'Search' and 'B'. The main content area has a heading 'Statistical Mo' and a descriptive text about the program. A table lists 6-mers along with their absolute counts, genomic counts, p-values, and associated genes. An arrow points to the entry for AT1G69490.

oligoMer	Absolute number of this oligoMer in query set	Absolute number in genomic set	Number of sequences in query set containing oligoMer	Number of sequences (out of 33282 in genomic set) containing oligoMer	p-value from binomial distribution	Query sequences containing this oligoMer
CATGCC	20	5574	17/52	4995/33282	7.36e-04	AT1G48270 AT1G12820 AT1G52740 AT1G14400 AT1G04400 AT1G17110 AT1G63990 AT1G45050 AT1G26260 AT1G43850 AT1G25350 AT1G65480 AT1G09000 AT1G01030 AT1G61040 AT1G79000 AT1G69180
GGCATG	20	5574	17/52	4995/33282	7.36e-04	AT1G48270 AT1G12820 AT1G52740 AT1G14400 AT1G04400 AT1G17110 AT1G63990 AT1G45050 AT1G26260 AT1G43850 AT1G25350 AT1G65480 AT1G09000 AT1G01030 AT1G61040 AT1G79000 AT1G69180
GATCAA	23	27657	16/52	17338/33282	9.51e-04	AT1G69490 AT1G51660 AT1G22920 AT1G10670 AT1G17110 AT1G63990 AT1G70510 AT1G26260 AT1G43850 AT1G47260 AT1G59640 AT1G79000 AT1G16710 AT1G26830 AT1G13400 AT1G69180
TTGATC	23	27657	16/52	17338/33282	9.51e-04	AT1G69490 AT1G51660 AT1G22920 AT1G10670 AT1G17110 AT1G63990 AT1G70510 AT1G26260 AT1G43850 AT1G47260 AT1G59640 AT1G79000 AT1G16710 AT1G26830 AT1G13400 AT1G69180

# summary

TOOL	INPUT	OUTPUT
bulk data retrieval:sequence	list of AGI codes	sequence
motif analyzer	list of AGI codes	DNA motif in promoter

# GO term enrichment

 Gene [Home](#) [Help](#) [Contact](#) [About Us](#) [Login/Register](#)[Search](#) [Browse](#) [Tools](#) [Portals](#) [Download](#) [Submit](#) [News](#) [ABRC Stocks](#)[Home](#) > [Tools](#) > [Bulk Data Retrieval](#) > [GO Annotations](#)

## GO annotation search, functional categorization and download

[\[ Help \]](#)

Paste locus identifiers (such as At1g01030) into the textbox and press one of the submit buttons below. The identifiers have to be separated by tabs, commas, carriage returns or spaces. Alternatively, you can upload a file, same formatting as for the textbox. Clicking on Get all GO annotations will display in detail all the GO annotations done to your set of genes. Clicking on Functional categorization will group the genes into broad functional categories based on the high level terms in GO hierarchy.

You may download the whole genome GO annotations from [TAIR FTP site](#).

Do you want to look for over-represented terms in your data set?

Try [GO Term Enrichment](#) : A GO tool for statistical comparisons of non-electronic annotations between two data sets

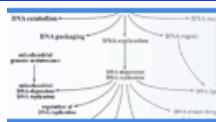
### Whole Genome Categorization

Get:

### GO Annotations

Locus Identifiers:

# GO term enrichment



*the Gene Ontology*

AmiGO

Search Browse BLAST Homolog Annotations Tools & Resources Help

Search GO   terms  genes or proteins  exact match

## GO Term Enrichment

### View Enriched Terms

The Term Enrichment tool finds significant shared GO terms or parents of those GO terms, used to describe the genes in the query/input set to help discover what those genes may have in common. The Term Enrichment tool makes use of the [GO-TermFinder](#) perl module written by Gavin Sherlock and Shuai Weng at Stanford University.

More information about this tool can be found in the [AmiGO manual](#).

**Caution:** Please note that by default, this tool uses annotation datasets that include [IEA](#) (electronically inferred) data. If you wish to work with only non-IEA data, please upload your own background set and select the "no" option [below](#).

### Input your gene products \* [required]

Enter your gene products below (symbols, synonyms, or accessions):

**OR** upload a [list containing gene products](#) :

# Download whole-genome annotation files from TAIR ftp site

<b>Download</b>	<b>Submit</b>
Download Overview	
ABRC Documents	
Genes	
<b>GO and PO Annotations</b>	
Maps	
Metabolic Pathways	
Polymorphisms	
Proteins	
Protocols	
Microarray Data	
Sequences	
Software	
User Requests	
Bulk Data Retrieval	

## Name

-  [Gene\\_Ontology](#)
-  [Plant\\_Ontology](#)

# Download whole-genome annotation files from TAIR ftp site

<b>Download</b>	<b>Submit</b>
Download Overview	
ABRC Documents	
<b>Genes</b>	
GO and PO Annotations	
Maps	
Metabolic Pathways	
Polymorphisms	
Proteins	
Protocols	
Microarray Data	
Sequences	
Software	
User Requests	
Bulk Data Retrieval	

Name
<a href="#">2010_projects</a>
<a href="#">Gene_Confidence.pdf</a>
<a href="#">Gene_families</a>
<a href="#">Locus_history</a>
<a href="#">OLD</a>
<a href="#">ORF_annotations</a>
<a href="#">SAGE</a>
<a href="#">SmallRNAs</a>
<a href="#">TAIR10_genome_release</a>
<a href="#">TAIR6_genome_release</a>
<a href="#">TAIR7_genome_release</a>
<a href="#">TAIR8_genome_release</a>
<a href="#">TAIR9_genome_release</a>
<a href="#">TIGR5_genome_release</a>
<a href="#">TIGR_release_versions</a>
<a href="#">est_mapping</a>
<a href="#">gene_aliases.20101027</a>

genome  
releases