

Biological interactions beyond the lab bench:

TAIR, PMN, piles of data, diverse plants,
and you!



Kate Dreher

TAIR, PMN
Carnegie Institution for Science

Piles of biological data about diverse plants . . .

- Plant nucleotide sequences in GenBank: **25,170,404 +**
- Plant genome sequencing projects: **~310**
- Articles published in plant journals (2007): **2400 +**

Piles of biological data about diverse plants . . .

- ❑ Plant nucleotide sequences in GenBank: **25,170,404 +**
- ❑ Plant genome sequencing projects: **~310**
- ❑ Articles published in plant journals (2007): **2400 +**
- ❑ **How can this information be organized and utilized?**

Biological databases try to help . . .



- TAIR – The Arabidopsis Information Resource
 - Comprehensive database providing access to genomic, proteomic, and metabolomic resources
 - **www.arabidopsis.org**

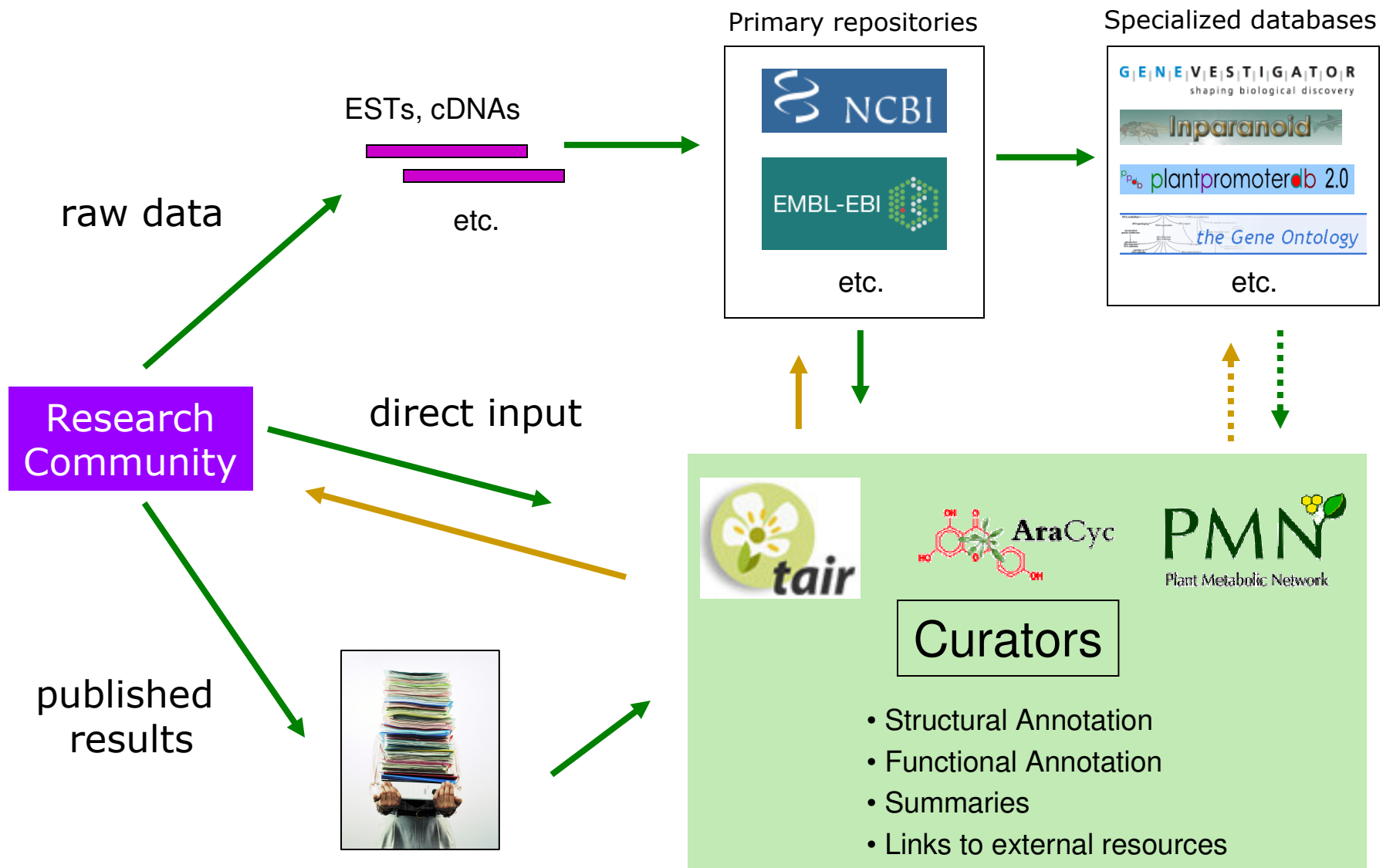


- AraCyc – Arabidopsis Metabolic EnCyclopedia
 - Database of metabolic pathways found in Arabidopsis
 - **www.arabidopsis.org/biocyc/**



- PMN – Plant Metabolic Network (**Coming soon!!**)
 - Collection of databases describing metabolic pathways in numerous plant species

Databases promote “biological interactions”



Databases promote “biological interactions”

- ❑ Connecting scientists to data:
 - Genomes
 - Transcripts
 - Proteins
 - Metabolism
 - Other organisms
- ❑ Connecting scientists to resources and tools:
 - DNA stocks / germplasm
 - Analysis and viewing applications, etc.
- ❑ Connecting scientists to each other
- ❑ Connecting scientists to everything at once . . .



Connecting scientists to everything . . .

TAIR logo

Home Help Contact About Us Login/Register

Gene Search

Search Browse Tools Stocks Portals Download Submit News

Search Overview
DNA/Clones
Ecotypes
Genes
GO Annotations
Keywords
Locus History
Markers
Microarray Element
Microarray Experiment
Microarray Expression
People/Labs
Polymorphisms/Alleles
Proteins
Protocols
Publication
Seed/Germplasm
Sequences

Tools Overview
GBrowse
Seqviewer
Mapviewer
AraCyc Metabolic Pathways
BLAST
WU-BLAST
FASTA
Patmatch
Motif Analysis
VxInsight
Java Tree View
Bulk Data Retrieval
Chromosome Map Tool
Restriction Analysis
Gene Symbol Registry

Portals Overview
Clones/DNA Resources
Education and Outreach
Gene Expression Resources
Genome Annotation
MASC/Functional Genomics
Mutant and Mapping Resources
Nomenclature
Proteomics Resources
Metabolomics Resources

Bulk Data Retrieval

News

Clones Received at ABRC

Gateway™

CDNWORK clones from J. Ecker (Salk)

b) 15 vectors from various donors

c) 2,425 transcription factor expression clones from S. P. Dinesh Kumar

d) 82 cloned genes and constructs



Connecting scientists to everything at once . . . to solve important biological problems

- ❑ Food prices are rising, hunger is increasing . . .
- ❑ We need more stress-tolerant crops!
 - You have an Arabidopsis mutant with reduced cold tolerance
 - You've mapped it to an interval on chromosome 4
 - You want to study the basic biology of cold tolerance
 - You want to help develop more cold-tolerant rice





Finding the mutation

Finding the mutation

■ Narrowing the mapping interval

- Search for additional markers in the region using the [Marker search](#) page

The screenshot shows a web interface for searching markers. It is divided into three main sections: 'Search by Name', 'Restrict by Features', and 'Restrict by Map Location'. The 'Search by Name' section has a 'Marker Name' dropdown and a 'starts with' input field. The 'Restrict by Features' section includes a list of genetic markers, a 'Polymorphic Between' section with ecotype dropdowns, and a 'Time Restriction' checkbox. The 'Restrict by Map Location' section includes a 'Chromosome' dropdown, a 'Map Type' dropdown, and a 'Range' section with 'between' dropdowns and marker name inputs. Three green boxes highlight specific search criteria: the 'All PCR' checkbox, the 'Polymorphic Between' section, and the 'Range' section.

Search by Name

Marker Name starts with
(leaving the input box blank will retrieve all entries)

Restrict by Features

Genetic Marker ☐ Any ☒ All PCR ☐ CAPS ☐ SSLP ☐ AFLP ☐ RFLP ☐ RAPD ☐ Hybridization Based

Polymorphic Between Ecotype Ler and Ecotype Ws

Time Restriction ☐ only search last 2 months

Restrict by Map Location

Chromosome 4 Map Type AGI Range between SM103_365,6 marker and SM58_108,7 marker

Finding the mutation

■ Search for additional markers in the region

- Use Marker search page

TAIR Genetic Marker Search Results
[new search](#) [download](#)
new genetic marker search check the boxes below and download results

Your query for genetic resulted in 1 match.

Displaying 1 - 1.

[Check All](#) [Uncheck All](#)

Marker Name

1 ☐ **F4D11S**

Genetic Marker:F4D11S

Date last modified	2001-03-23																						
Name	F4D11S																						
Tair Accession	GeneticMarker:2005326																						
Type	CAPS																						
Length	1.3																						
Is PCR Marker	true																						
Special Conditions	none																						
Chromosome	4																						
Associated Polymorphisms																							
Digest Patterns																							
	<table><thead><tr><th>Polymorphism</th><th>species variant</th><th>restrict enzyme</th><th>num sites</th><th>fragment length</th><th>attribution</th></tr></thead><tbody><tr><td>F4D11S</td><td>LANDSBERG ERECTA</td><td>SSTI</td><td>1</td><td>1.000; 0.300 kb</td><td></td></tr><tr><td>F4D11S</td><td>WASSILEWSKJA</td><td>SSTI</td><td>2</td><td>0.700; 0.300; 0.300 kb</td><td></td></tr></tbody></table>	Polymorphism	species variant	restrict enzyme	num sites	fragment length	attribution	F4D11S	LANDSBERG ERECTA	SSTI	1	1.000; 0.300 kb		F4D11S	WASSILEWSKJA	SSTI	2	0.700; 0.300; 0.300 kb					
Polymorphism	species variant	restrict enzyme	num sites	fragment length	attribution																		
F4D11S	LANDSBERG ERECTA	SSTI	1	1.000; 0.300 kb																			
F4D11S	WASSILEWSKJA	SSTI	2	0.700; 0.300; 0.300 kb																			
Map Locations																							
	<table><thead><tr><th>chrom</th><th>map</th><th>map type ?</th><th>coordinates</th><th>orientation</th><th>attrib</th></tr></thead><tbody><tr><td>4</td><td>AGI</td><td>nuc_sequence</td><td>15780107 - 15781422 bp</td><td>unknown</td><td>details</td></tr><tr><td>4</td><td>F4D11</td><td>assembly_unit</td><td>64757 - 66072 bp</td><td>unknown</td><td></td></tr></tbody></table>	chrom	map	map type ?	coordinates	orientation	attrib	4	AGI	nuc_sequence	15780107 - 15781422 bp	unknown	details	4	F4D11	assembly_unit	64757 - 66072 bp	unknown					
chrom	map	map type ?	coordinates	orientation	attrib																		
4	AGI	nuc_sequence	15780107 - 15781422 bp	unknown	details																		
4	F4D11	assembly_unit	64757 - 66072 bp	unknown																			
Map Links	Map Viewer	Sequence Viewer	GBrowse																				
Associated Nucleotide Sequences																							
Flanking Sequences	primer 1: TCCTTCGTATCACAACTCC																						
	primer 2: GCGATAACACATTCTCAGCG																						

Finding the mutation

■ Search for additional markers in the region

- Use Marker search page

TAIR Genetic Marker Search Results
[new search](#) [download](#)
new genetic marker search check the boxes below and download results

Your query for genetic resulted in 1 match.

Displaying 1 - 1.

[Check All](#) [Uncheck All](#)

Marker Name

1 ☐ **F4D11S**

Genetic Marker:F4D11S

Date last modified	2001-03-23																						
Name	F4D11S																						
Tair Accession	GeneticMarker:2005326																						
Type	CAPS																						
Length	1.3																						
Is PCR Marker	true																						
Special Conditions	none																						
Chromosome	4																						
Associated Polymorphisms																							
Digest Patterns																							
	<table><thead><tr><th>Polymorphism</th><th>species variant</th><th>restrict enzyme</th><th>num sites</th><th>fragment length</th><th>attribution</th></tr></thead><tbody><tr><td>F4D11S</td><td>LANDSBERG ERECTA</td><td>SSTI</td><td>1</td><td>1.000; 0.300 kb</td><td></td></tr><tr><td>F4D11S</td><td>WASSILEWSKIIA</td><td>SSTI</td><td>2</td><td>0.700; 0.300; 0.300 kb</td><td></td></tr></tbody></table>	Polymorphism	species variant	restrict enzyme	num sites	fragment length	attribution	F4D11S	LANDSBERG ERECTA	SSTI	1	1.000; 0.300 kb		F4D11S	WASSILEWSKIIA	SSTI	2	0.700; 0.300; 0.300 kb					
Polymorphism	species variant	restrict enzyme	num sites	fragment length	attribution																		
F4D11S	LANDSBERG ERECTA	SSTI	1	1.000; 0.300 kb																			
F4D11S	WASSILEWSKIIA	SSTI	2	0.700; 0.300; 0.300 kb																			
Map Locations																							
	<table><thead><tr><th>chrom</th><th>map</th><th>map type ?</th><th>coordinates</th><th>orientation</th><th>attrib</th></tr></thead><tbody><tr><td>4</td><td>AGI</td><td>nuc_sequence</td><td>15780107 - 15781422 bp</td><td>unknown</td><td>details</td></tr><tr><td>4</td><td>F4D11</td><td>assembly_unit</td><td>64757 - 66072 bp</td><td>unknown</td><td></td></tr></tbody></table>	chrom	map	map type ?	coordinates	orientation	attrib	4	AGI	nuc_sequence	15780107 - 15781422 bp	unknown	details	4	F4D11	assembly_unit	64757 - 66072 bp	unknown					
chrom	map	map type ?	coordinates	orientation	attrib																		
4	AGI	nuc_sequence	15780107 - 15781422 bp	unknown	details																		
4	F4D11	assembly_unit	64757 - 66072 bp	unknown																			
Map Links	Map Viewer	Sequence Viewer	GBrowse																				
Associated Nucleotide Sequences																							
Flanking Sequences	primer 1: TCCTTCGTATCACAACTCC																						
	primer 2: GCGATAACACATTCTCAGCG																						

Finding the mutation

- **Narrow the mapping interval**
 - Find BACs for complementation using Seqviewer



Finding the mutation (Seqviewer)

The screenshot shows the TAIR SeqViewer Whole Genome View interface. At the top, there is a navigation bar with links: Home, Help, Contact, About Us, Login/Register. A search bar is present with a dropdown menu set to 'Gene' and a 'Search' button. Below the navigation bar is a menu with tabs: Search, Browse, Tools (highlighted with a green box), Stocks, Portals, Download, Submit, and News. The main content area is titled 'TAIR SeqViewer Whole Genome View' with a 'HELP' link. It displays '2 hits, click [here](#) for a list.' Below this, five horizontal green bars represent chromosomes, numbered 1 to 5. Various gene names are labeled along these chromosomes: PHYA, SRP54A, M235, UFO, MS_1_1, ARR3, SBG9, ADH on chromosome 1; RGA_RNS1, MI421, PHYB, ER, COP1, LTP, MI79A on chromosome 2; CA1, MNSOD, DWF1, ABI3, GL1, NIT1.2, ASN1 on chromosome 3; M51, GA1.1, HY4, MI465, AG, RPS2, MI431, AP2 on chromosome 4; TFL1, TT4, NGA139, PHYC, LFY3, M555 on chromosome 5. A blue dot is visible on chromosome 1, and a red double bar is on chromosome 4. Below the chromosomes, a text prompt says 'Click on green chromosome to open a closeup view'. On the left side, there are settings for 'Zoom level:' (set to 200 Kb) and 'Display:' with checkboxes for Markers, Polymorphisms, T-DNA/Tn, Gene Models, Transcripts, and Annotation Units, all of which are checked. On the right side, there is a search section with 'Search by' radio buttons for 'name' (selected) and 'sequence (15-150 nt)'. A dropdown menu next to 'name' is set to 'marker' and is highlighted with a green box. Below this, a text input field contains 'SM103_365,6' and 'SM58_108,7', also highlighted with a green box. There is a 'Paste in name(s) (up to 250) or sequence(s) (up to 4):' label with a 'help' link. Below the input field, there is a text prompt 'or upload a file of names or sequences: help' and a 'Browse...' button. At the bottom of the search section are 'Submit' and 'Reset' buttons. At the very bottom of the page, a footer states 'Version: TAIR 8.0 genome sequence, released April 2008'.

TAIR SeqViewer Whole Genome View [HELP](#)

2 hits, click [here](#) for a list.

PHYA SRP54A M235 UFO MS_1_1 ARR3 SBG9 ADH 1

RGA_RNS1 MI421 PHYB ER COP1 LTP MI79A 2

CA1 MNSOD DWF1 ABI3 GL1 NIT1.2 ASN1 3

M51 GA1.1 HY4 MI465 AG RPS2 MI431 AP2 4

TFL1 TT4 NGA139 PHYC LFY3 M555 5

Click on green chromosome to open a closeup view

Zoom level: 200 Kb

Display:

- ☒ Markers
- ☒ Polymorphisms
- ☒ T-DNA/Tn
- ☒ Gene Models
- ☒ Transcripts
- ☒ Annotation Units

Search by ☒ name **marker** or ☐ sequence (15-150 nt)

Paste in name(s) (up to 250) or sequence(s) (up to 4): [help](#)

SM103_365,6
SM58_108,7

or upload a file of names or sequences: [help](#)

Browse...

Submit Reset

Version: TAIR 8.0 genome sequence, released April 2008

Finding the mutation (Seqviewer)

some 4

SM103_365,6

PERL0841615 PERL0841616

PERL0841626 PERL0841627

PERL0841628 PERL0841629

SALK_065203.53.75.X

SALK_116728.46.95.N

SALK_052122.33.20.X

AT4G35730.1 AT4G35731.1

AT4G35733.1 AT4G35734.1

AT4G35740.1 AT4G35741.1

EB3R0DY01EB8LL EB3R0DY01EB8LL

EB3R0DY02JRDIS EB3R0DY02JRDIS

FB102G05F (3') 5996

F8D20

F4B14

T19K4

Marker

Poly

T-DNA/Tn

Gene

Transcript

AnnotUnit

Annotation Unit: F4B14

Date last modified

2008-04-23

Name

F4B14

Clone

vector type insert type clone name

BAC genomic F4B14

TAIR Accession

AssemblyUnit:3601514

Chromosome

4

Neighboring Annotation Unit

Stock Information ?

Select

Stock # ?

Availability ?

Donor

Order from ABRC

Reset

Map Locations

chrom

map

map type

coordinates

orientation

attrib

4

AGI

nuc_sequence

16936442 - 17035902 bp

forward

details

1

F28P22

assembly_unit

19845 - 20338 bp

forward

Map Links

[Map Viewer](#)
[Sequence Viewer](#)
[GBrowse](#)

Other Information

Sequence_status

Length

Orientation

annotated GenBank release

99461.0 bp

F

Sequence

Bio Source

Source

Date

GenBank Accession

Sequence

genomic

AGI-TIGR

2001-04-10

genomic

Gene Features

gene name

description

AT4G35783.1

DVL17/RTFL6 (ROTUNDIFOLIA LIKE 6); similar to DVL18/RTFL5 (ROTUNDIFOLIA LIKE 5) [Arabidopsis thaliana] (TAIR:AT5G59510.1); contains InterPro domain DVL (InterPro:IPR012552)

AT4G35860.2

GTP-binding protein ATGB2

AT4G35837.1

unknown protein

Obtaining DNA and seed stocks

- ❑ Register with TAIR as an individual
- ❑ Affiliate with a lab
- ❑ Before ordering . . .
 - Wait for a curator to manually validate your registration



Membership has its privileges . . .

- ❑ Connect to the community via the People / Labs search page
 - It's not quite Facebook, but, all the cool plant biologists are there . . .

Search	Browse	Tools	Stocks	Portals	Download	Submit	News
Search Overview							
DNA/Clones							
Ecotypes							
Genes							
GO Annotations							
Keywords							
Locus History							
Markers							
Microarray Element							
Microarray Experiment							
Microarray Expression							
People/Labs							
Polymorphisms/Alleles							
Proteins							
Protocols							
Publication							
Seed/Germplasm							
Sequences							

Information Resource

Resource (TAIR) maintains a database of genetic and molecular information for the model plant *Arabidopsis thaliana*. Data available from TAIR includes gene structure, gene product information, DNA and seed stocks, genome maps, genetic and physical maps, and information about the Arabidopsis research community. Gene structures are updated 1-2 times per year using the latest published research literature. Gene structures are updated 1-2 times per year using methods as well as community submissions of new and updated gene structures. TAIR provides extensive linkouts from our data pages to other Arabidopsis resources.

Resource Center at The Ohio State University collects, reproduces, and distributes DNA resources of *Arabidopsis thaliana* and related species. Ordering for the ABRC are fully integrated into TAIR.

at the Carnegie Institution for Science Department of Plant Biology and funded by the National Science Foundation.

Breaking News

TAIR8 Release

The Arabidopsis Information Resource is pleased to announce the release of the latest version of the Arabidopsis genome annotation (TAIR8) at TAIR and NCBI.

TAIR Survey Now Open!

Please click on the survey button at the top of the page to give us your comments and feedback on how well TAIR is serving your needs and what additional tools and data you would like to see. The survey will be open for only a short time, please act now to be sure your input is included.



Finding the mutation

- ▣ **Looking for candidate genes**
 - Identify genes in the region using Seqviewer



Finding the mutation (Seqviewer)

Closeup View #1 , Chromosome 4
Remove

show legend

Zoom to: 200 Kb
Find

Search by name (e.g. UFO)
Go

Select range (e.g. 1500-2500)
List Genes In Range

Click here to recenter the view

SM103_365,6
SM56

Wed May 07 21:08:24 PDT 2008

Chromosome 4: 16932578 to 17132577 nt.
Includes annotation units F8D20 F4B14 T19K4 F23E13
Total loci in this region: 76

Download as text file

	A	B	C	D	E	F	G
1	TAIR SeqViewer	Wed May 07 21:08:32 PDT 2008					
2	Chromosome 4: 16932578 to 17132577 nt.						
3	Includes annotation units F8D20 F4B14 T19K4 F23E13						
4	Total loci in this region: 76						
5							
6	Location	Locus	Gene Model: Description				
7	16931044-16933361	AT4G35730.1	AT4G35730.1:similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G34220.2)				
8	16934241-16935507	AT4G35733.1	AT4G35733.1:similar to F-box protein-related / C-type lectin-related [Arabidopsis tha				
9	16936125-16940337	AT4G35740.1	RECQL3:None; AT4G35740.1:RecQL3 (Recq-like 3); ATP binding / ATP-dependent h				
10	16936125-16940337	AT4G35740.2	RECQL3:None; AT4G35740.2:RecQL3 (Recq-like 3); ATP binding / ATP-dependent h				
11	16940512-16941867	AT4G35750.1	AT4G35750.1:Rho-GTPase-activating protein-related; similar to unknown protein [Ar				
12	16942326-16944682	AT4G35760.1	AT4G35760.1:electron carrier/ protein disulfide oxidoreductase; similar to unnamed p				
13	16944946-16946197	AT4G35770.1	SEN1:Senescence-associated gene that is strongly induced by phosphate starvation				
14	16944946-16946197	AT4G35770.2	SEN1:Senescence-associated gene that is strongly induced by phosphate starvation				
15	16944946-16946197	AT4G35770.3	SEN1:Senescence-associated gene that is strongly induced by phosphate starvation				
16	16946526-16950599	AT4G35780.1	AT4G35780.1:protein kinase family protein; similar to protein kinase family protein [A				
17	16952411-16952805	AT4G35783.1	AT4G35783.1:DVL17/RTFL6 (ROTUNDIFOLIA LIKE 6); similar to DVL18/RTFL5 (RO				
18	16952968-16955212	AT4G35785.1	AT4G35785.1:transformer serine/arginine-rich ribonucleoprotein, putative; similar to t				

Finding the mutation

- **Search for cold-tolerance-related genes in the interval**
 - Use [Gene search page](#) and limit by features



Finding the mutation (Gene search page)

Search by Associated Keyword ?

Keyword Term ?

contains cold

Keyword Type

Any
GO Molecular Function
GO Biological Process
GO Cellular Component

Evidence ?

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

Restrict by Features ?

Gene Model Type ?

Any
pre tna
transposable element gene
protein coding

Advanced

gene structure predicted
has associated literature
is sequenced
is not sequenced

Time Restriction ?

☐ only search last 2 months

Restrict by Map Locations ?

Chromosome ?

4

Map Type ?

AGI

Range ?

between v103_365, marker and v58_108,7 marker

Finding the mutation (Gene search page)

Locus ?	Description ?	Gene Model(s) ?	Other Names ?	Keywords ?
1 <input type="checkbox"/>	AT4G35790 Encodes a protein with phospholipase D activity. Involved in phospholipase metabolism. Mutants are affected in hydrogen peroxi....	AT4G35790.1	F4B14.60 F4B14_60 ARABIDOPSIS THALIANA PHOSPHOLIPASE D DELTA	plasma membrane, phospholipase D activity, response to cold
		ATPLDDELTA	PLDDELTA PHOSPHOLIPASE D	microtubule cytoskeleton, plasma membrane, phospholipase D activity, response to cold, programmed cell death, phosphatidic acid metabolic process, lateral root elongation zone, primary root elongation zone, cotyledon vascular system, leaf vascular system
		AT4G35790.3		phospholipase D activity
		AT4G35790.2		phospholipase D activity
2 <input type="checkbox"/>	AT4G36020 Encodes a cold shock domain protein. Involved in cold acclimation by blocking the secondary structure of mRNA which in turn f....	AT4G36020.1	T19K4.150 T19K4_150 COLD SHOCK DOMAIN PROTEIN 1 CSDP1	cellular component unknown, RNA binding, double-stranded DNA binding, nucleic acid binding, single-stranded DNA binding, response to cold, DNA duplex unwinding, RNA secondary structure unwinding



Getting acquainted with the candidates

Getting acquainted with the candidates

- Scan information on the Locus page
 - Read description

Locus: AT4G35790	
Date last modified	2003-05-02
TAIR Accession	Locus:2125314
Representative Gene Model ?	AT4G35790.1
Other names:	ARABIDOPSIS THALIANA PHOSPHOLIPASE D DELTA, ATPLDDELTA, F4B14.60, F4B14_60, PLDDELTA
Description ?	Encodes a protein with phospholipase D activity. Involved in phospholipase metabolism. Mutants are affected in hydrogen peroxide mediated cell death.




Getting acquainted with the candidates

- Scan information on the Locus page
 - Look at GO and PO annotations and Annotation detail evidence codes

Annotations ?	Category	Relationship Type ?	Keyword ?
	GO Biological Process	involved in	response to cold, programmed cell death, phosphatidic acid metabolic process
	GO Cellular Component	colocalizes with	microtubule cytoskeleton, plasma membrane
		located in	plasma membrane
	GO Molecular Function	has	phospholipase D activity
	Plant structure	expressed in	lateral root elongation zone, primary root elongation zone, cotyledon vascular system, leaf vascular system
Annotation Detail			

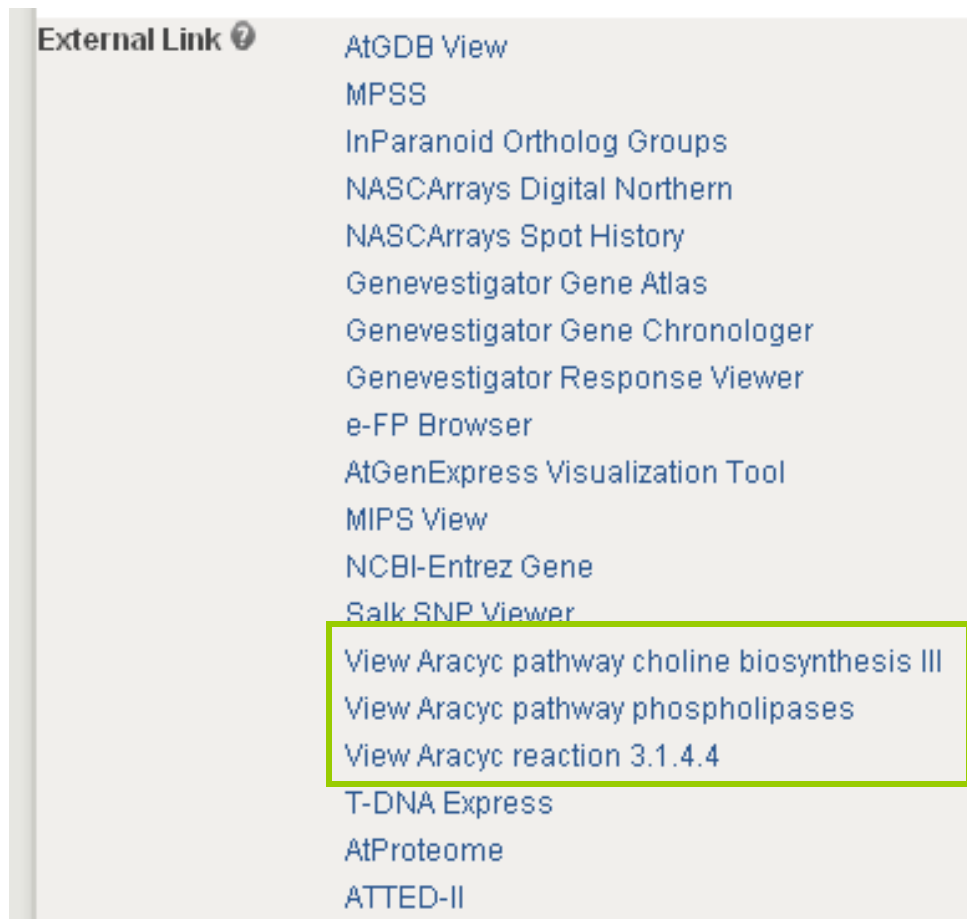
Getting acquainted with the candidates

- Scan information on the Locus page
 - Look at GO and PO annotations and Annotation detail evidence codes

Keyword Category	Relationship Type	Keyword	Gene	Evidence Code  Evidence Description  Evidence With: Evidence Reference 	Annotated By/ Date Last Modified
biological process	involved in	programmed cell death	ATPLDDELTA	<i>inferred from mutant phenotype:</i> analysis of physiological response: none: Zhang, et al. (2003)	The Arabidopsis Information Resource/ 2004-11-24
biological process	involved in	response to cold	AT4G35790.1	<i>inferred from expression pattern:</i> Protein levels (e.g. Western blots): none: Kawamura, et al. (2003)	Daniel MacLean/ 2006-06-18
cellular component	located in	plasma membrane	AT4G35790.1	<i>inferred from direct assay:</i> co-fractionation: none: Dunkley, et al. (2006)	Donghui Li/ 2007-08-09
molecular function	has	phospholipase D activity	ATPLDDELTA	<i>inferred from sequence or structural similarity:</i> Sequence similarity (homologue of/most closely related to): none: Elias M, et al. (2002)	The Arabidopsis Information Resource/ 2003-03-29

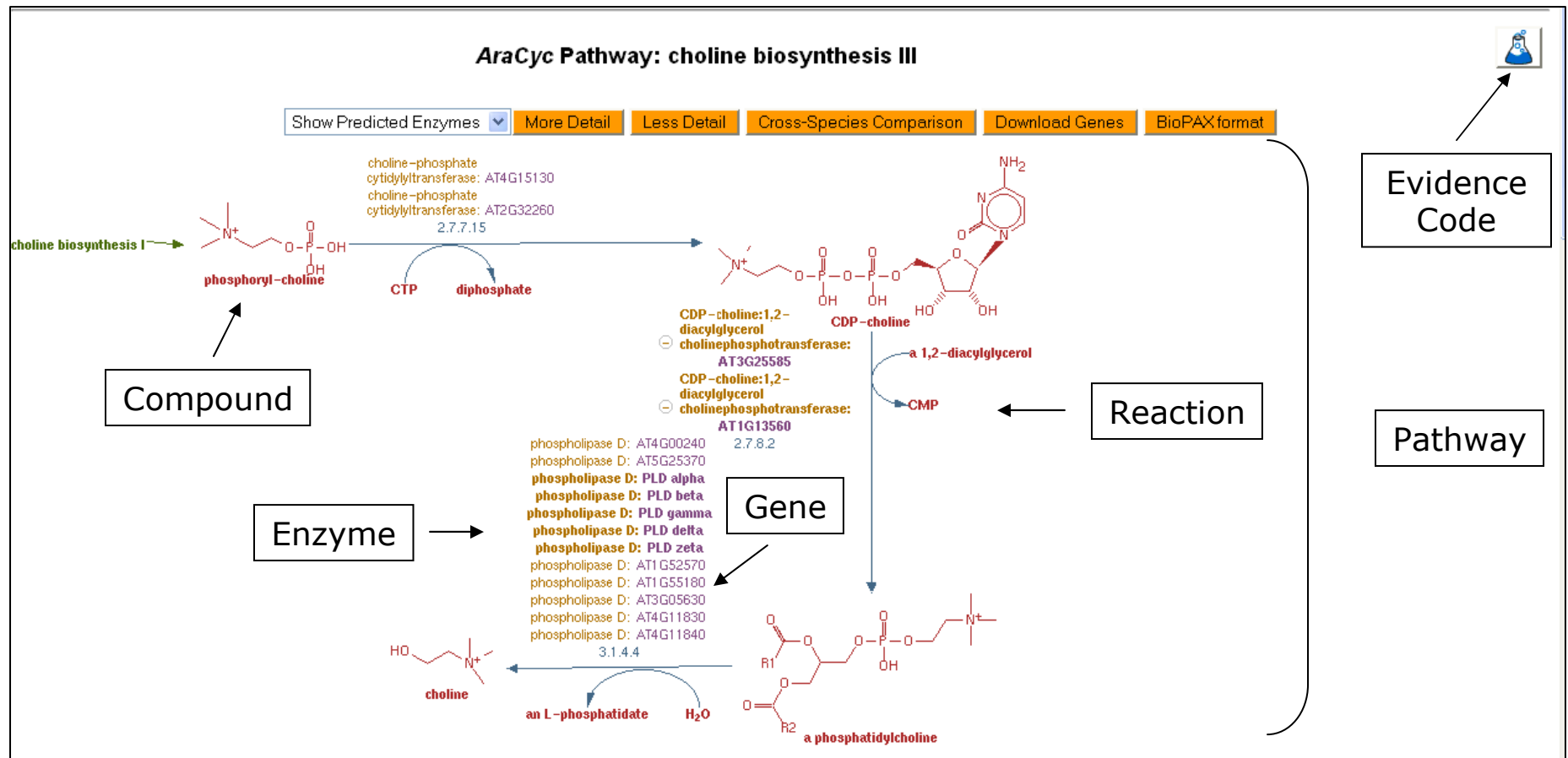
Getting acquainted with the candidates

- Take advantage of links on the Locus page
 - Connect to the AraCyc metabolic database (and the PMN in the future)



Getting acquainted with the candidates

- Connect to the AraCyc metabolic database (and the PMN in the future)



Getting acquainted with the candidates

■ Connect to the AraCyc metabolic database (and the PMN in the future)

Superclasses: [Biosynthesis](#) -> [Fatty Acids and Lipids](#) -> [Choline Biosynthesis](#)

Summary:

General information: Choline is a fundamental metabolite in plants because of its contribution to the synthesis of the membrane phospholipid phosphatidylcholine, which accounts for 40 to 60% of lipids in non-plastid plant membranes [[Mou02](#)]. Choline is also a precursor for the formation of glycine betaine ([glycine betaine biosynthesis III \(plants\)](#)) in certain plants such as spinach, where this osmoprotectant is accumulated and confers also tolerance to salinity, drought, and other environmental stresses. In addition choline has been recognized as an essential nutrient for humans [[McNeil01](#)].

The choline biosynthetic pathway enables plants to decouple choline synthesis from lipid metabolism (Kennedy pathway - [triacylglycerol biosynthesis](#)) and provides them with the metabolic flexibility to adapt to environmental conditions where large and variable amounts of choline are beneficial for survival [[Rontein01](#)].

Pathway information: The first step in choline biosynthesis is the direct decarboxylation of serine to ethanolamine [[Rontein01](#)], which is catalyzed by a serine decarboxylase unique to plants [[Rontein03](#)]. Ethanolamine is widely recognized as the entrance compound to choline biosynthesis.

The pathway variant displayed (nucleotide pathway) represents the biosynthetic route as found in diverse plant families. The synthesis of choline from ethanolamine may take place at three parallel pathways, where three consecutive N-methylation steps are carried out either on free-bases [[Prudhomme92](#)], phospho-bases [[Nuccio00](#)], phosphatidyl-bases [[McNeil01](#)] or a mixture of the latter [[Datko88a](#)] [[Datko88](#)] [[Hitz81](#)].

The synthesis of intermediates on both the phospho-base and phosphatidyl-base level includes the nucleotide pathway via CDP-phosphoaminoalcohol and the methylation pathway. However, it has been pointed out that the synthesis of phosphatidylethanolamine and phosphatidylcholine is characterized by a high degree of interaction and furcation on the various levels of arising intermediates. Consequently, it has been assumed that the reactions embedded in the nucleotide and methylation pathways may be two characteristics of one overall phosphoaminoalcohol pathway for the formation of phosphatidylcholine in plants [[Kinney93](#)].

The release of choline from the different pathway levels is also species-specific. Phosphocholine can either be directly dephosphorylated to release choline as observed in spinach [[Summers93](#)] or incorporated into phosphatidylcholine with the subsequent release of choline, as in tobacco [[McNeil00](#)]. The latter reaction has been shown to be specifically catalyzed by phospholipase D ([phospholipases](#)) in castor bean [[Wang94](#)]. Although a well-defined physiological role of phospholipase D (PLD) still await further research, progress has been made to assign some members of the heterogeneous family of PLD's to distinct cellular functions [[Kirk99](#)]. The remaining enzymes involved in this pathway, phosphoaminoalcohol cytidyltransferase and CDP-aminoalcohol phosphotransferase, cover a broader spectrum of substrates. This may be beneficial to process the heterogeneous mixture of possible substrates but it also indicates that the pathway flux is probably controlled more upstream [[Kinney93](#)].

Superpathways: [superpathway of choline biosynthesis](#)

Variants: [choline biosynthesis II](#) , [choline biosynthesis I](#)

Unification Links: [MetaCyc:PWY-3561](#)

Legend for Pathway Diagram

If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.

References

[Datko88](#): Datko AH, Mudd SH, (1988) "Enzymes of phosphatidylcholine synthesis in Lemna, soybean, and carrot." Plant Physiol. (1988), 88, 1338-1348.

[Datko88a](#): Datko AH, Mudd SH, (1988) "Phosphatidylcholine synthesis. Differing patterns in soybean and carrot." Plant Physiol. (1988), 88, 854-861.

[Hitz81](#): Hitz WD, Rhodes D, Hanson AD, (1981) "Radiotracer evidence implicating phosphoryl and phosphatidyl bases as intermediates in betaine synthesis by water-stressed barley leaves." Plant Physiol. (1981), 68, 814-822.



Confirming the candidate gene(s)

Confirming the candidate gene(s)

- Generate primers to sequence the region
 - Get sequence information from Locus page

Locus ?	Description ?	Gene Model(s) ?	Other Names ?	Keywords ?
1 AT4G35790	Encodes a protein with phospholipase D activity. Involved in phospholipase metabolism. Mutants are affected in hydrogen peroxi....	AT4G35790.1	F4B14.60 F4B14_60 ARABIDOPSIS THALIANA PHOSPHOLIPASE D DELTA	plasma membrane, phospholipase D activity, response to cold
Locus: AT4G35790				
Chromosome 4		full length CDS full length genomic full length cDNA		
Nucleotide Sequence ?				
		AT4G35790.3	metabolic process, lateral root elongation zone, primary root elongation zone, cotyledon vascular system, leaf vascular system	
			phospholipase D activity	
		AT4G35790.2	phospholipase D activity	

Confirming the candidate gene(s)

- Generate primers to sequence the region
 - Use Portal pages or Google the site to look for primer design resources

The Arabidopsis Information Resource (TAIR) maintains a database of biology data for the model higher plant *Arabidopsis thaliana*. Data includes the complete genome sequence along with gene structure, gene metabolism, gene expression, DNA and seed stocks, genome markers, publications, and information about the Arabidopsis community. Product function data is updated every two weeks from the latest and community data submissions. Gene structures are updated using computational and manual methods as well as community submissions. TAIR also provides extensive linkouts from our data pages to other resources.

Breaking News
[TAIR8 Release](#)
The Arabidopsis Information Resource is pleased to announce the release of the latest version of the Arabidopsis genome annotation (TAIR8) at TAIR and NCBI.

New Clones Received at ABRC

Confirming the candidate gene(s)

- Generate primers to sequence the region
 - Check primer sequence and orientation using Seqviewer

Whole Genome View Options:

SeqViewer Nucleotide View

[SeqViewer Home](#) | [Release Note](#) |

ATG = Translational Start/Stop

ATGC = Exon

atgc = Annotation on other strand [Hide](#)

atgc = UTR

atgc = Intron

Choose strand to view:

Forward Strand

Choose objects to be highlighted:

Genes

◀

Clicking on the arrows will shift the view by 5 kb

▶

ATGC = Your query sequence

10

20

30

40

50

60

70

80

90

100

Gene(s) +

Marker(s)

16959701

TTATCTCTGGGATCGACATCATCGGTATCGGTGGGTCTAGCACAGGCGTTACAGGCGGTGAAGAGACGGCGCAAAATGTTCTGAGAACATATCCATGTTAG

^ A

16959901

GTAACTCTAGCCTTAACAAATTTCAAATCGAGGTCAACCGTGTAGAAGCATAACGTCCTCCGATACTTTCTCCGC

| T

16960101

aggaacgaagaagaagaagaagaaagctaacaaaaaccgaaaccctaacgaacagggtcgcgttggttgctgctgcagaaaccaaagatgaattt

| 4

16960301

gaaaagtttttttttttttttcttttagaaaaacgaagatatttatttttttaggtgaaatgatttgagagagaagacaagcaaggaaggagaagaga

| G

16960501

ggacaggtggaggggacgacgtgagggagttggttgcctcactgacgacattattccgtgtagatctcagaggagtcggttgccacggtctcttcattagct

| 3

16960701

tctttaatcactgaactttacttaattagttttacaaaaactactgatgacgaatttgatttttagaacggttgtagcaatagctttgtagacgttaaagaaat

5

16960901

aaaaagctaataacaataatgctaaagttgacaatgatgatgtgaagagttcttctctttttctggacaaataattaagaagacaagtcaaaaaccagcta

7

16961101

tgcgaatctggtgatattaaccaccacacatccattggatagttttccaacatttgcttcattttattatatacaaaaataaaaattggaaaaatgattaatt

9

16961301

atcacataaaataaaaaatcccaaaaaccatttcttaaatggtgtggataaaaacaagaaataaaccatattggaaagtcgagctgtttaaacttttaagtg

0

16961501

gtcacacgtcaatcgataattgtcggaacaaggaaaaaaggcattctcgattcgtatctactcaatgaagcaacagcagtaacataaagaatcctttca

. ^ A

16961701

tggattacaaaatgaaccggtttcacatggtttactttctgaaatagaacaacacgacgcaaaaagccaaaaactaaaccatacgaaaagcccaagctcgat

1 | T

16961901

gttttcggtttttctatgctttctactatatttttttttggaggtgaggttgagtgatttatgagtcaggtttggaataatatacaactcagcaaaaaag

4

Confirming the candidate gene(s)

- Get seed for additional lines with mutations in the candidate loci
 - Identify Germplasm resources at the Locus page

Locus: AT4G35790					
Germplasm Showing 9 of 9 entries	Name/Image	Polymorphisms	Background	Stock Name	Select
	SAIL_1261_E04	SAIL_1261_E04.v1		CS846680	<input type="checkbox"/>
	Phenotypes ? None available				
	SALK_023247	SALK_023247.40.80.x		SALK_023247	<input type="checkbox"/>
Phenotypes ? None available					

Confirming the candidate gene(s)

- Get seed for additional lines with mutations in the candidate loci
 - Get a snapshot of insertion sites using GBrowse

Search

VISTA TIP - if you turn on the Vista track and everything below this band disappears, click "configure" next to "Report & Analysis - Annotate Vista Plot", then click "configure" on the new page.

Landmark or Region:

Data Source: Arabidopsis thaliana TAIR8

Reports & Analysis:

Annotate VISTA Plot

Scroll/Zoom: Show 4.653 kbp

Overview of Chr4

0M 1M 2M 3M 4M 5M 6M 7M 8M 9M 10M 11M 12M 13M 14M 15M 16M 17M 18M

Details

Locus
AT4G35790
protein_coding_gene

Protein Coding Gene Models
AT4G35790.1
AT4G35790.2
AT4G35790.3

T-DNAs/Transposons

FLAG_538E12
SALK_092469.40.65.N
SALK_023808.23.25.X
SALK_023247.40.80.X
BX547751
SALK_1
BX547

Confirming the candidate gene(s)

- Get seed for additional lines with mutations in the candidate loci
 - View sequencing trace from insertion using Seqviewer

| is end nearest insertion point
 === = sequence match
 --- = no sequence match

T-DNA/Tn annotation appears below the sequence

Choose strand to view:

Choose objects to be highlighted:

80 90 100 Gene (s) +
 | | | Marker (s)

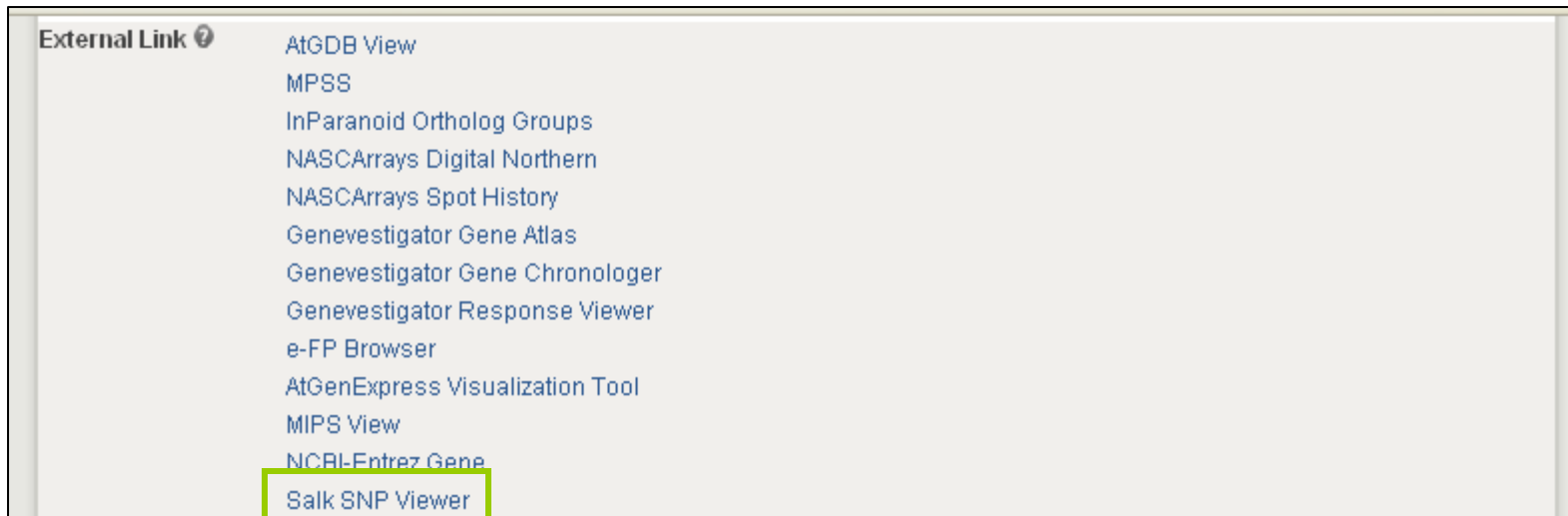
```

cttagtgagagcaataaagcgtaa ^ A
=====|T
aatacgtgagagcaataaagcgtaa | 4
=====|G
16955301 ttatgggtgtctgcaacaacaaagaaaaataacaataatcgtaatcagtcacgggattgatatttccgggtctgcggtctgggtttaaaccgctaatt 3
aaatattttgatctgtcctaagtggtccatgaatcttcgggttgggttcgggttaaacagaaacggatagggcacatctaatacctaataaaagtcgaagcaga 5
16955501 aatatttgggttgaacgcagttacacagcatcattatccgattttgaaaaagaaaaccgagaaaaatagaatatcctcaaatattttgcctcttcactca 7 ^ A
ctcagcatctctacattttgcacccaacaaatatttgacgacaatcacaaaagaatgtataaaaaagaaacacacagttaaatgtacaacaactcaaa 8 | T
16955701 caacactttcagctaatattattcaggatcaaaagaaaccccaacaaatgctgaaatttacattgaacattcacattacCGTGGTTAAAGTGTCAGGAA 5 | 4
GAGCCATGGAATGAGCTCCAATGATCTTACCACCAACATCTGGGAAGGTCTCGTAATCAGGAAGAGGGCTTACTTTACCATCAACGTCTACTGTAGAGG . | G
16955901 ATACTTTATTAAAGTGACCTTTGAGCTCTGAGAATTTCCGGGTCTATGAATCTTTCCAGTTTCTTCAGAGATTGTGTTACCTTCTTCAGACATTCCAGA 1 | 3
TCTGATGGCTCCACAACTCATCTCCAGTTTTCCTAAATGCTCTGCCCATAGTGACATTCTGTATCCATACACctgtgtagatagatagagatcatc | 5
16956101 aagctctttatgtcttacgaagactcaatatataatgggtcaagaatctgggagtagaacaacaaagttctctcaagtatctaagaacgtgatttttt | 7
cgcaagatcgaatgctatcccaatttctttactgtttttaaattccaggatcaggaatttgaaactttccccatgtagatatagagacacttttgaggt | 9
16956301 ttgggttatgatagatgatgatgatgaagatgtaggaggaagcaaacCTGGCCACGTGGGTGTCTTCCCTTGTGAGCCCATGTATGATTAGGTTGGT | 0
ATGCGCCCATGGCGATTTCAGTATCTTTGGTGCCTGCCATAGATCTTTGGTTGATATTAGCAGATCCCATGAGTACATACTCATCTACTATCATCCC | .
16956501 TTTTGGGTGCACGTAAATCATGAAACGCTGGAAATTATAAGAATCTGATACctgaaaaatgcagttaaaatgatcataaagaaattgttccactgattac | 1
tacaggaactttatgtatgattttatctatctgtaagctacagtttttggtaagtgaacacattcttagtcttagaagtttgaagattatacCACACTGCCA |
16956701 TTGGTGGCTGGCATATCATCTGGAAGCTGCTCTCGTTTACCAAGGCAGTAAAGTTAAGGTAATCGAGAGGATGAGCATCTGATTGCACCGCTTTCAGTT |
CTTTTGTATTAACATCATATCATCTGCATAGTTTGGCTctacgaaatgaaaagacagaagatgtagcagaatgaagtttctcttttaagaaat |
16956901 ggaacaacctattttggtattggcacttatcttacactcagctaaacaaagttacCTGCCAATATAGAAATTTCTTGACAGGGCCAGACTTTGGGTGCGC |
TTCAGGCCCAATGGTATGACAACATATACGGCAATCTTTCCTTAGCTCTGATTTTGTAAACATCTTTAGTGCCAACTCCATAGGAATAAGATTGTCA |
|=====FLAG_538E12=====|
16957101 GCTCctgctcatacaagacgaacacagaatcagaaaaatccggcggtgaaaaagtactaagctttatttaatacagttcaggttaactgtgaggtgaaaaat |
=====FLAG_538E12=====|
agagcccaaaatacaaaactagtaactgagaacaattctagacttcatgctgtctgcaatgtgagtattttatttcttagtctaagaactgtactcaag |
=====FLAG_538E12=====|
16957301 atcaaaatgcacCTGCGTCTCTATAAGGAAGGCCAAGCATAAGGAAGACCCAGGAAATACTGATTCTCGATATATATGAAATGCTGAGCAGATCTGATTGT |
=====FLAG_538E12=====|
CTGGATGTATGCAGTCTGGATGCTTTTATCTACAACAAGACGCTTGGCACATTCCAGATGctgcagttaaattatattgaagactttgtaagtgtgcaa |
=====FLAG_538E12=====|
16957501 ttctataccatgtggtggcttataaagagaatttcattgaaataacggtgattctcaaggaagataattaataatagcaataaagttgattgaccaa |

```

Confirming the candidate gene(s)

- Obtain additional polymorphic lines
 - Find natural variations using SALK SNP viewer from Locus page



- Use Ecotype search page to access appropriate seed

Confirming the candidate gene(s)

- Generate constructs for complementation
 - Check for alternative gene models at the [Locus page](#)

Locus: AT4G35790			
Date last modified	2003-05-02		
TAIR Accession	Locus:2125314		
Representative Gene Model ?	AT4G35790.1		
Other names:	ARABIDOPSIS THALIANA PHOSPHOLIPASE D DELTA, ATPLDDELTA, F4B14.60, F4B14_60, PLDDELTA		
Description ?	Encodes a protein with phospholipase D activity. Involved in phospholipase metabolism. Mutants are affected in hydrogen peroxide mediated cell death.		
Other Gene Models ?	ATPLDDELTA	AT4G35790.3 (splice variant)	AT4G35790.2 (splice variant)

[Map Links ?](#) [Map Viewer](#) [Sequence Viewer](#) [GBrowse](#)

- And then examine them using [GBrowse](#)
- **NOTE:** The xxxxx.1, xxxxx.2, etc. splice variants are numbered sequentially based on when they are identified; the last digit is not associated with confidence in any way
- **NOTE:** The "Representative Gene Model" does not provide any indication of confidence; the longest coding region is designated at the "Representative" to allow the maximum number of domain predictions, etc.

Confirming the candidate gene(s)

- Generate constructs for complementation
 - Decide on appropriate gene model in GBrowse

Tracks

Analysis <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> VISTA Plot			
Assembly <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> Annotation Units			
Community Annotation <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input checked="" type="checkbox"/> Eugene Predictions	<input checked="" type="checkbox"/> Gnomon Predictions	<input checked="" type="checkbox"/> Hanada et al. 2007 Gene Models	<input type="checkbox"/> Quesneville et al. Natural transposons
DNA <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> 6 Frame Translation	<input type="checkbox"/> DNA/GC		
Expression <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> CATMA primers	<input type="checkbox"/> Microarray ATH1-121501 probes	<input type="checkbox"/> MPSS seed 1	
<input type="checkbox"/> CATMA transcriptome data (analysis) (Aubourg et al. 2007)	<input type="checkbox"/> MPSS flower (Lu et al. 2005)	<input type="checkbox"/> MPSS seed 2	
Gene <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> CDS	<input type="checkbox"/> Noncoding RNAs	<input type="checkbox"/> Pseudogenes	
<input checked="" type="checkbox"/> Locus	<input checked="" type="checkbox"/> Protein Coding Gene Models	<input type="checkbox"/> Transposable element genes	
Genomic Features <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> Natural transposons			
Methylation <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> HMBD col BU/UB	<input type="checkbox"/> mCIP col/ddc BU	<input type="checkbox"/> mCIP ddc BU/UB	<input type="checkbox"/> mCIP met1 BU/UB
<input type="checkbox"/> mCIP col BU/UB	<input type="checkbox"/> mCIP col/met1 BU	<input type="checkbox"/> mCIP ddc/col BU	<input type="checkbox"/> mCIP met1/col BU
Sequence Similarity <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> Arabidopsis cDNAs	<input type="checkbox"/> Brassica BAC clones	<input type="checkbox"/> Brassica ESTs	<input type="checkbox"/> JAtY Clone Library
<input type="checkbox"/> Arabidopsis ESTs	<input type="checkbox"/> Brassica BAC ends	<input type="checkbox"/> Brassica GSSs	
Variation <input type="checkbox"/> All on <input type="checkbox"/> All off			
<input type="checkbox"/> Marker	<input type="checkbox"/> Polymorphisms	<input type="checkbox"/> T-DNAs/Transposons	

Configure tracks... Update Image

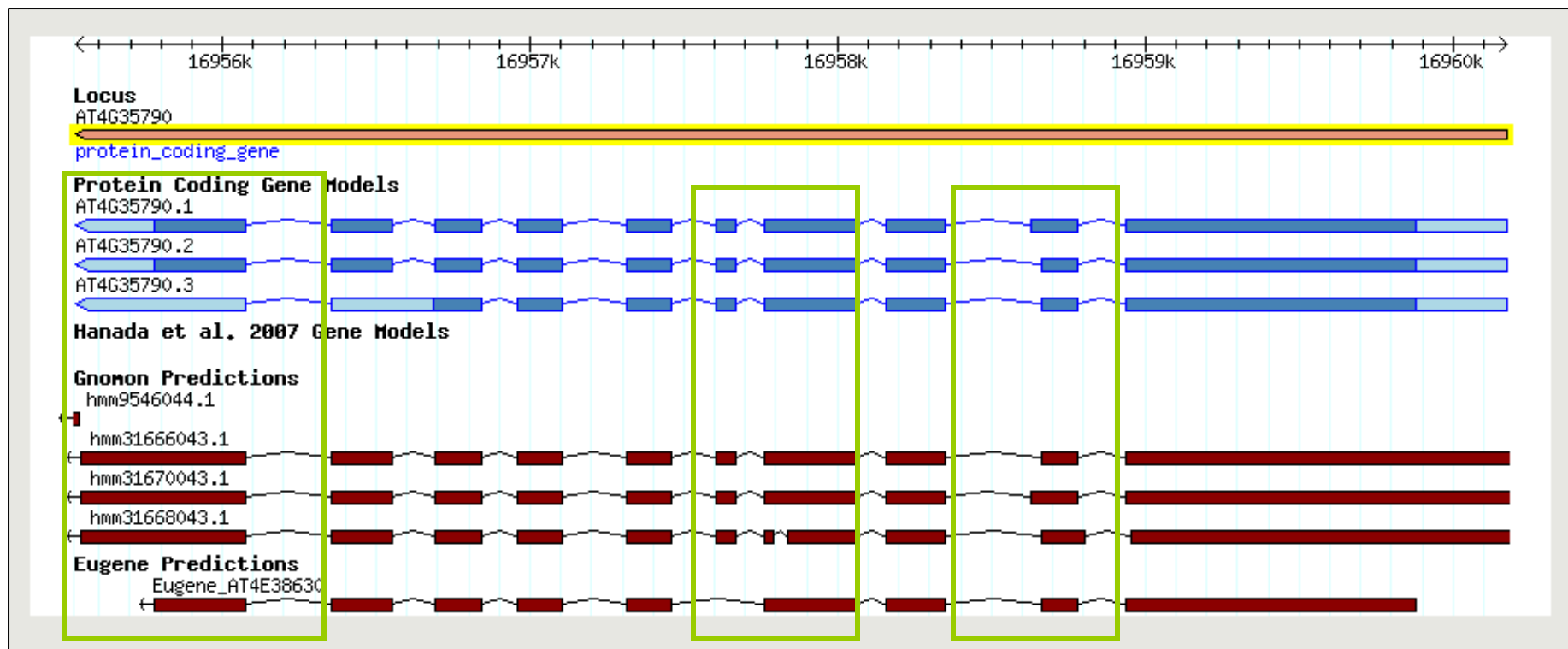
Confirming the candidate gene(s)

- Generate constructs for complementation
 - Check other gene prediction programs

□ Tracks		
□ Analysis <input type="checkbox"/> All on <input type="checkbox"/> All off		
<input type="checkbox"/> VISTA Plot		
□ Assembly <input type="checkbox"/> All on <input type="checkbox"/> All off		
<input type="checkbox"/> Annotation Units		
□ Community Annotation <input type="checkbox"/> All on <input type="checkbox"/> All off		
<input checked="" type="checkbox"/> Eugene Predictions	<input checked="" type="checkbox"/> Gnomon Predictions	<input checked="" type="checkbox"/> Hanada et al. 2007 Gene Models

Confirming the candidate gene(s)

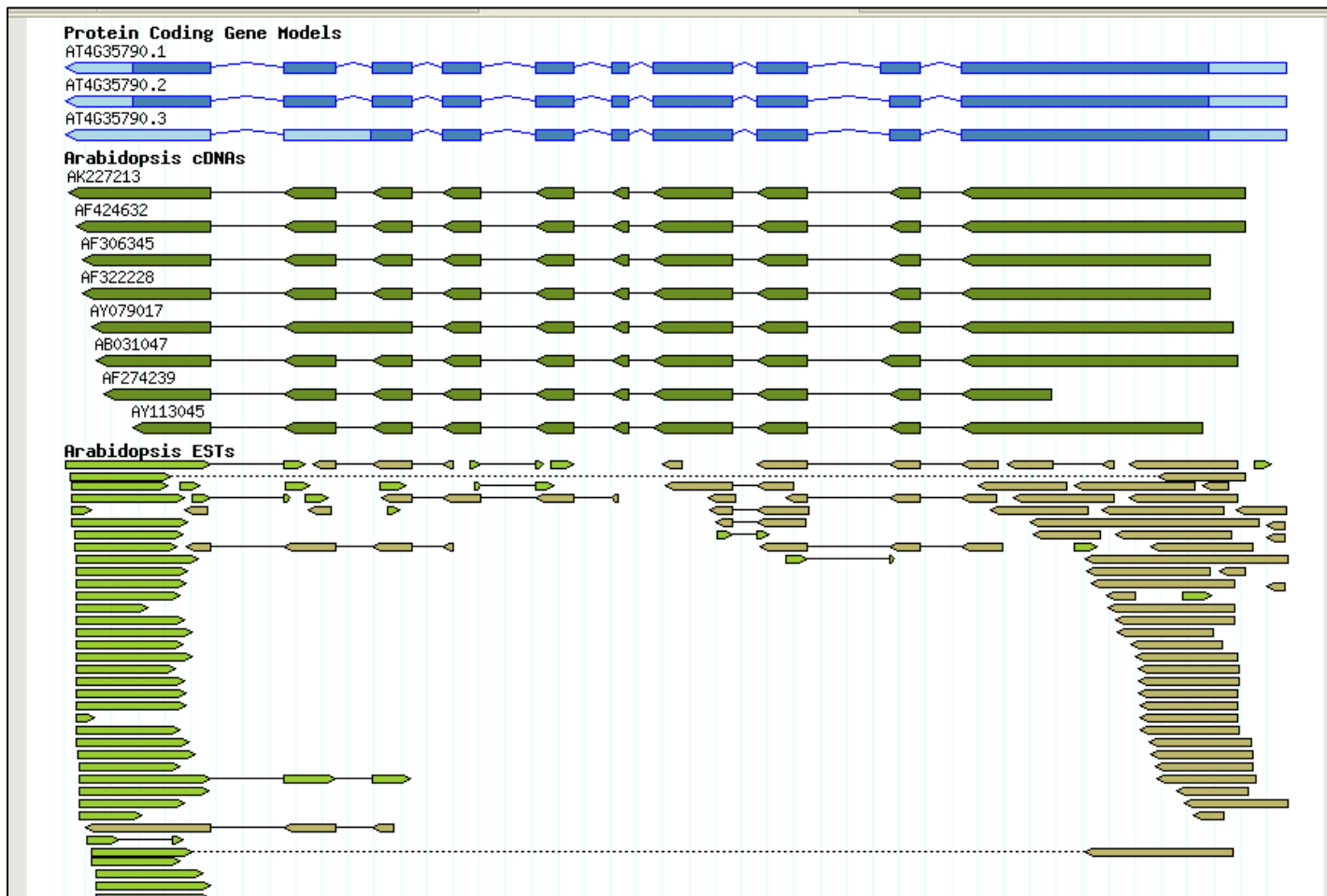
- Generate constructs for complementation
 - Check other gene prediction programs



Confirming the candidate gene(s)

- Generate constructs for complementation
 - Look at cDNA and EST support

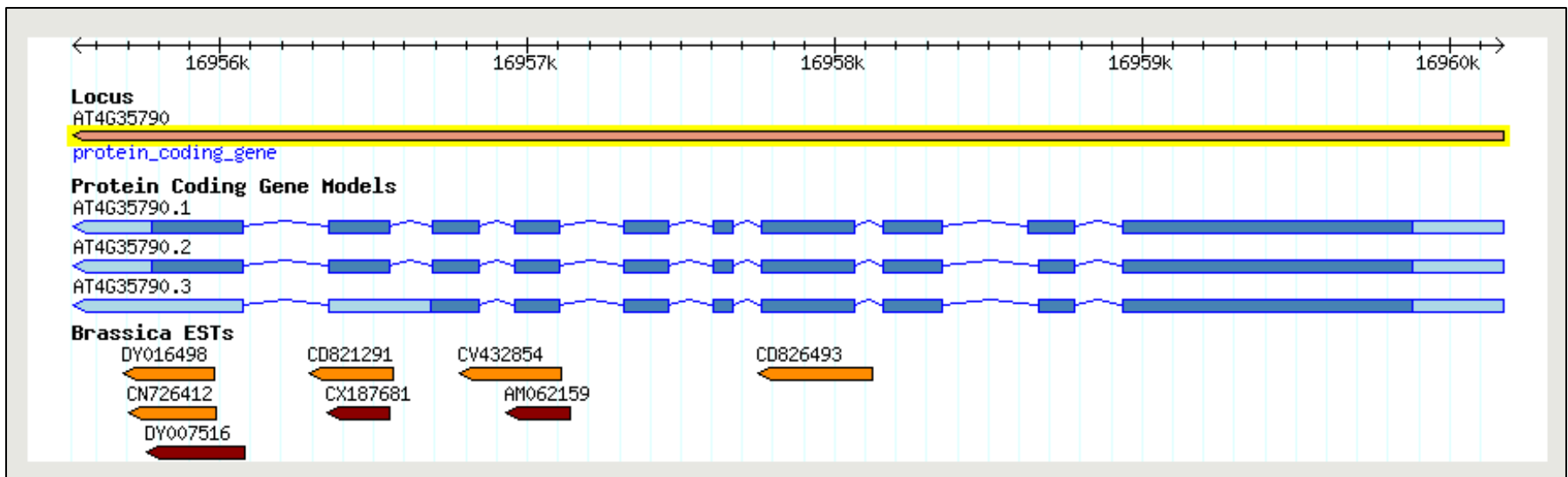
cDNAs



ESTs

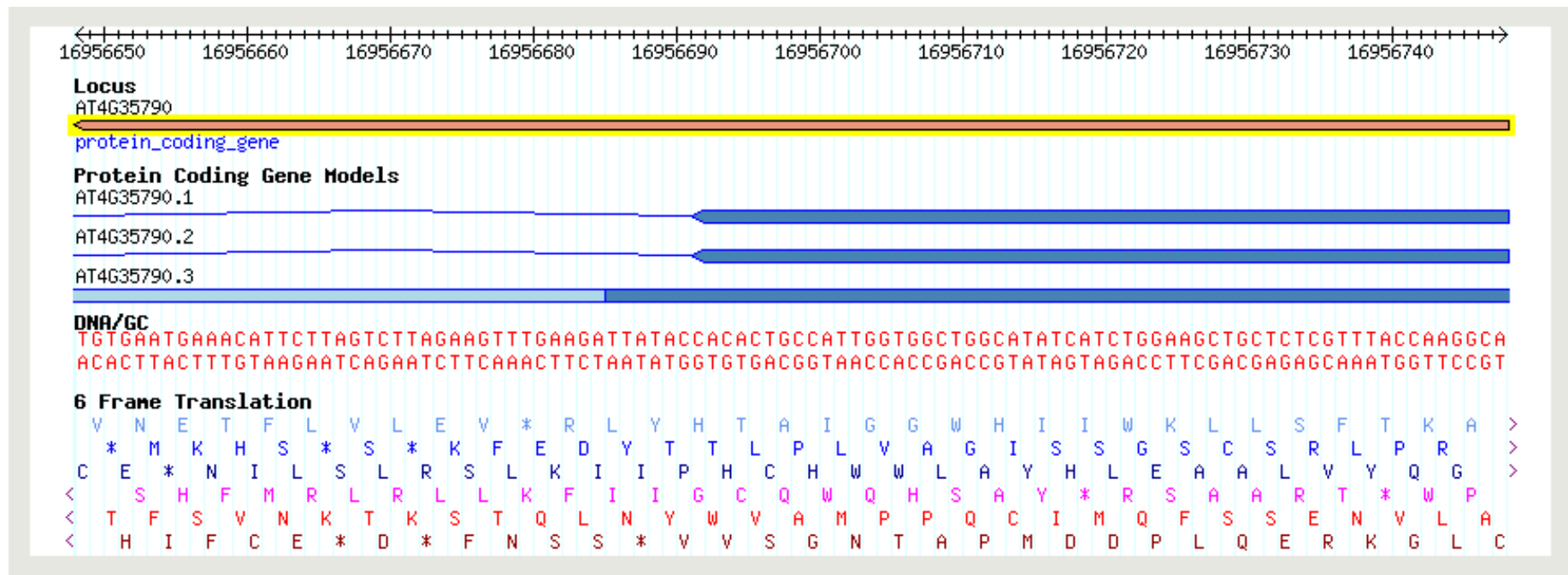
Confirming the candidate gene(s)

- Generate constructs for complementation
 - Look at *Brassica* EST support



Confirming the candidate gene(s)

- Generate constructs for complementation
 - Check translation frames and alternate amino acid sequences



Confirming the candidate gene(s)

- Generate constructs for complementation
 - Browse ABRC catalog for expression vectors

The screenshot shows the TAIR website interface. At the top is a navigation bar with tabs: Search, Browse, Tools, Stocks, Portals, Download, Submit, and News. The 'Stocks' tab is active, displaying a dropdown menu with options: Stocks Overview, ABRC Home, Browse ABRC Catalog, Supplier, Search Stocks, and Search Seed/Genes. The 'Browse ABRC Catalog' option is highlighted. To the left, the main heading 'The Arabidopsis Information Resource' is visible, followed by a brief description of the resource. To the right, a 'Breaking News' section is partially visible.

DNA STOCKS

I. LIBRARIES

- A. Genomic
- B. cDNA

II. CLONES

- A. Full-length cDNAs
- B. Cosmids
- C. Individually Cloned Genes and Constructs
- D. Mapping Clones
- E. Clone Sets
- F. YACs
- G. BACs
 - 1. Arabidopsis thaliana
 - 2. BACS from Close Relatives of Arabidopsis thaliana

H. ESTs

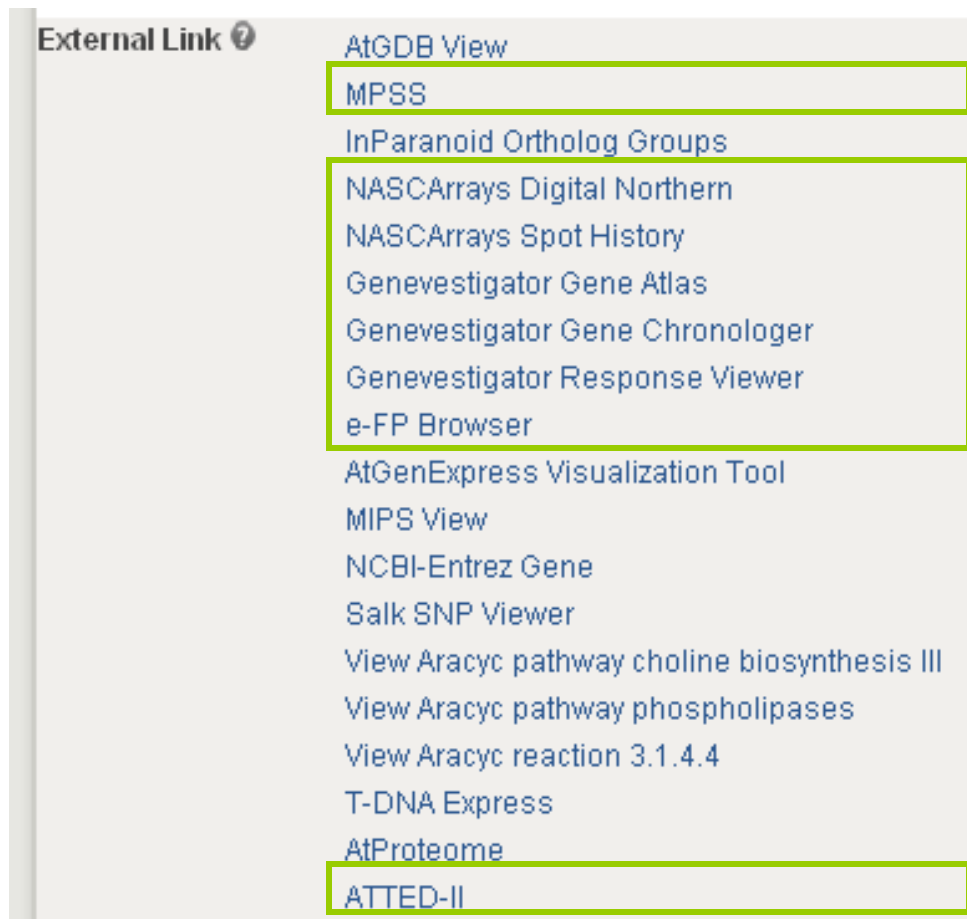
III. VECTORS



Characterizing the successful candidate

Characterizing the successful candidate

- Take advantage of links on the Locus page
 - View expression data



Characterizing the successful candidate

- Take advantage of links on the Locus page
 - Look for interacting proteins using link to IntAct



Characterizing the successful candidate

- Visit the Protein page
 - See predicted domains

Protein: AT4G35790.1					
Date last modified 		2008-04-22			
TAIR Accession 		AASequence:1009124182			
External IDs 		UniProtKB	GenPept	Similar Proteins in Genbank	
		Q9C5Y0	18419668	NCBI BLink	
Properties		Calculated MW 	98916.2		
		Calculated PI 	7.1735		
		Length (aa) 	868		
Domains 		Database	Structural Class Type 	Accession	Interpro
					Position
		superfam	Alpha and beta proteins (a+b)	56024	436-469
		superfam	Alpha and beta proteins (a+b)	56024	844-856
		superfam	All beta proteins	49562	76-178
		smart		SM00239	C2 calcium-dependent membrane targeting 15-153
		pfam		PF00168	C2 calcium-dependent membrane targeting 16-138
		smart		SM00155	Phospholipase D/Transphosphatidylase;Molecular Fun 368-403
		pfam		PF00614	Phospholipase D/Transphosphatidylase;Molecular Fun 713-740
		pfam		PF00614	Phospholipase D/Transphosphatidylase;Molecular Fun 368-403
		prosite		PS50004	C2 calcium-dependent membrane targeting 30-138

Characterizing the successful candidate

- Visit the Protein page
 - Use the link to SUBA for subcellular localization predictions

Organelle Targeting Predictions		
Predictor	Prediction	Data
iPSORT (Reference) (Web Service)	other	
LOCtree (Reference) (Web Service)	Chloroplast	Confidence: 5
MitoPred (Reference) (Web Service)		Confidence:
Mitoprot 2 (Reference) (Web Service)	other	Raw Score (Mit): 0.13 Raw Score (Cpt): 0.18 Cleaved Sequence:
MultiLoc (Reference) (Web Service)	other	Confidence: 0.96
Peroxp	perox	
Predotar (Reference) (OSC Service)		Raw Score (Mit): 0.01 Raw Score (Cpt): 0.01 Raw Score (Ret): Raw Score (None): 0.99
SubLoc (Reference) (Web Service)	cyto	Confidence: 2
TargetP (Reference) (Web Service)	other	Confidence: 3 Raw Score (Cpt): 0.17 Raw Score (Mit): 0.13 Raw Score (Sec): 0.17 Raw Score (Other): 0.74
WoLFPSORT (Reference) (Web Service)	Cytosol	Confidence: 8

Characterizing the successful candidate

□ Read the publications associated with the locus

Publication ?	title	source	associated gene models	date
	A leucine-rich repeat protein is required for growth promotion and enhanced seed production mediated by the endophytic fungus <i>Piriformospora indica</i> in <i>Arabidopsis thaliana</i> .	THE PLANT JOURNAL	AT4G35790.1 AT4G35790.3 AT4G35790.2	2007
	Early PLD(alpha)-mediated events in response to progressive drought stress in <i>Arabidopsis</i> : a transcriptome analysis.	JOURNAL OF EXPERIMENTAL BOTANY	ATPLDELTA	2007
	Mapping the <i>Arabidopsis</i> organelle proteome.	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	AT4G35790.1	2006
	Mass spectrometric approach for identifying putative plasma membrane proteins of <i>Arabidopsis</i> leaves associated with cold acclimation.	THE PLANT JOURNAL	AT4G35790.1	2003
	Rice phospholipase d isoforms show differential cellular location and gene induction.	PLANT AND CELL PHYSIOLOGY	ATPLDELTA	2003
	The Oleate-Stimulated Phospholipase D, PLD(delta), and Phosphatidic Acid Decrease H ₂ O ₂ -Induced Cell Death in <i>Arabidopsis</i> .	THE PLANT CELL	ATPLDELTA	2003
	Kinetic analysis of <i>Arabidopsis</i> phospholipase Ddelta: Substrate preference and mechanism of activation by Ca ²⁺ and phosphatidylinositol 4, 5-bisphosphate.	JOURNAL OF BIOLOGICAL CHEMISTRY	ATPLDELTA	2002
	Molecular diversity of phospholipase D in angiosperms.	BMC GENOMICS	ATPLDELTA	2002
	A novel phospholipase d of <i>Arabidopsis</i> that is activated by oleic Acid and associated with the plasma membrane.	PLANT PHYSIOLOGY	ATPLDELTA	2001
	Involvement of a novel <i>Arabidopsis</i> phospholipase D, AtPLDdelta, in dehydration-inducible accumulation of phosphatidic acid in stress signalling.	THE PLANT JOURNAL	ATPLDELTA	2001
	View Complete List (10 of 11 displayed)			

Characterizing the successful candidate

- Test predicted molecular function

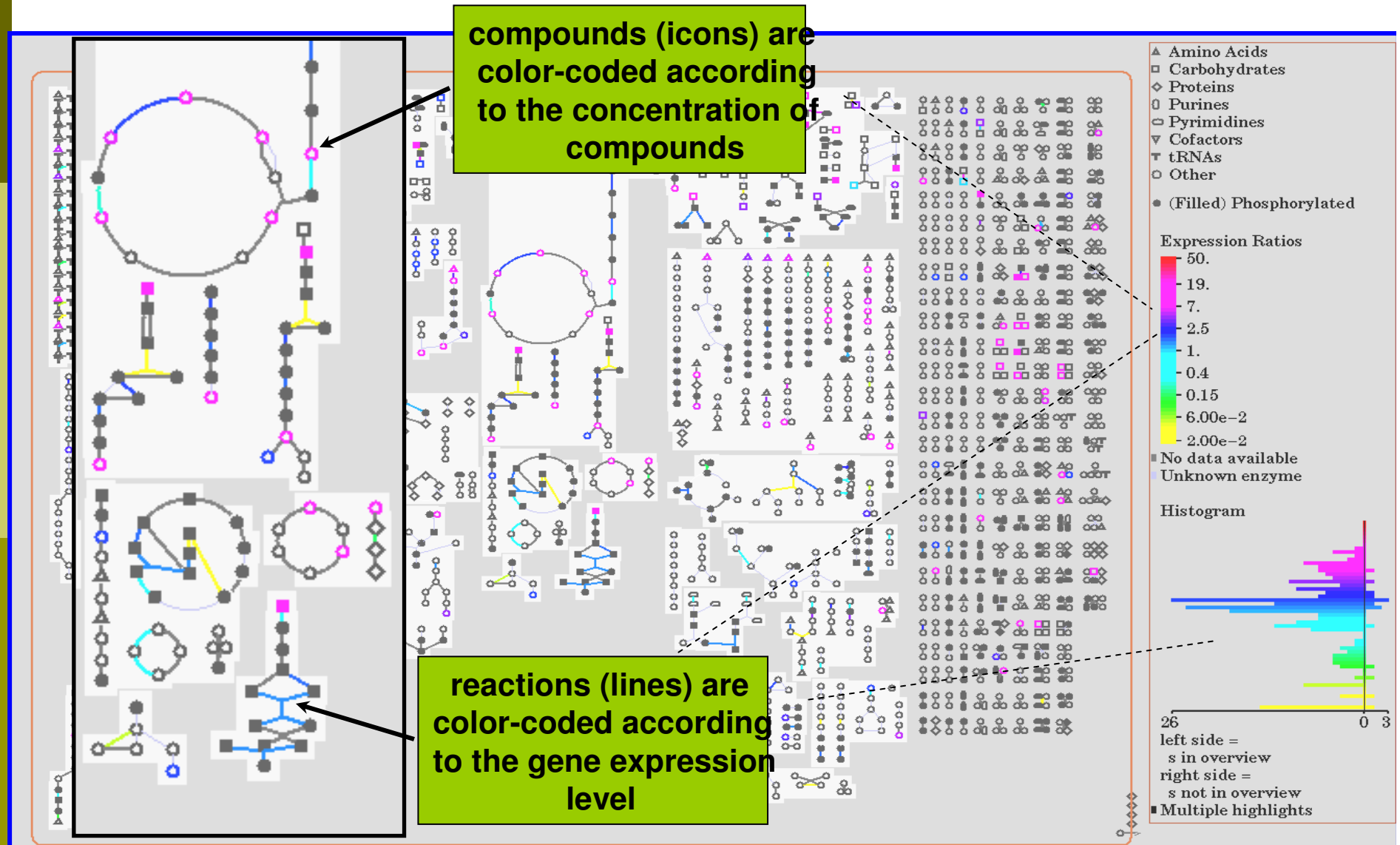
molecular function	has	phospholipase D activity	ATPLDDELTA <i>inferred from sequence or structural similarity:</i> Sequence similarity (homologue of/most closely related to): none: Elias M, et al. (2002)
-----------------------	-----	-----------------------------	---

- Order cDNA and vector for enzyme assay in *E.coli* using the [DNA/clones search page](#)

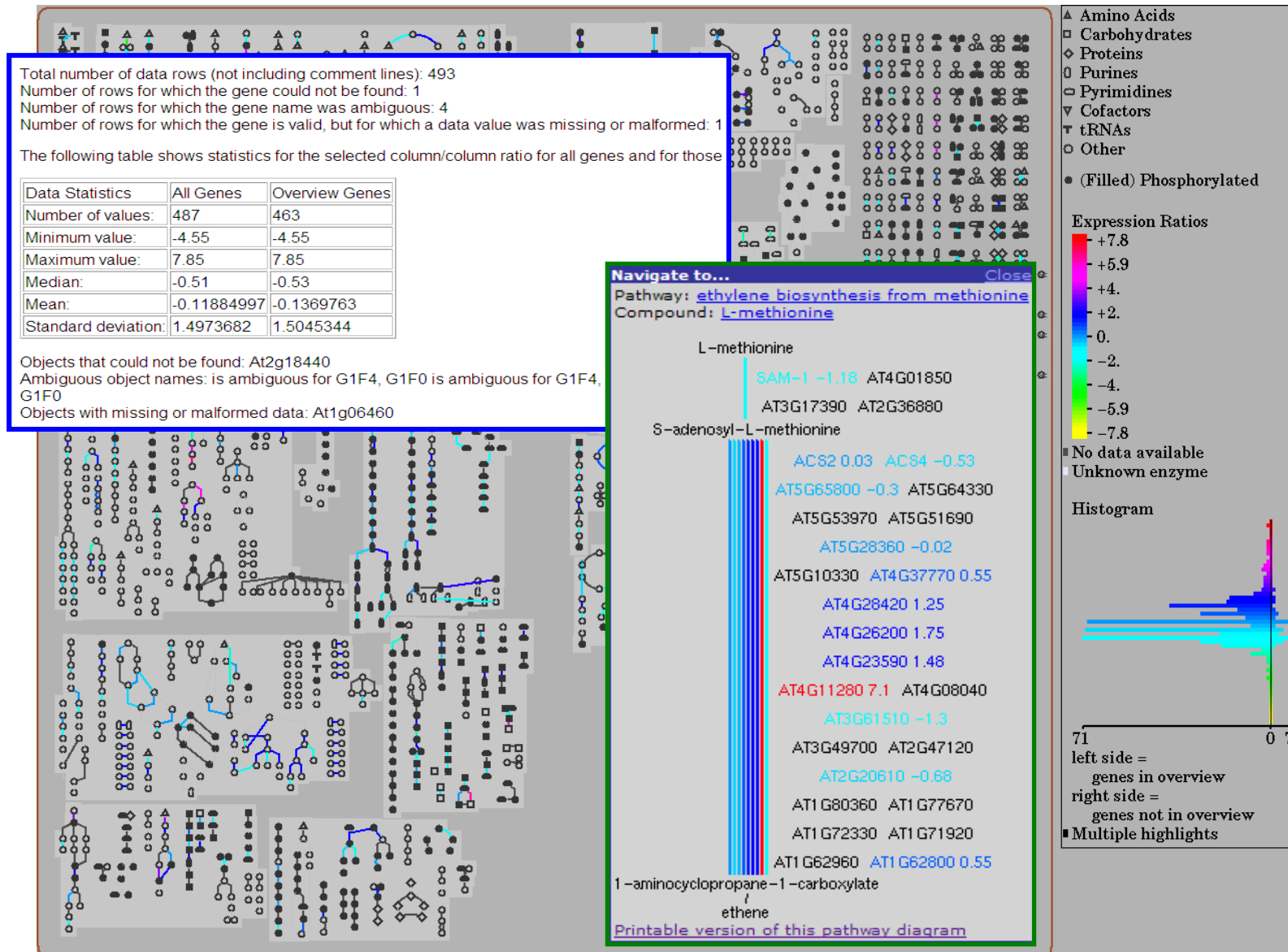
Characterizing the successful candidate

- Examine metabolomic and transcript profile of mutant plants
 - Use OMICS viewer in AraCyc to examine the results

Characterizing the successful candidate



Characterizing the successful candidate



Characterizing the successful candidate

- Identify and analyze related genes / proteins in Arabidopsis
 - Perform BLAST to find homologs in Arabidopsis

Characterizing the successful candidate

- Identify and analyze related genes / proteins in Arabidopsis
 - Use [AraCyc Advanced Query page](#) to generate a list of similar enzymes

Specify your query below

In database Arabidopsis thaliana COL (ARA) search for Proteins remove the entire condition below

Where NAME contains the substring phospholipase D

add a condition

insert a new search component here

Specify the contents of the output of your query below

Column 1	Column 2	
<input checked="" type="radio"/> Sort based on this column	<input type="radio"/> Sort based on this column	add a column
NAME	GENE	

Select your desired output format: ☒ HTML ☐ Tab Delimited Text (columns are separated by tabs)

Submit Query Reset Query

Characterizing the successful candidate

- Obtain nucleotide and amino acid sequences for family members
 - Use BULK Data Retrieval tool

Search	Browse	Tools	Stocks	Portals	Download	Submit	News
<h2>The Arabidopsis Information Resource</h2> <p>The Arabidopsis Information Resource (TAIR) maintains a database of genetic and biology data 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 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 by 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.</p> <p>The Arabidopsis Biological Resource Center at The Ohio State University collects, preserves and distributes seed and DNA resources of <i>Arabidopsis thaliana</i> and other species. Stock information and ordering for the ABRC are fully integrated into TAIR.</p>					Download Overview		
					Genes		
					GO and PO Annotations		
					Maps		
					Pathways	Information	
					Proteins	used to	
					Protocols	base of the	
					Microarray Data	me	
					Sequences) at TAIR and	
					Software		
					User Requests	ived at ABRC	
					Bulk Data Retrieval	TM	
						from J.	
						Ecker (Salk)	

Characterizing the successful candidate

- Obtain nucleotide and amino acid sequences for family members
 - Use BULK Data Retrieval tool

The screenshot displays the 'Sequences' section of the BULK Data Retrieval tool. It features a text area for 'Locus Identifiers or Sequences' containing a list of identifiers: AT4G35790, AT4G11850, AT3G16785, AT3G15730, AT2G42010, AT5G25370, AT4G11840, AT4G11830, AT4G00240, AT3G05630, and AT1G55180. Below this is an 'Upload file:' section with a 'Browse...' button. The 'Dataset:' dropdown is set to 'AGI transcripts'. The 'Output Options:' section shows 'Select output format:' with 'Fasta' selected (indicated by a green dot) and 'Tab-delimited text (format: id tab description)' as an option. At the bottom are three buttons: 'Reset', 'Get Sequences', and 'Send Sequences to EBI ClustalW'. A dropdown menu is open on the right, listing various sequence types: 'AGI transcripts' (highlighted), 'AGI coding sequences', 'AGI genomic locus sequences', 'AGI protein sequences', 'Upstream Sequences - 500bp', 'Upstream Sequences - 1000bp', 'Upstream Sequences - 3000bp', 'Downstream Sequences - 1000bp', 'Downstream Sequences - 3000bp', 'Intergenic Sequences', 'Intron Sequences', 'AGI 3' UTRs', and 'AGI 5' UTRs'.

Sequences

Locus Identifiers or Sequences:

- AT4G35790
- AT4G11850
- AT3G16785
- AT3G15730
- AT2G42010
- AT5G25370
- AT4G11840
- AT4G11830
- AT4G00240
- AT3G05630
- AT1G55180

Upload file: Browse...

Dataset: AGI transcripts

Output Options:

Select output format:

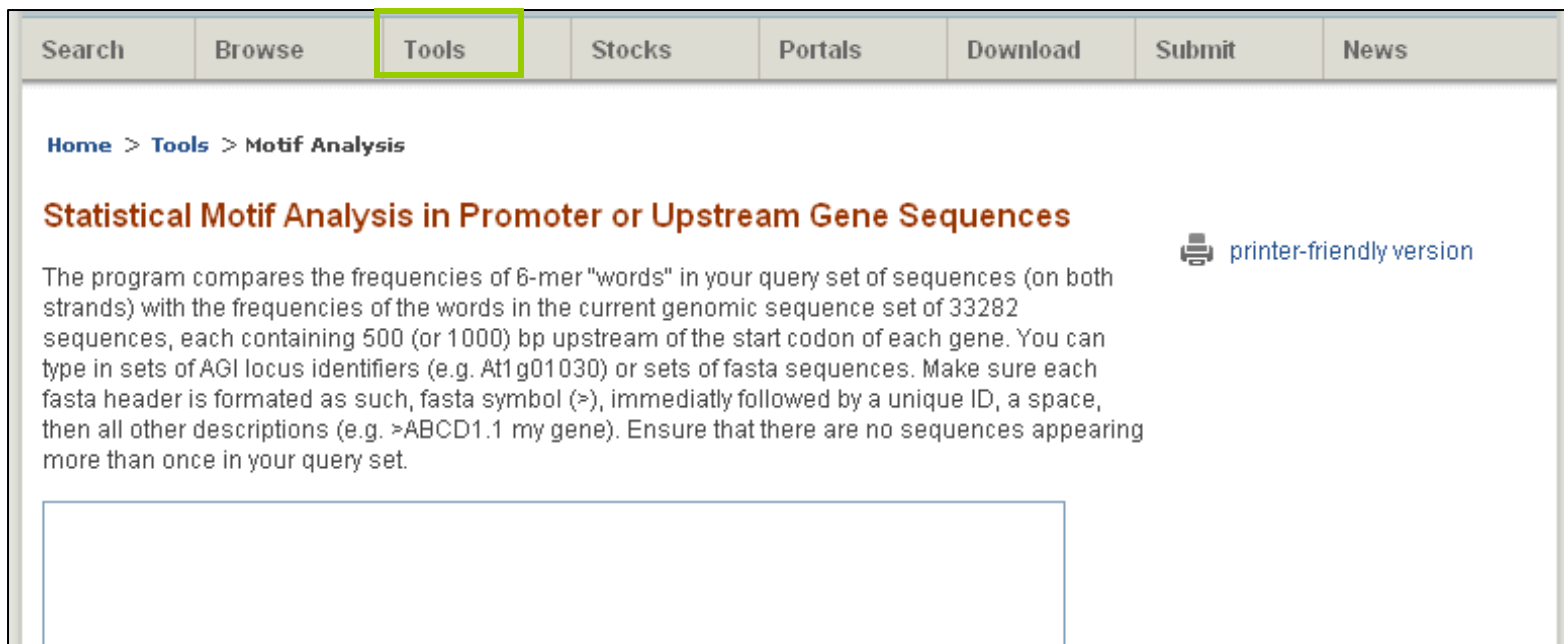
- ☒ Fasta
- ☐ Tab-delimited text (format: id tab description)

Reset Get Sequences Send Sequences to EBI ClustalW

AGI transcripts
AGI transcripts
AGI coding sequences
AGI genomic locus sequences
AGI protein sequences
Upstream Sequences - 500bp
Upstream Sequences - 1000bp
Upstream Sequences - 3000bp
Downstream Sequences - 1000bp
Downstream Sequences - 3000bp
Intergenic Sequences
Intron Sequences
AGI 3' UTRs
AGI 5' UTRs

Characterizing the successful candidate

- ❑ Search for common upstream regulatory elements
 - Use Motif Analysis to identify over-represented 6-mers in promoters




The screenshot shows a web application interface for Motif Analysis. At the top is a navigation bar with tabs: Search, Browse, Tools (highlighted with a green border), Stocks, Portals, Download, Submit, and News. Below the navigation bar is a breadcrumb trail: Home > Tools > Motif Analysis. The main heading is "Statistical Motif Analysis in Promoter or Upstream Gene Sequences". To the right of the heading is a printer icon and a link to "printer-friendly version". The text describes the program's function: comparing 6-mer frequencies in a query set against a genomic sequence set of 33282 sequences. It provides instructions on input formats (AGI locus identifiers or FASTA sequences) and formatting requirements. A large empty text box is at the bottom for user input.

Search Browse **Tools** Stocks Portals Download Submit News

Home > Tools > Motif Analysis

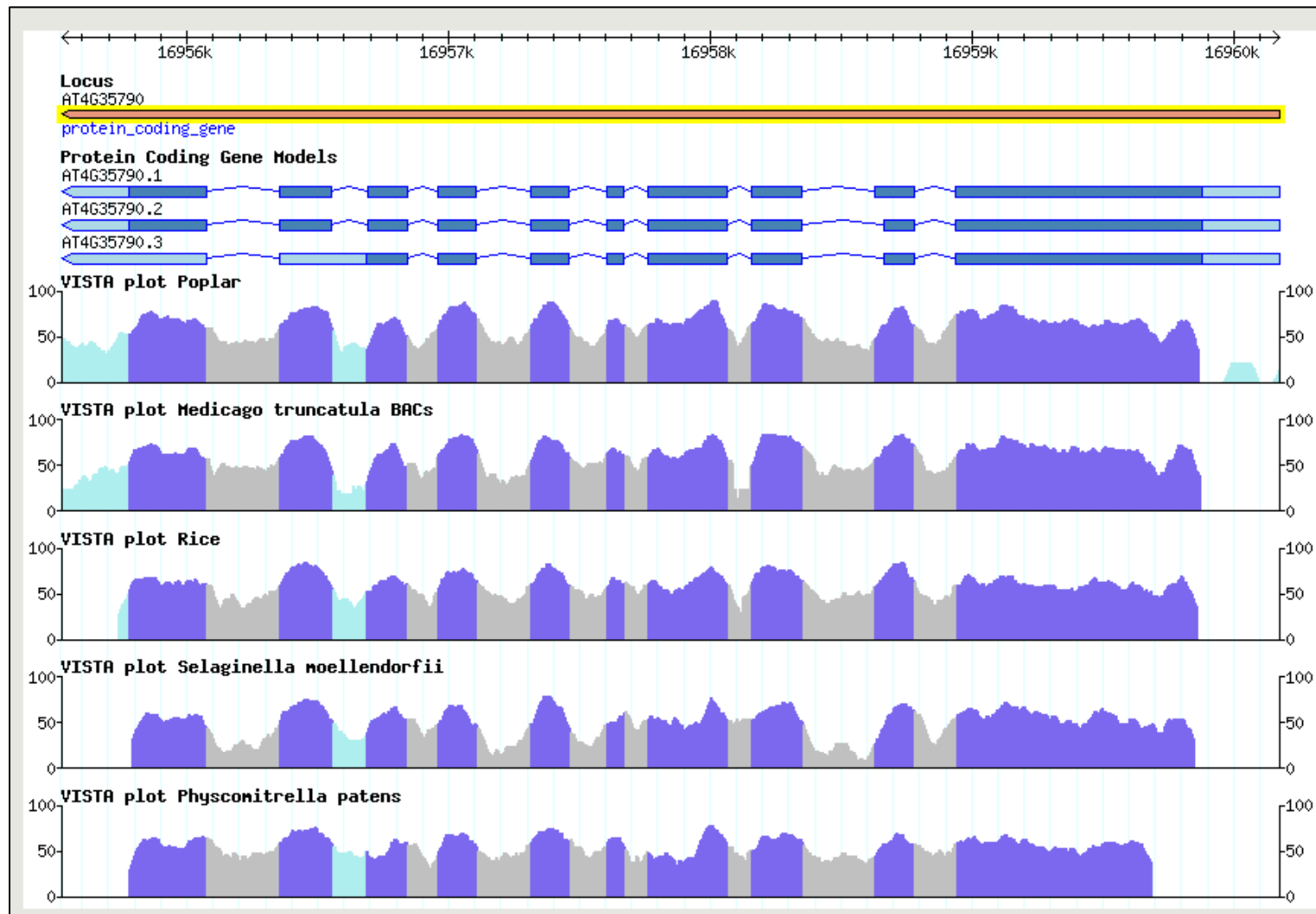
Statistical Motif Analysis in Promoter or Upstream Gene Sequences

 [printer-friendly version](#)

The program compares the frequencies of 6-mer "words" in your query set of sequences (on both strands) with the frequencies of the words in the current genomic sequence set of 33282 sequences, each containing 500 (or 1000) bp upstream of the start codon of each gene. You can type in sets of AGI locus identifiers (e.g. At1g01030) or sets of fasta sequences. Make sure each fasta header is formatted as such, fasta symbol (>), immediately followed by a unique ID, a space, then all other descriptions (e.g. >ABCD1.1 my gene). Ensure that there are no sequences appearing more than once in your query set.

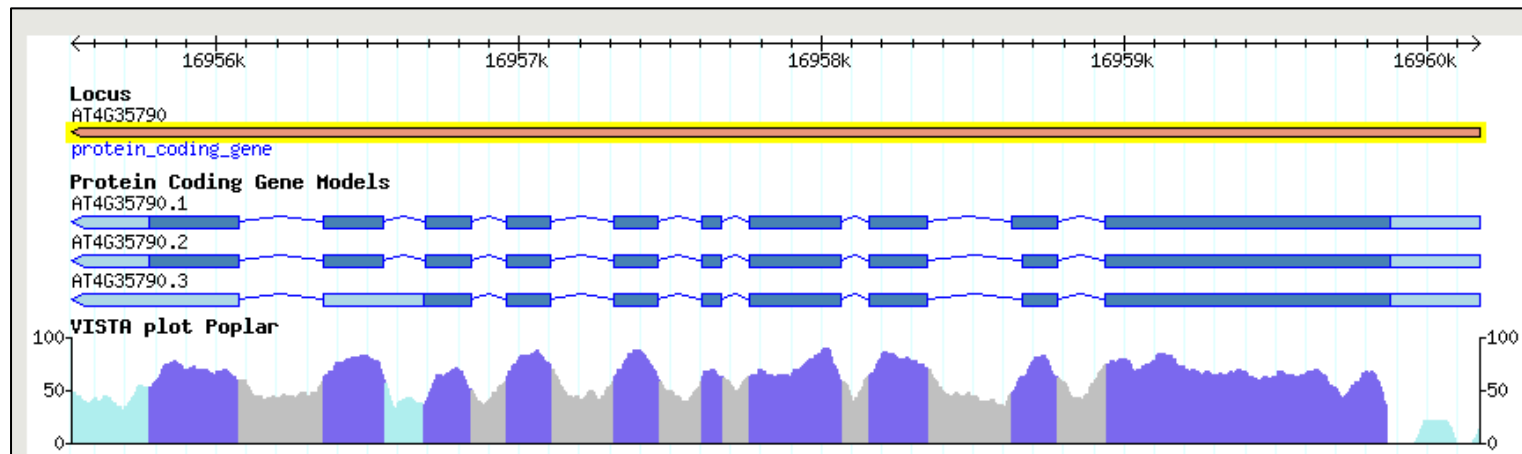
Characterizing the successful candidate

- View orthologs using VISTA tracks in GBrowse



Characterizing the successful candidate

□ View orthologs using VISTA tracks in GBrowse



Location on Arabidopsis	Location on Poplar	Alignment
chr4:16,955,522-16,959,857 (-) Sequence (softmasked) length: 4336bp	scaffold_57:67,032-73,442 (+) Sequence (softmasked) length: 6411bp VISTA Browser View in: <input type="text" value="Select Browser"/>	rankVISTA Alignment: Arabidopsis-Poplar MFA: Arabidopsis-Poplar CNS: Arabidopsis-Poplar rVISTA: Arabidopsis-Poplar PDF: Arabidopsis-Poplar
chr4:16,955,522-16,959,872 (+) Sequence (softmasked) length: 4351bp Overlap=4351bp	LG_II:920,536-926,812 (+) Sequence (softmasked) length: 6277bp VISTA Browser View in: <input type="text" value="Select Browser"/>	rankVISTA Alignment: Arabidopsis-Poplar MFA: Arabidopsis-Poplar CNS: Arabidopsis-Poplar rVISTA: Arabidopsis-Poplar PDF: Arabidopsis-Poplar

Characterizing the successful candidate

- Find other related proteins in different species using BLAST

Arabidopsis thaliana WU-BLAST2 Search

WuBlast

BLAST program

Datasets NEW

Input:

☒ query sequence

☐ locus name
(cds)
(At1g01030)

Or **upload** a file

FASTA, GCG or RAW [formats](#) accepted.

If sequence is less than 30 letters you need to set the S value to less than 100 or you can miss matches.

BLASTN : NT query to NT db

Green plant GB all (Protein)

At4g35790

Browse...

Characterizing the successful candidate

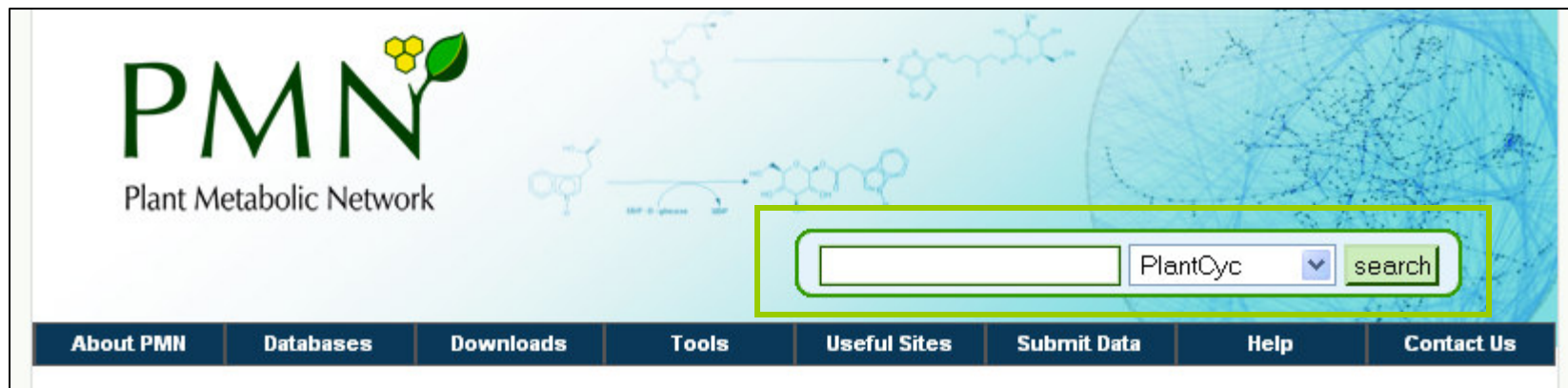
- Make connections to proteins in other species using GO terms on the Locus page

Annotations ?	Category	Relationship Type ?	Keyword ?
	GO Biological Process	involved in	response to cold, programmed cell death, phosphatidic acid metabolic process
	GO Cellular Component	colocalizes with	microtubule cytoskeleton, plasma membrane
		located in	plasma membrane
	GO Molecular Function	has	phospholipase D activity
	Plant structure	expressed in	lateral root elongation zone, primary root elongation zone, cotyledon vascular system, leaf vascular system
Annotation Detail			



Characterizing the successful candidate

- Make connections to proteins in other species using PlantCyc



Plant Metabolic Pathway Databases

PlantCyc allows access to information about shared and unique metabolic pathways present in over XXX plant species.

PlantCyc

Each species-specific metabolic pathway database provides access to the subset of biochemical pathways identified in a particular plant and enables users to examine experimental gene expression, proteomic, and metabolomic data overlaid on a comprehensive OMICS viewer metabolic map.

AraCyc	<input type="button" value="Search"/>	<input type="button" value="Metabolic Map"/>	<input type="button" value="More Info"/>
CapCyc (Pepper)	<input type="button" value="Search"/>	<input type="button" value="Metabolic Map"/>	<input type="button" value="More Info"/>
CoffeeCyc	<input type="button" value="Search"/>	<input type="button" value="Metabolic Map"/>	<input type="button" value="More Info"/>
EggplantCyc	<input type="button" value="Search"/>	<input type="button" value="Metabolic Map"/>	<input type="button" value="More Info"/>

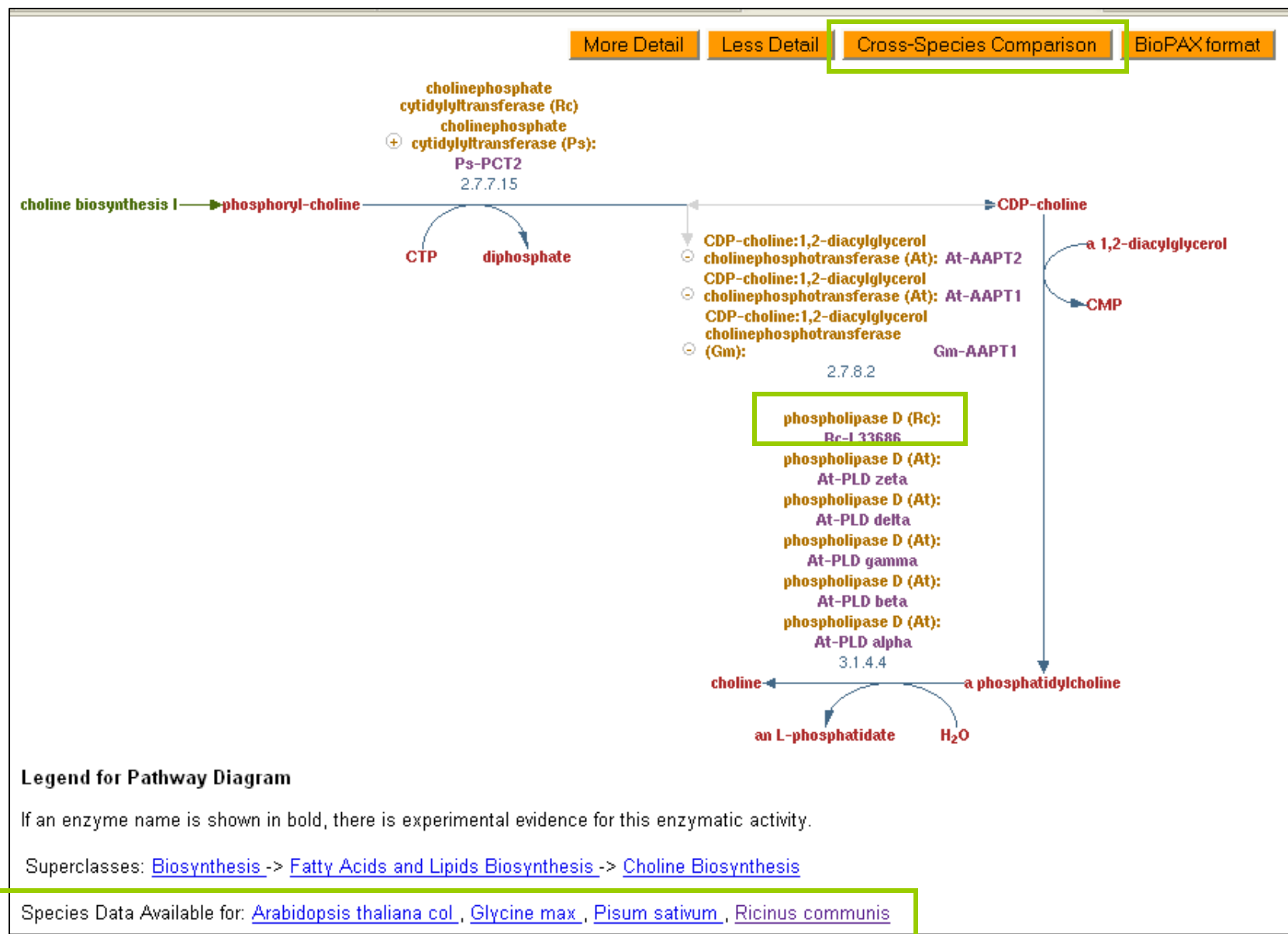
News

PMN International Summer Road Show!

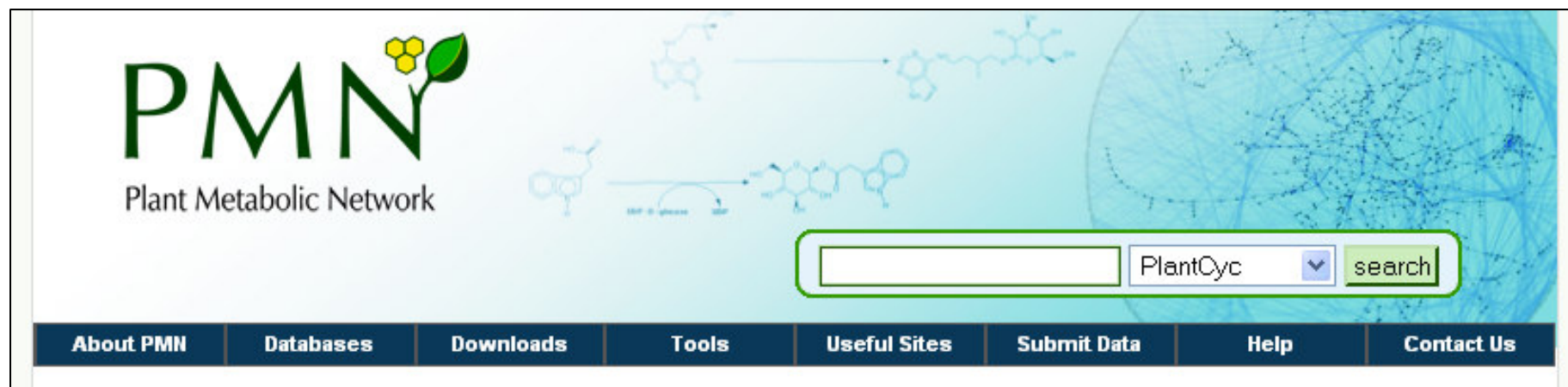
PMN representatives will visit several conferences in North America and Asia this summer.

- Come hear the director, Peifen Zhang, introduce the Plant Metabolic Network to the public for the first time at the **ASPB meeting** (June 27 - July 2) in **Merida, Mexico!**
- Additional appearances in **Canada** and **Japan** are highlighted on the [conferences](#) page.

Characterizing the successful candidate



Generating more “candidates” at the PMN



The image shows the header section of the Plant Metabolic Network (PMN) website. On the left is the PMN logo, which consists of the letters 'PMN' in a large, green, serif font, with a stylized green plant stem and two yellow flowers to the right. Below the logo is the text 'Plant Metabolic Network' in a smaller, black, sans-serif font. To the right of the logo is a light blue background featuring faint chemical structures and a network diagram. A search bar is located in the center-right, with a text input field, a dropdown menu set to 'PlantCyc', and a green 'search' button. Below the search bar is a dark blue navigation bar with white text links: 'About PMN', 'Databases', 'Downloads', 'Tools', 'Useful Sites', 'Submit Data', 'Help', and 'Contact Us'.

PMN
Plant Metabolic Network

Search: PlantCyc

[About PMN](#) [Databases](#) [Downloads](#) [Tools](#) [Useful Sites](#) [Submit Data](#) [Help](#) [Contact Us](#)

Generating more “candidates” at the PMN



- ❑ Leverage the known information about plant metabolism to predict the metabolic complement (pathways and enzymes) of agronomically and societally important crops as genome sequencing efforts advance.
 - a.k.a. Predict PoplarCyc, WheatCyc, GrapeCyc, MaizeCyc
 - Import predictions into PlantCyc
- ❑ Continually improve and refine databases as experimental support increases

Generating more “candidates” at the PMN

- Create initial data sets
 - Develop a “gold standard” set of plant metabolic enzymes
 - Primarily use experimentally verified enzymes
 - Initial set largely from Arabidopsis
 - Develop a “pipeline” : reciprocal BLAST- based protocol for predictively annotating **enzymes** in newly sequenced genomes
 - Try to determine appropriate E-value cut-offs for different classes of enzymes
 - Develop an extensive set of curated primary and **secondary metabolic pathways** from a diverse set of plants in PlantCyc


Generating more “candidates” at the PMN

- Predict new metabolic pathway databases
 - Use Pathologic program (SRI International) to predict new databases using:
 - Newly annotated **enzyme** set
 - PlantCyc **metabolic pathway** set
 - Use existing literature to “validate” pathways
 - Try to assess whether pathway is biologically feasible
 - Provide researchers with experimental hypotheses to test

Generating more “candidates” at the PMN

- ❑ Continually improve metabolic pathway databases
- ❑ Iterative Process
 - 1. Create PlantCyc: high quality multi-species enzyme / pathway data set
 - 2a. Generate new nnnnCycs using PlantCyc, new genomes, and Pathologic software
 - 2b. Repredict and refine existing “Cycs” (e.g. AraCyc) using PlantCyc, updated genomes, and Pathologic software
 - 3. Obtain experimental support for predicted pathways and enter newly described pathways (from literature)
 - 4. Update PlantCyc high quality enzyme and pathway data set

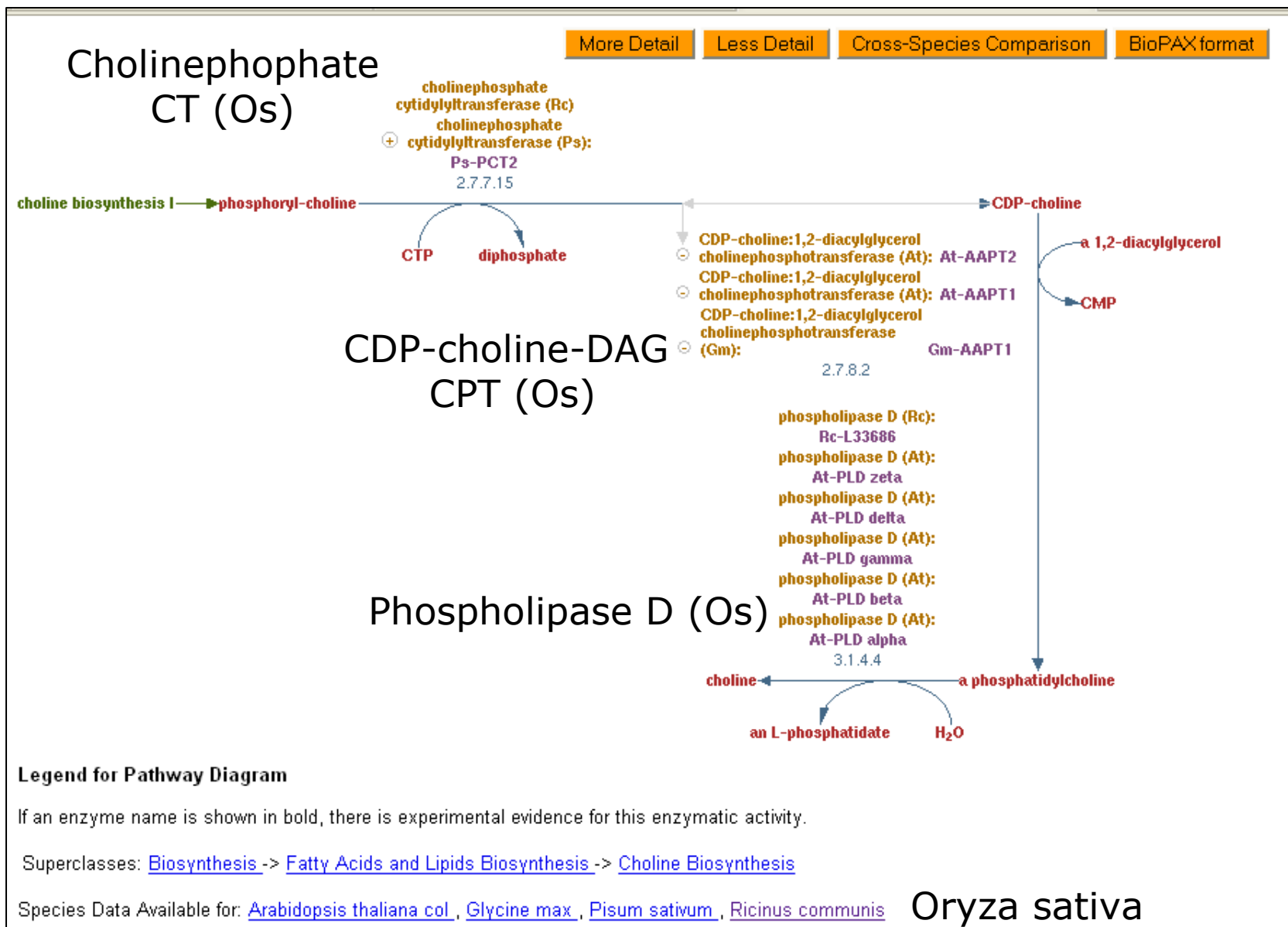
Generating more “candidates” at the PMN

- ❑ Continually improve metabolic pathway databases
 - ❑ Iterative Process
 - 1. Create PlantCyc: **higher** quality enzyme and pathway data set
 - 2a. Generate new nnnnCycs using PlantCyc, new genomes, and Pathologic software
 - 2b. Repredict and refine existing “Cycs” (e.g. AraCyc) using PlantCyc, updated genomes, and Pathologic software
 - 3. Obtain experimental support for predicted pathways and enter newly described pathways (from literature)
 - 4. Update PlantCyc high quality enzyme and pathway data set
- 

Generating more “candidates” at the PMN


- RiceCyc
 - Generated at and managed by Gramene curators
 - Member of the Plant Metabolic Network
 - Beneficiary of the database growth and refinement efforts

Generating more “candidates” at the PMN



Characterizing the successful candidate

□ Iterative Process

- 
- Obtain information from TAIR, AraCyc and the PMN about:
 - Gene sequences
 - Expression patterns
 - Protein functions
 - Metabolic processes
 - Relationships between evolutionarily and functionally “related” proteins within and across species
 - Use tools to further analyze data and make predictions
 - Make testable hypotheses
 - Get exciting experimental results
 - Transfer them to new organisms and share them with others



Transferring your knowledge to rice

Transferring your knowledge to rice

- Gain access to rice resources through MODs portal page

Search Browse Tools Stocks **Portals** Download Submit News

Home > Portals > Genome Annotation > Gene Structural Annotation > Other Plant Genomes

Genome Snapshot

Gene Structural Annotation

- Genome Assembly
- Genome Annotation
- Genome Publications and Documents
- Submit Structural Annotation to TAIR

Gene Functional Annotation

- Ontologies
- TAIR Gene Search
- Search/Browse TAIR Keywords
- TAIR GO Annotation Download
- Search/Browse Gene Ontology Database
- Search/Browse Plant Ontology Database
- Submit Functional Annotation to TAIR
- FTP Downloads

Genome Annotation Tools

- Gene Structural Annotation Tools

Other Genome Databases and Sequencing Efforts

This page provides a list of **plant-specific** and **non-plant** model organism databases, as well as links to other sites that maintain **links to numerous organism-specific resources**.

This page also has links to information about **other plant genome sequencing projects**

Plant Model Organism Databases

BeanGenes - beans
Phaseolus and *Vigna* species

The Brachypodium distachyon Information Resource - model grass
Brachypodium distachyon

Chlamydomonas Center - green alga
Chlamydomonas

ForesttreeDB - white spruce, loblolly pine, and poplar
Picea glauca, *Pinus taeda*, and *Populus* species

GrainGenes - grasses
Triticum and *Avena* species



Gramene - cultivated rice, wild rice, maize, wheat, barley, sorghum, pearl millet, foxtail, and oats
Oryza sativa and *Zizania* species, *Zea mays*, *Triticum aestivum*, *Hordeum vulgare*, *Sorghum bicolor*, *Pennisetum glaucum*, *Alopecurus pratensis*, *Avena* species, and other grasses



Sharing your knowledge with everyone . . .

Sharing your resources and tools . . .

- ABRC Submission
 - Seed stocks
 - DNA stocks

Search	Browse	Tools	Stocks	Portals	Download	Submit	News
<h2>The Arabidopsis Information Resource</h2> <p>The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant <i>Arabidopsis thaliana</i>. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, metabolism, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every two weeks from the latest published research literature 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 and updated genes. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.</p> <p>The Arabidopsis Biological Resource Center at The Ohio State University collects, reproduces, 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.</p> <p> TAIR is located at the Carnegie Institution for Science Department of Plant Biology and funded by the National Science Foundation. </p>						Submit Overview	
						ABRC Stock Donation	
						External Links	
						Gene Class Symbol Registration	
						Gene Family Data	
						Gene Functional Annotations	
						Gene Structure Annotations	
						Locus Identifier Request	
						Marker and Polymorphism Data	
						Metabolic Pathway Data	
						Phenotypes	
						Protocols	
						2010 Functional Genomics Gene List	

Sharing your data . . .

- Data submission forms
 - Gene Family Data
 - Gene Functional Annotations
 - Gene Structure Annotations
 - Locus Identifier Request
 - Marker and Polymorphism Data
 - Metabolic Pathway Data
 - Phenotypes
 - Protocols
 - 2010 Functional Genomics Gene List

Sharing your data . . .

- Opportunities at the PMN
 - Help create new “Cycs”
 - Help validate new and existing “Cycs”
 - Literature-based support
 - New experimental support
 - *Please contact me if you’re interested

Publishing your results . . .

- Please help each other (and the curators)
 - List the AGI locus identifier for each gene in the publication
 - Provide a unique identifier, e.g. SAIL T-DNA number, ABRC or NASC stock number, etc. for each germplasm used in the publication
 - Check the gene registry and register your gene symbols
 - Be careful when assigning allele numbers
 - Textpresso to the rescue . . .
(www.textpresso.org/arabidopsis/)
 - Submit any new sequence data to GenBank (e.g. new cDNA or EST sequence) even other sequences already exist

Publishing your results . . .

- TAIR-Journal collaboration
 - Authors submit functional annotations at time of paper acceptance
 - Plant Physiology requires participation as of March 2008
 - Other journals may join in the future

Sharing your expertise . . .

□ Community Annotation (present)

■ Add Comments

- Locus page
- Gene model page
- Protein page
- Marker page
- Polymorphism page
- Germplasm page
- Clone page
- Etc.
 - Only 372 comments since 2002

Locus
page →

Comments ? (shows only the most recent comments by default)	entered by	date	comment
	Robert Last	06/02/2005	Additional relevant references: 1. Cloning A nuclear gene encoding mitochondrial...

Associated Nucleotide Sequences			
Flanking Sequences		primer 1: CTCCACCAATCATGCAAATG primer 2: TGATGTTGATGGAGATGGTCA	
Comments (shows only the most recent comments by default)	entered by	date	comment
	Ooi-Kock Teh	07/21/2005	Works better with these primers 5'ctccaccaatcatgcaaagt3' 5'tgatgtgatggagatggtcag3' Anneal...

← Genetic
marker
page

Sharing your expertise . . .

- Community Annotation (future?)
 - Voluntary “gene expert” editors?
 - Meetings with curators at conferences or on campuses?
 - Wiki-style annotation?

Sharing your expertise . . .

- Community Annotation (future?)
 - Wiki-style annotation?
 - Pros
 - Allows rapid sharing of information
 - Gives scientists a direct route to enter data
 - Errors can be quickly corrected
 - Cons
 - Information entered in a Wiki page is not part of a database
 - Information entered in a Wiki page is not amenable to structured searches and advanced queries
 - Computational analyses cannot be performed easily on Wiki content
 - Inaccurate information may be posted
 - Wiki pages provided by other databases have not been popular
 - Existing “Add Comment” option provides some of the beneficial features of Wiki-style annotation

Biologists, biocurators, and biological databases: building a bright future of synergistic interactions

- ❑ You provide the data.
- ❑ You provide the tools.
- ❑ You provide the resources.



Biologists, biocurators, and biological databases: building a bright future of synergistic interactions

- ❑ You provide the data.
- ❑ You provide the tools.
- ❑ You provide the resources.
- ❑ TAIR and the PMN strive to help you easily share your work to promote scientific progress around the world
- ❑ We continue to need your advice to make that a reality!



Acknowledgements

TAIR, AraCyc, and the PMN

Eva Huala (*Director and Co-PI*)

Sue Rhee (*PI and Co-PI*)

Current Curators:

- Peifen Zhang (*Director and lead curator- metabolism*)
- Tanya Berardini (*lead curator – functional annotation*)
- David Swarbreck (*lead curator – structural annotation*)

- Debbie Alexander (*curator*)
- A. S. Karthikeyan (*curator*)
- Philippe Lamesch (*curator*)
- Donghui Li (*curator*)

Recent Past Contributors:

- Christophe Tissier (*curator*)
- Hartmut Foerster (*curator*)

Tech Team Members:

- Bob Muller (*Manager*)
- Larry Ploetz (*Sys. Administrator*)
- Raymond Chetty
- Anjo Chi
- Vanessa Kirkup
- Cynthia Lee
- Tom Meyer
- Shanker Singh
- Chris Wilks

Metabolic Pathway Software:

- Peter Karp and SRI group

NSF



Thank you . . .



www.arabidopsis.org

curator@arabidopsis.org



www.arabidopsis.org/biocyc

curator@arabidopsis.org



coming SOON!

kadreher@tairgroup.org