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?**

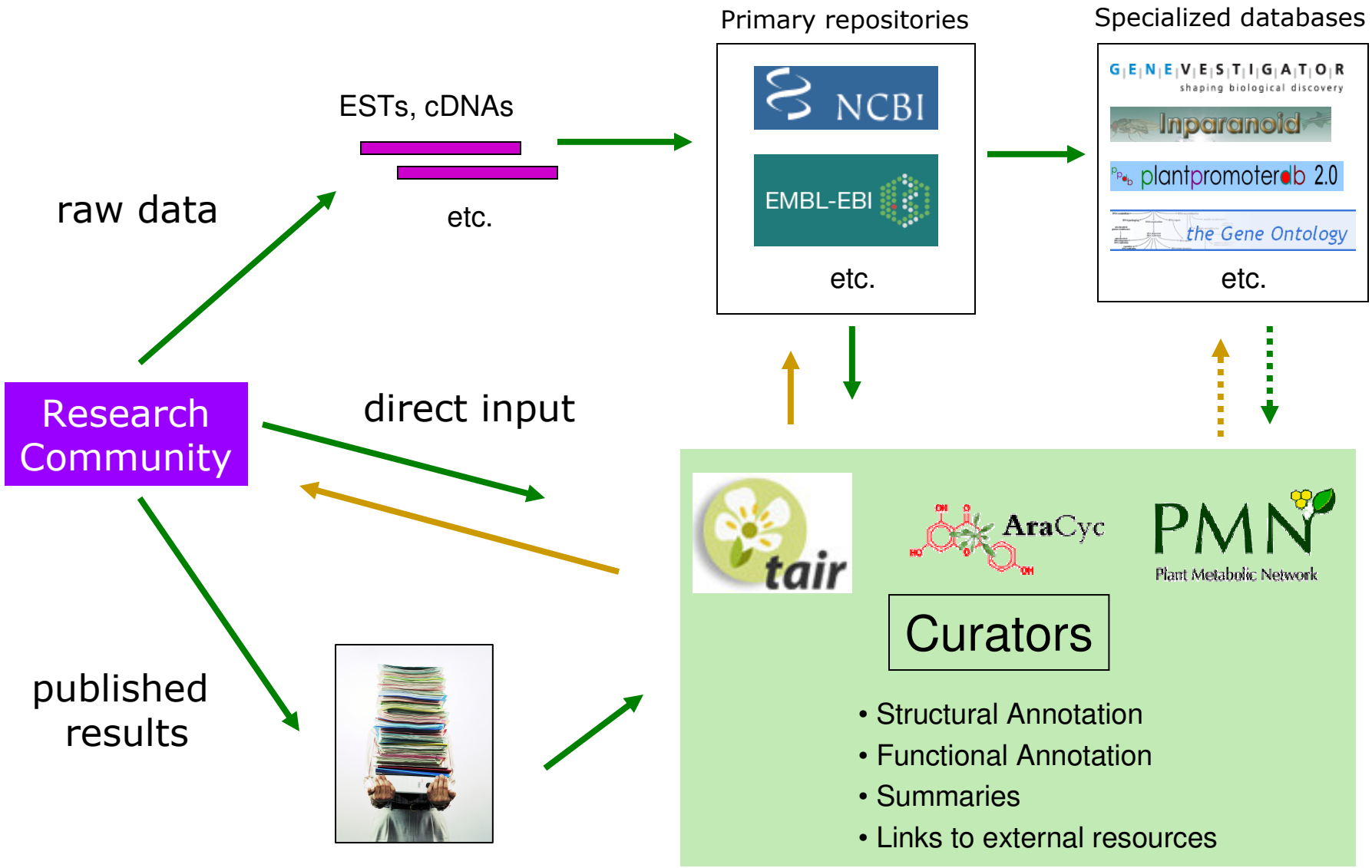
Biological databases try to help . . .

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

- AraCyc – **Ara**bidopsis Metabolic En**Cyc**lopedia
  - 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"

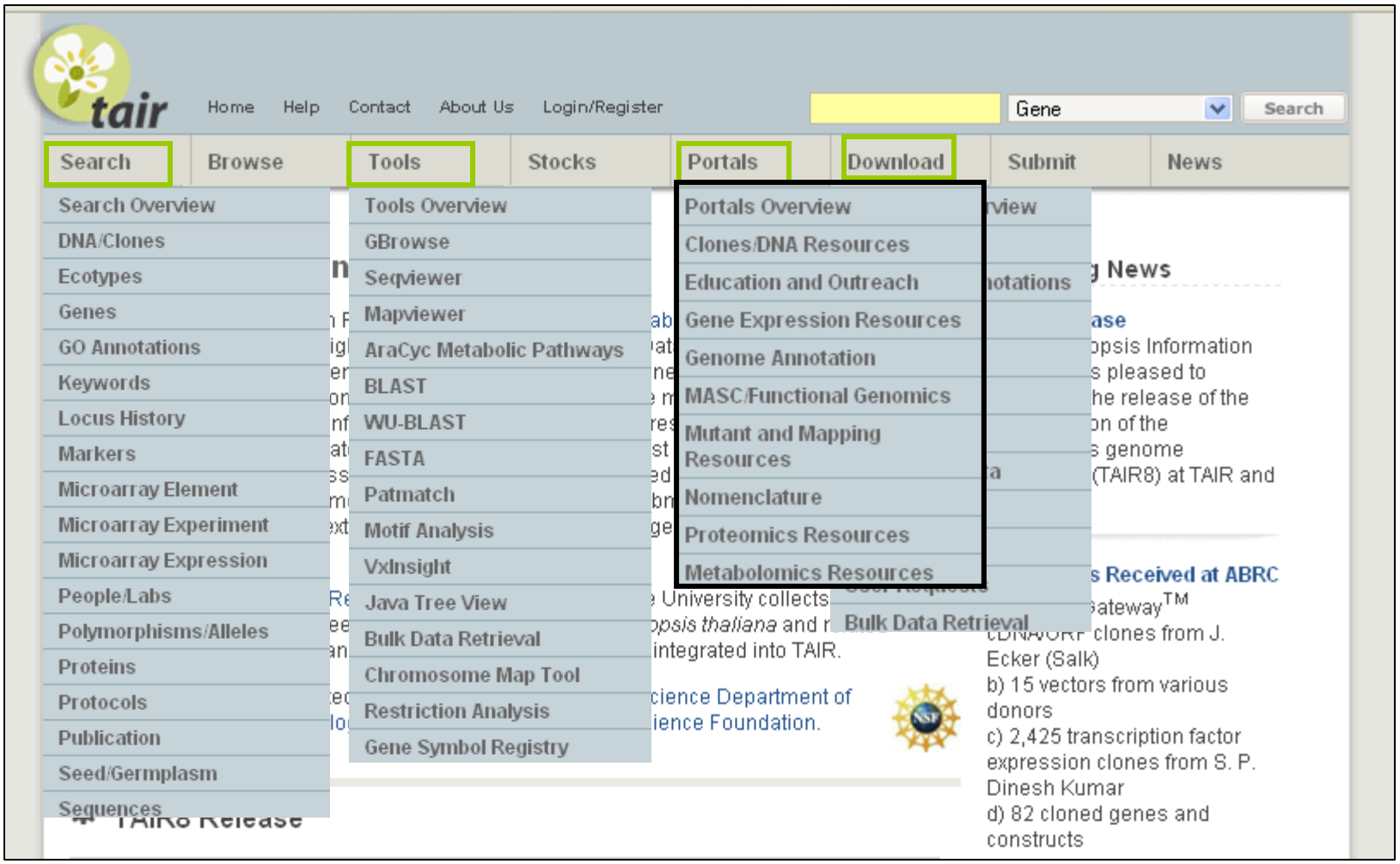
- Connecting scientists to data:
  - Genomes
  - Transcripts
  - Proteins
  - Metabolism
  - Other organisms
- Connecting scientists to resources and tools:
  - DNA stocks / germplasm
  - Analysis and viewing applications, etc.
- Connecting scientists to each other
- Connecting scientists to everything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

tair

AraCyc

PMN
Plant Metabolic Network

Connecting scientists to everything . . .



tair

AraCyc

PMN
Plant Metabolic Network

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

tair

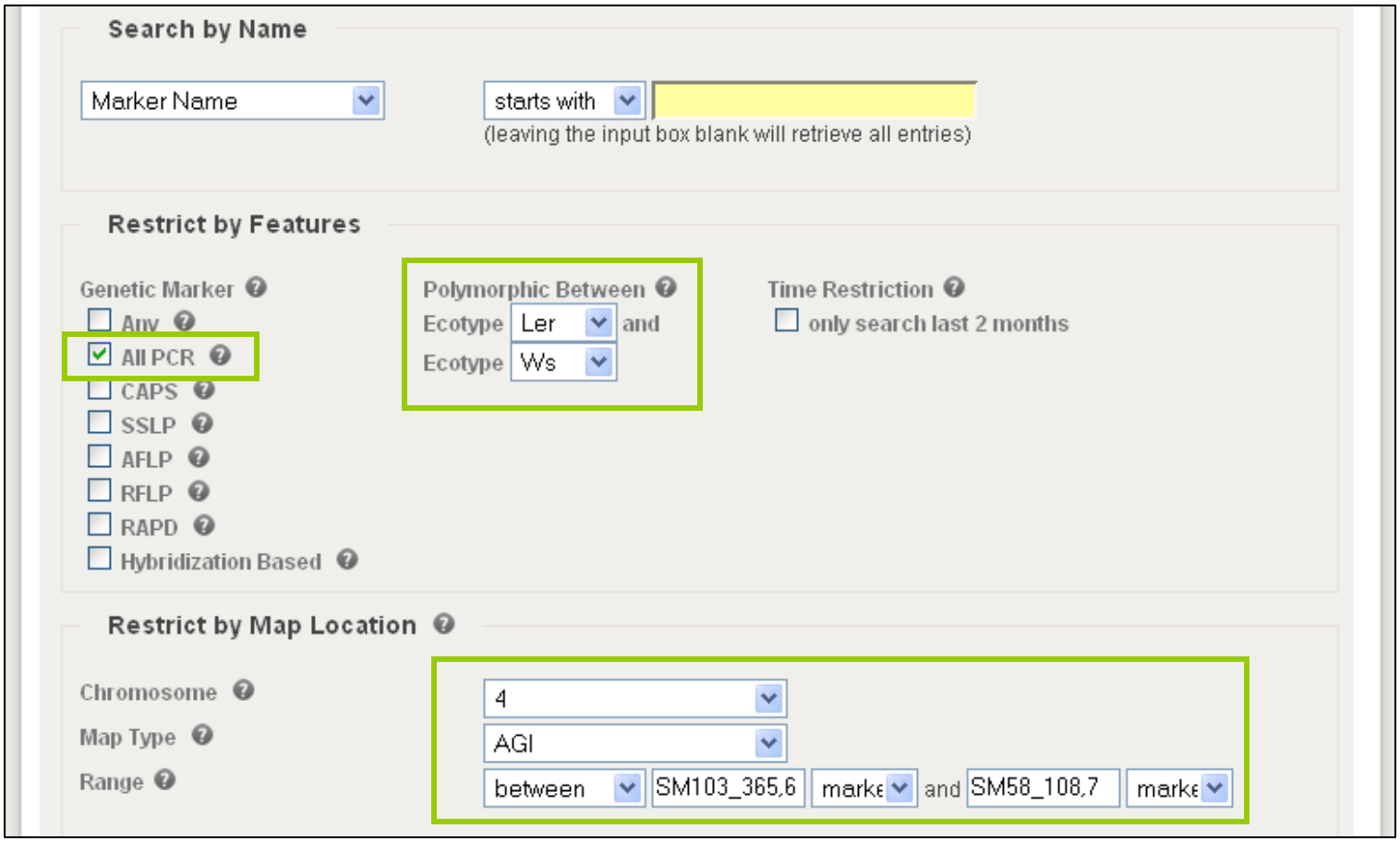
AraCyc

PMN
Plant Metabolic Network

Finding the mutation

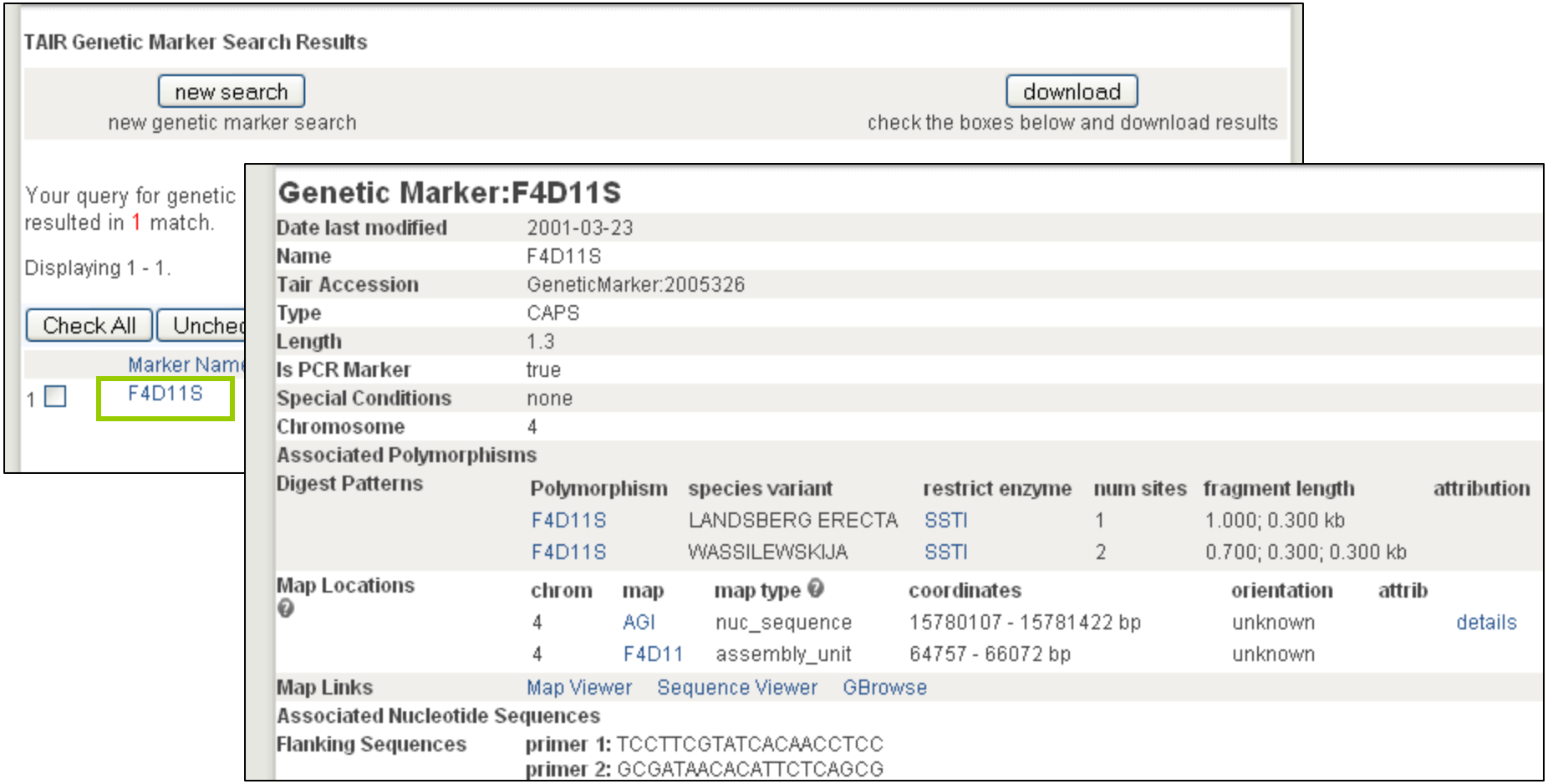
Finding the mutation

- **Narrowing the mapping interval**
  - Search for additional markers in the region using the Marker search page



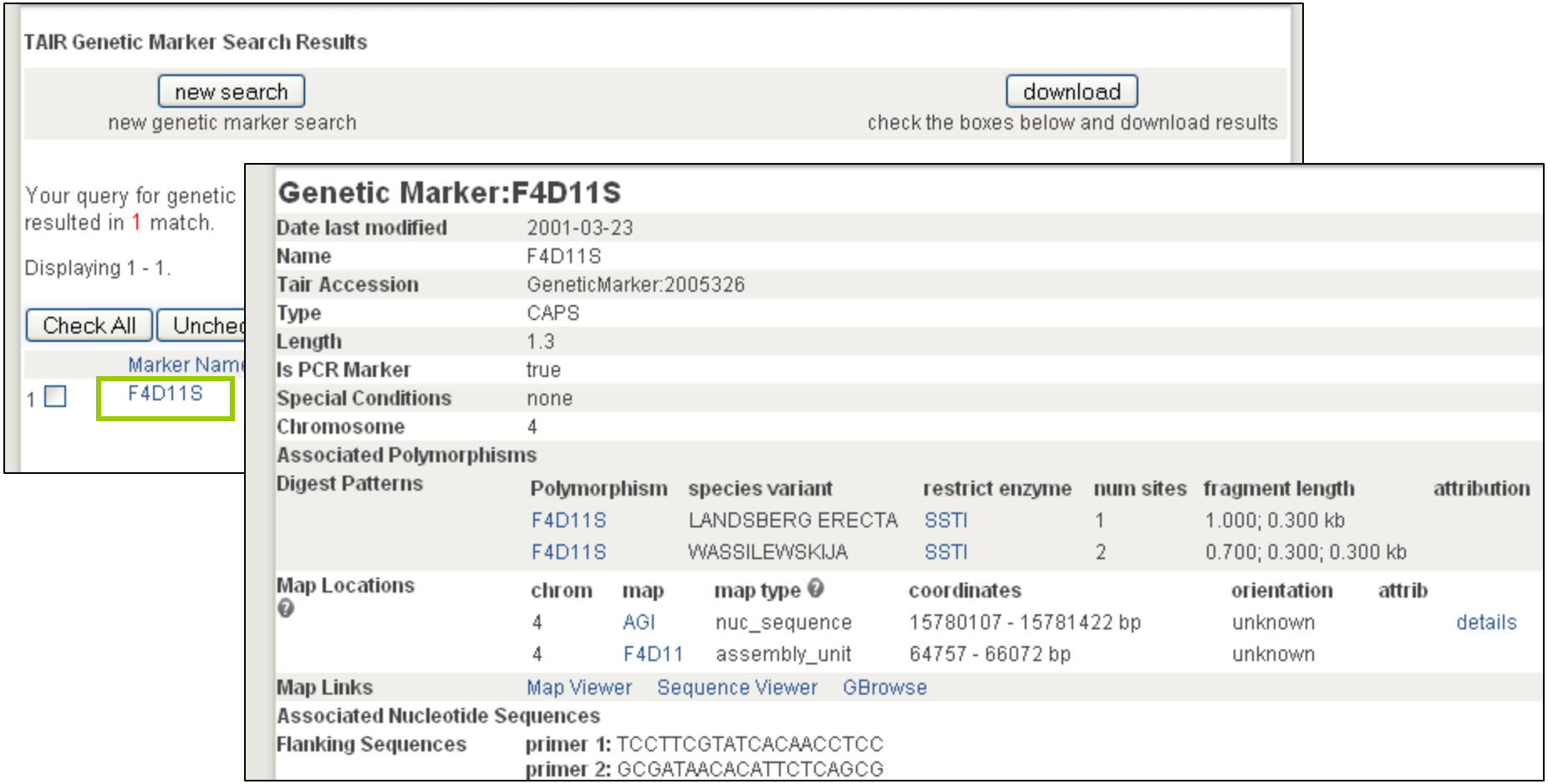
Finding the mutation

- **Search for additional markers in the region**
  - Use Marker search page



Finding the mutation

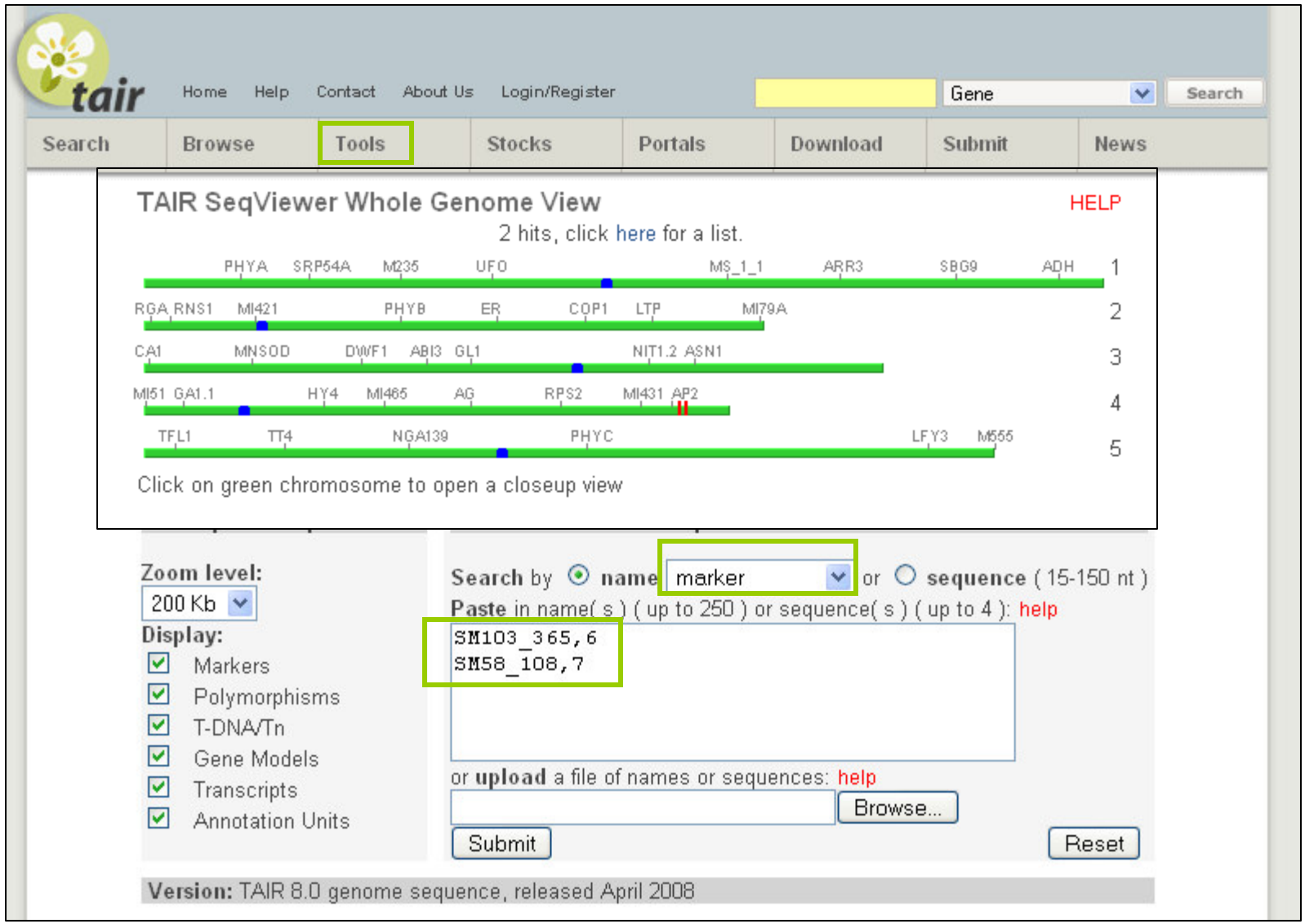
- **Search for additional markers in the region**
  - Use Marker search page



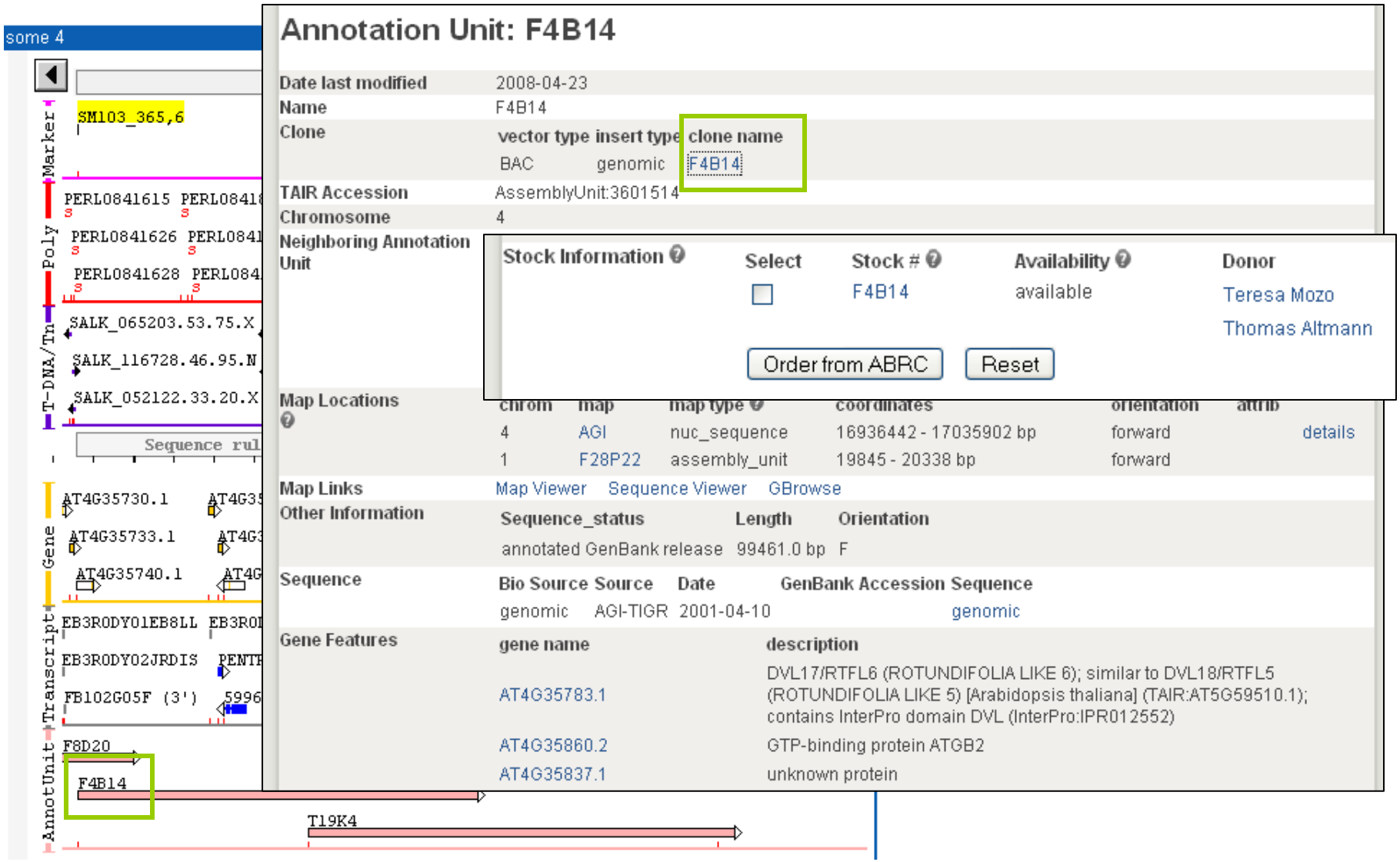
Finding the mutation

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

Finding the mutation (Seqviewer)



Finding the mutation (Seqviewer)



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

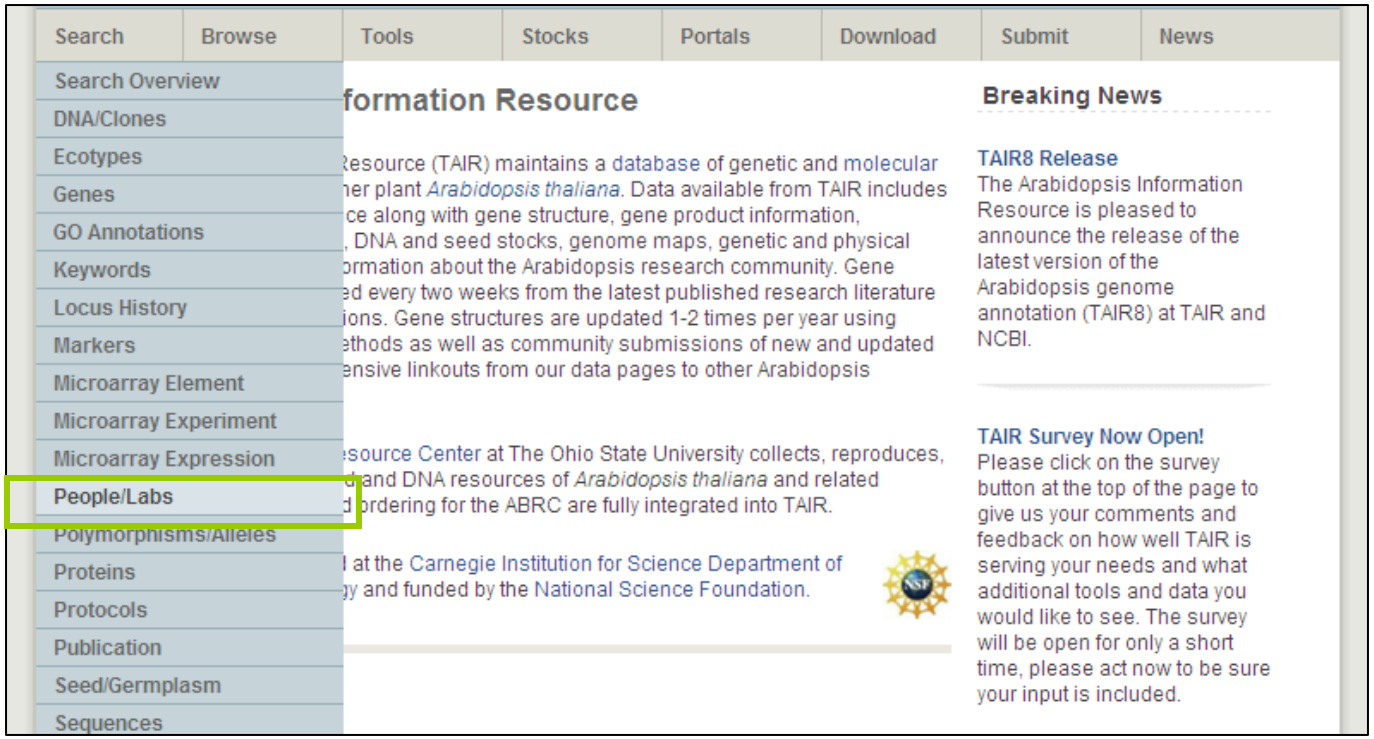
tair

AraCyc

PMN
Plant Metabolic Network

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 . . .



tair

AraCyc

PMN
Plant Metabolic Network

Finding the mutation

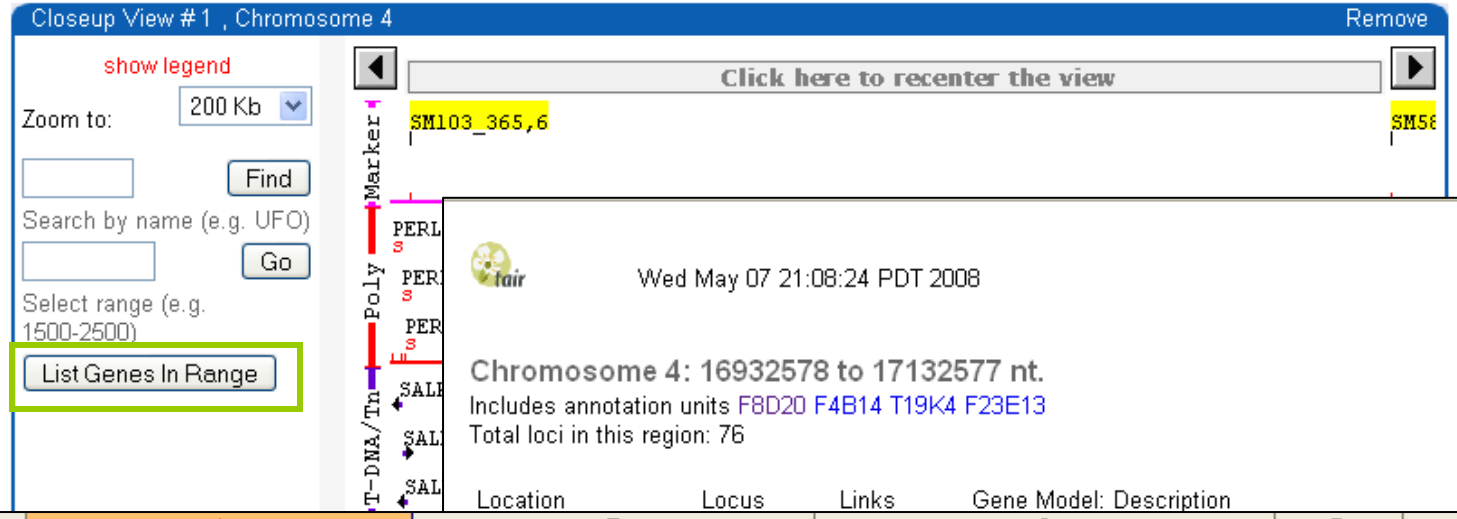
- **Looking for candidate genes**
  - Identify genes in the region using Seqviewer

tair

AraCyc

PMN
Plant Metabolic Network

Finding the mutation (Seqviewer)

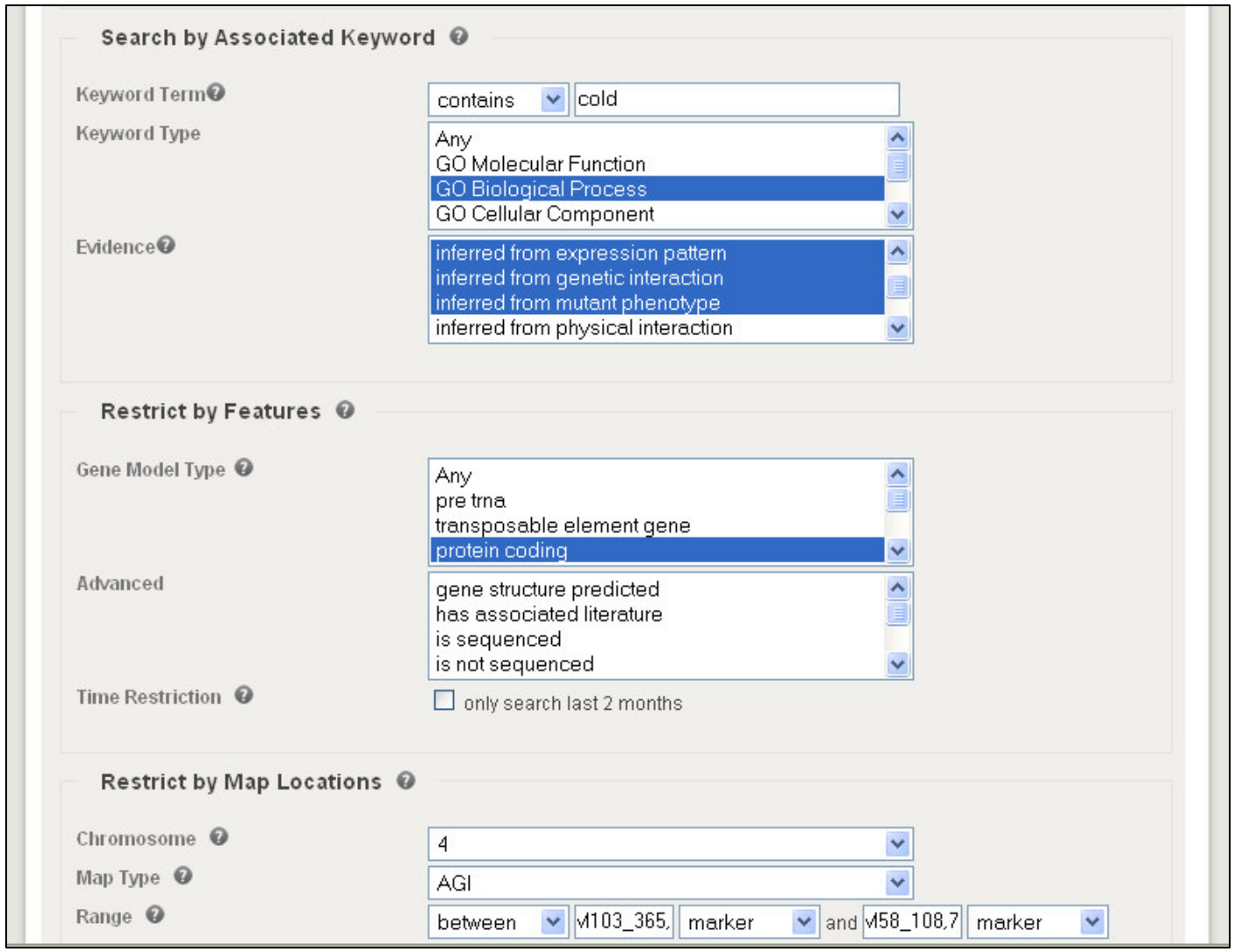


	A	B	C	D	E	F	G
1	TAIR SeqViewer	Wed May 07 21:08:32 PDT 2008					
2	Chromosome 4: 16932578 to 17132577 nt.						
3	Includes annotation units F8D20 F4B14 T19K4 F23E13						
4	Total loci in this region: 76						
5							
6	Location	Locus	Gene Model: Description				
7	16931044-16933361	AT4G35730.1	AT4G35730.1:similar 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 [Ara				
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 starvatio				
14	16944946-16946197	AT4G35770.2	SEN1:Senescence-associated gene that is strongly induced by phosphate starvatio				
15	16944946-16946197	AT4G35770.3	SEN1:Senescence-associated gene that is strongly induced by phosphate starvatio				
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)



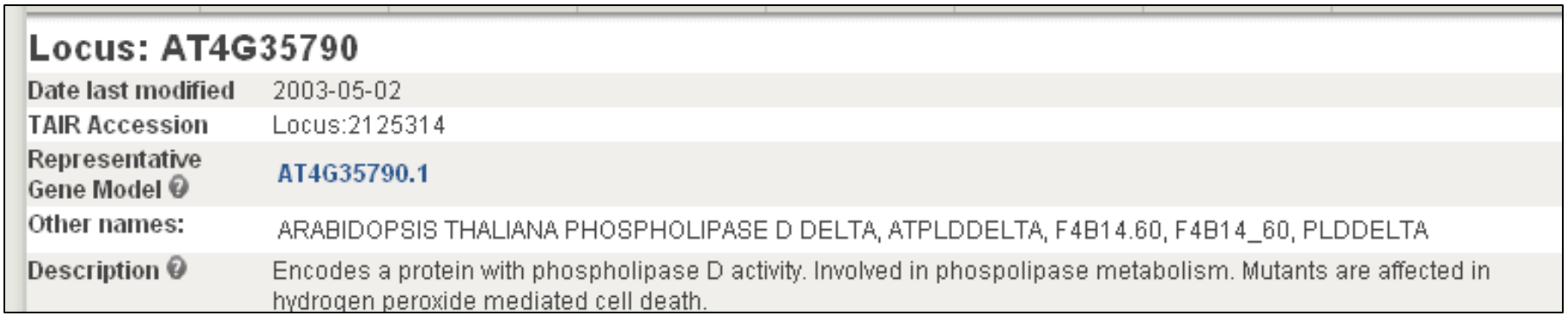
Finding the mutation (Gene search page)

	Locus	Description	Gene Model(s)	Other Names	Keywords
1 ☐	AT4G35790	Encodes a protein with phospholipase D activity. Involved in phospolipase 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 ☐	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

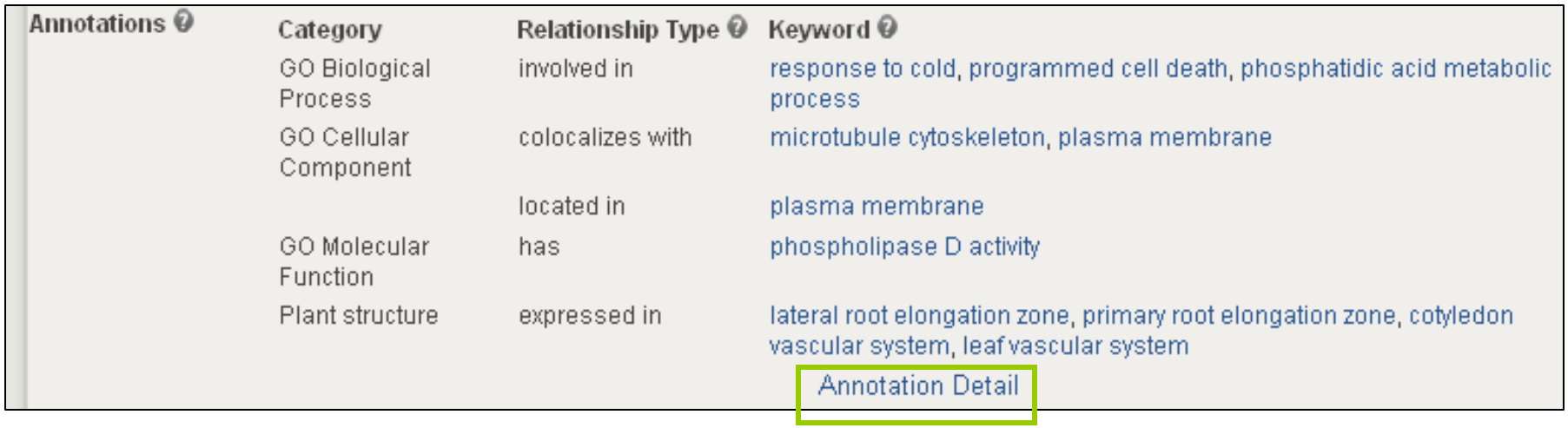


Locus: AT4G35790

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 phospolipase 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

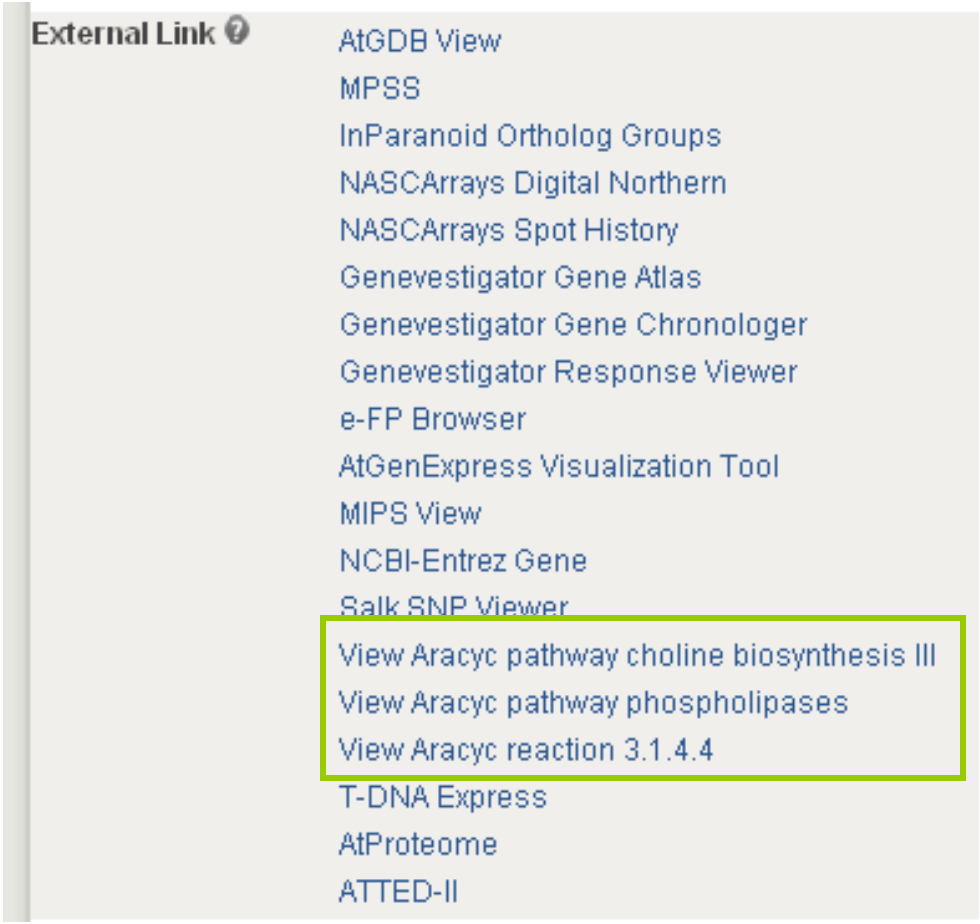
Getting acquainted with the candidates

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

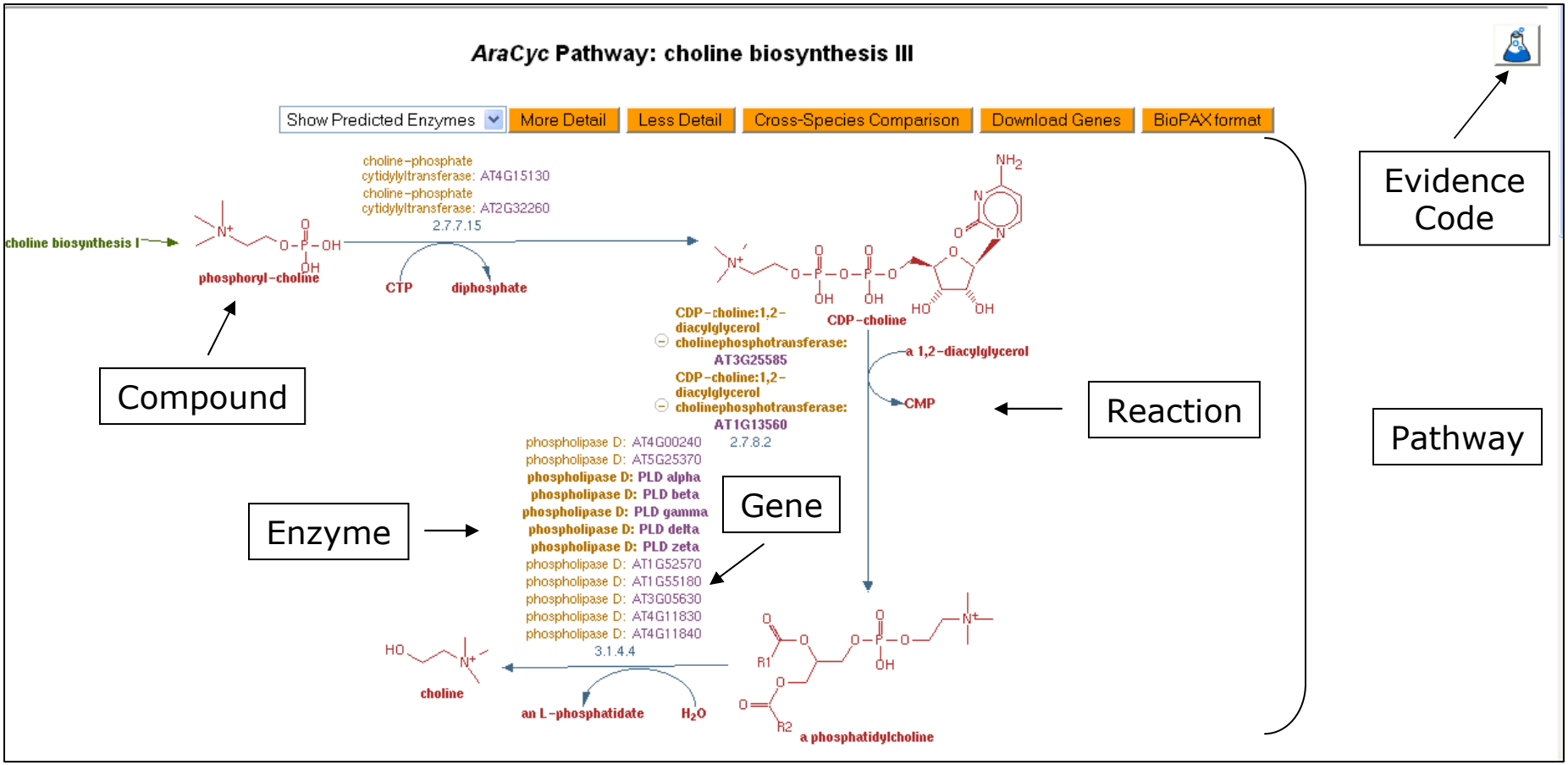
Getting acquainted with the candidates

- 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 cytidylyltransferase 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.

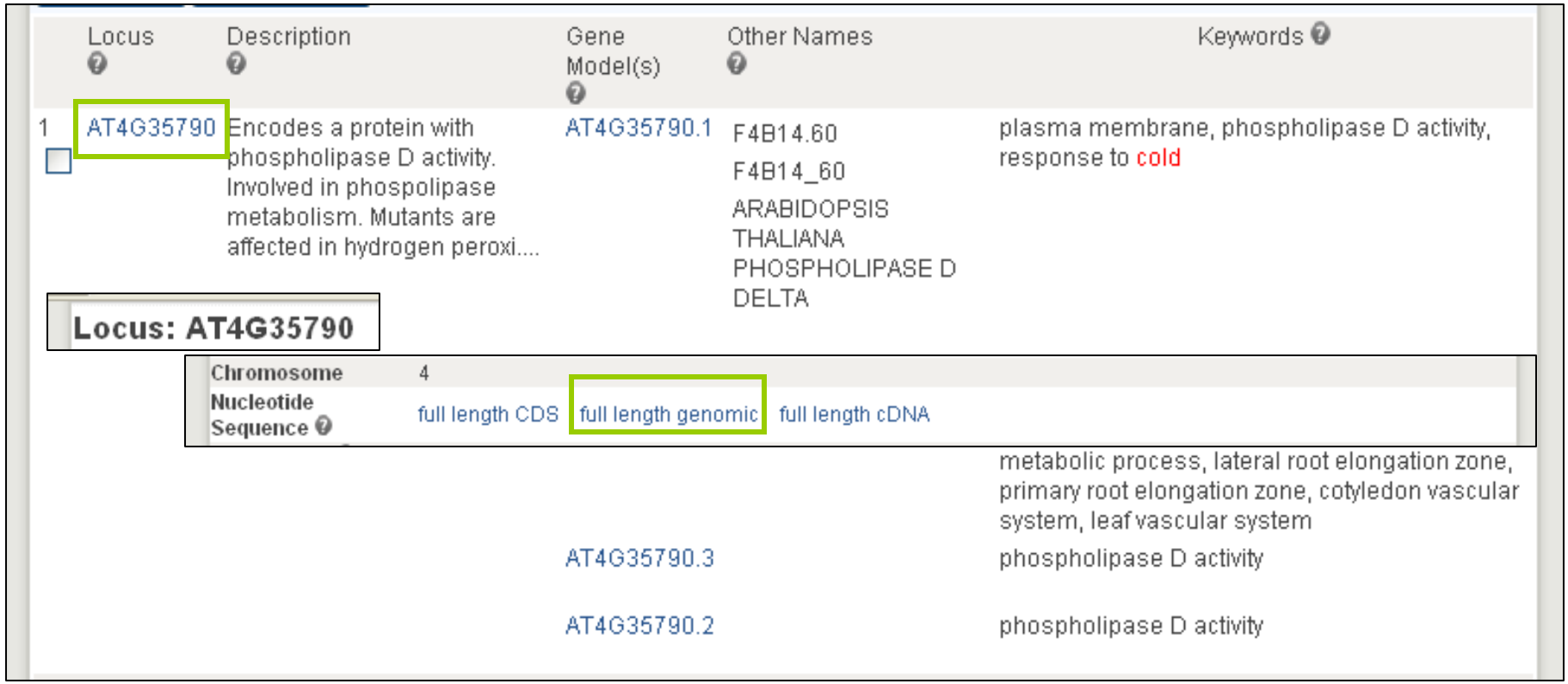
Datko88a: Datko AH, Mudd SH, (1988) "Phosphatidylcholine synthesis. Differing patterns in soybean and carrot." Plant Physiol. (1988), 88, 854-861.

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)

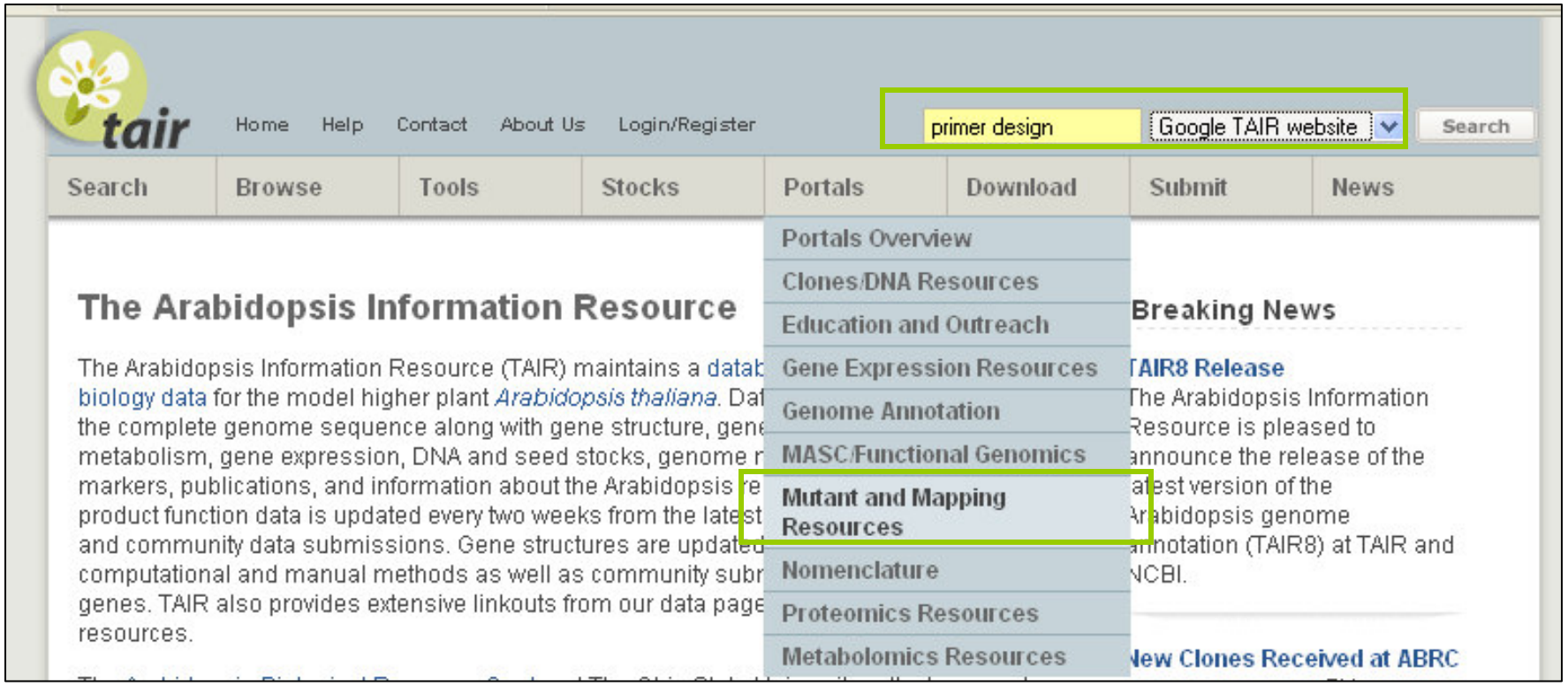
Confirming the candidate gene(s)

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



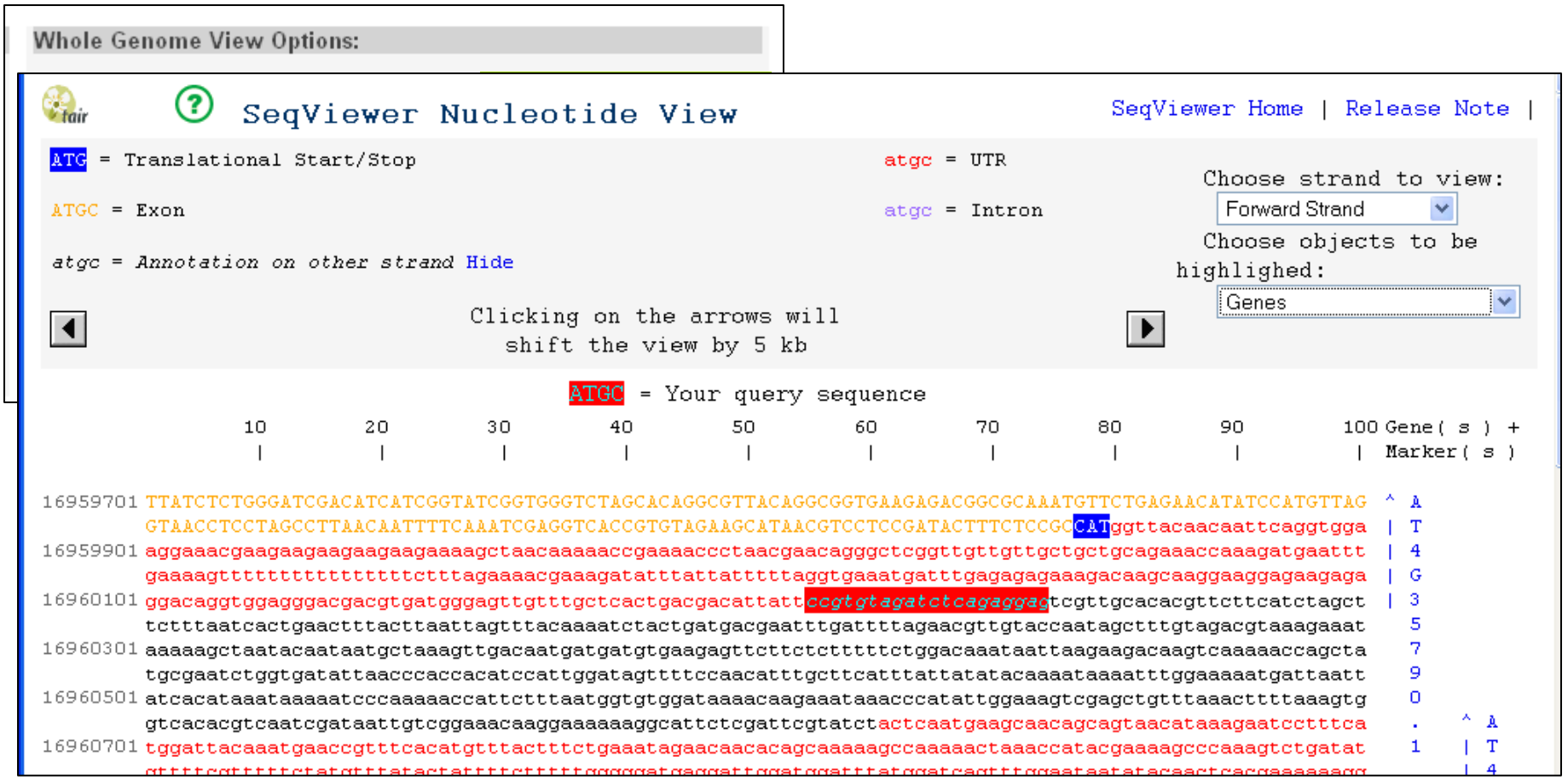
Confirming the candidate gene(s)

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



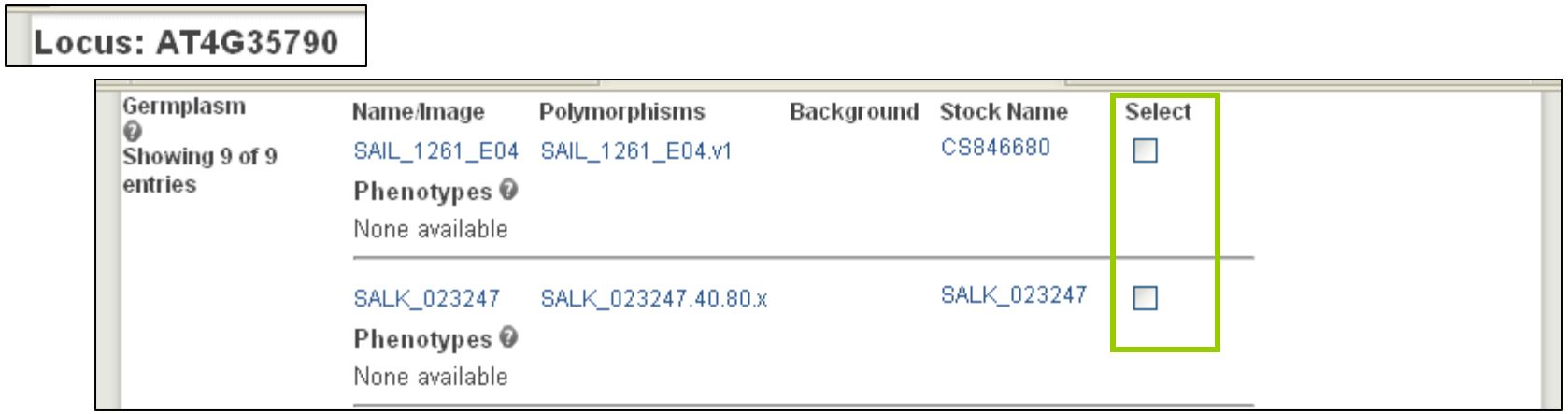
Confirming the candidate gene(s)

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



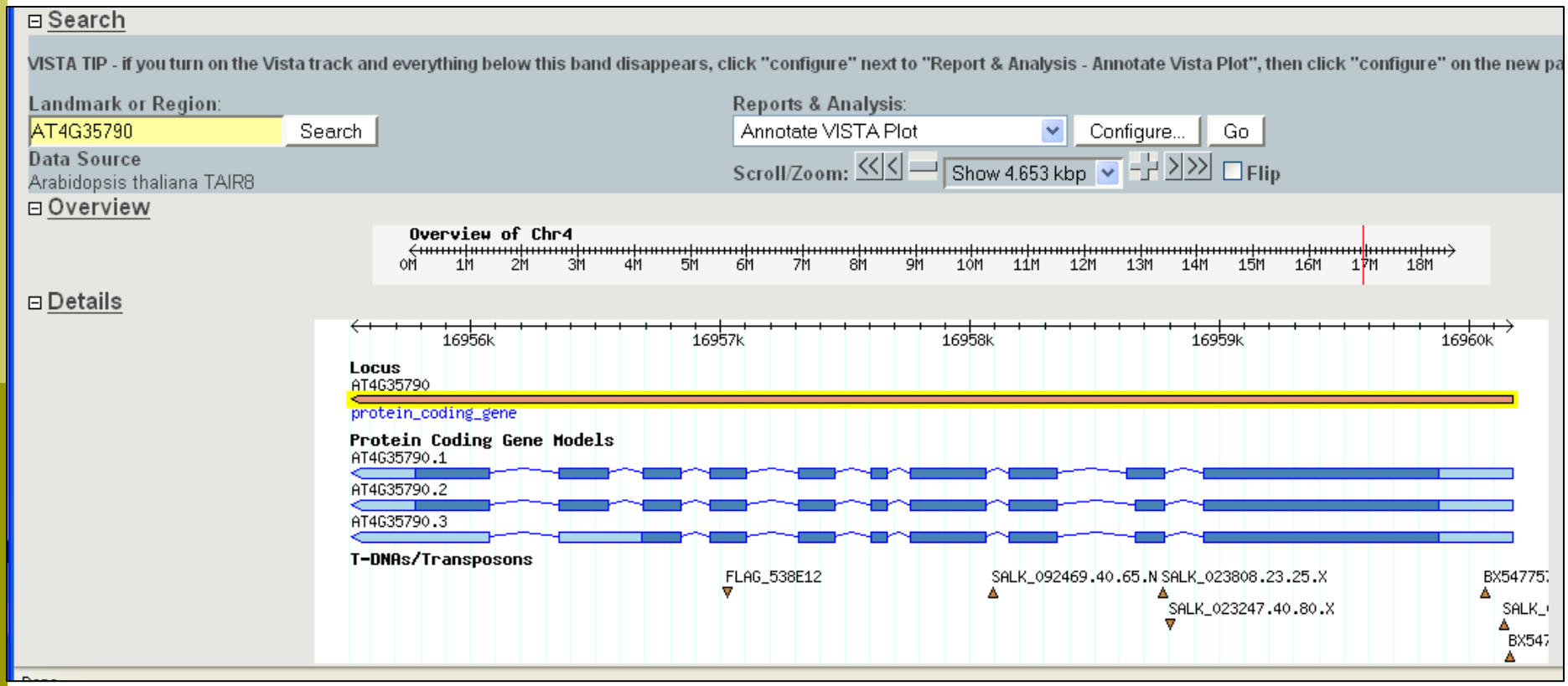
Confirming the candidate gene(s)

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



Confirming the candidate gene(s)

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

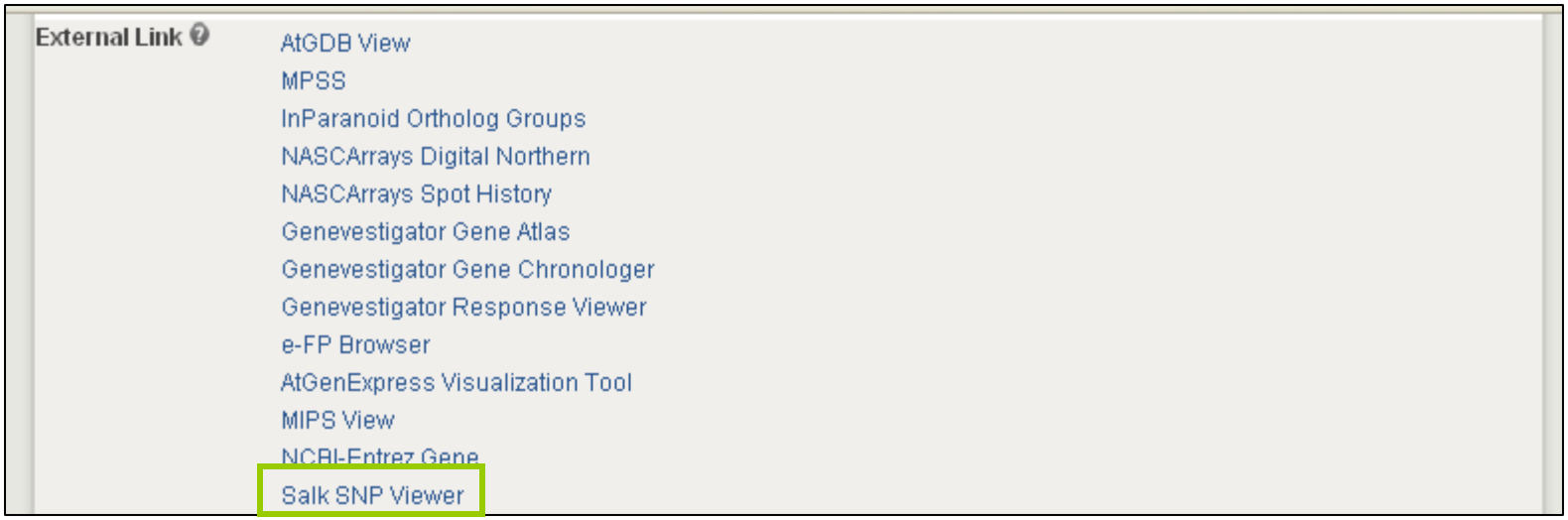


Confirming the candidate gene(s)

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

Confirming the candidate gene(s)

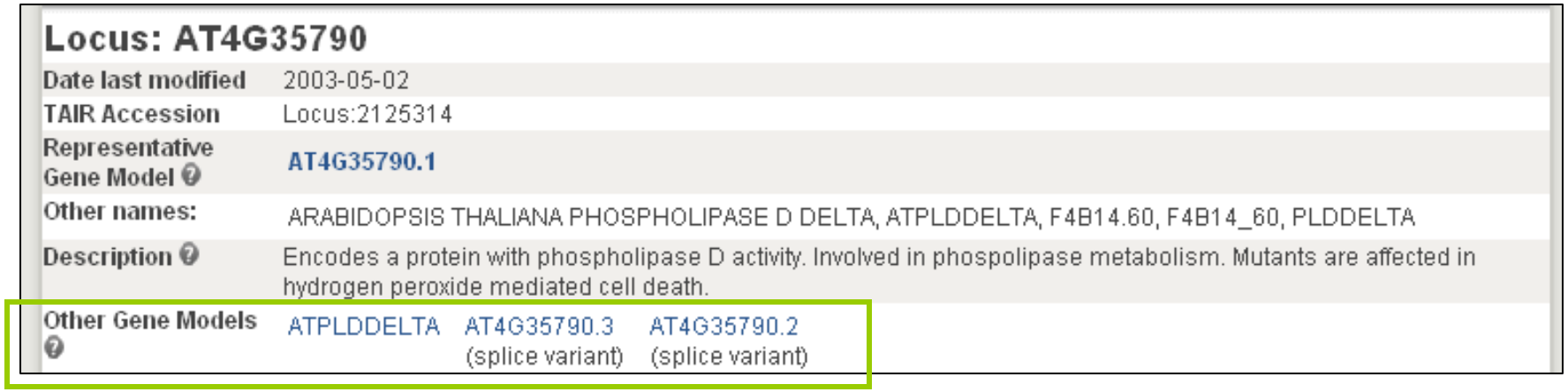
- Obtain additional polymorphic lines
  - Find natural variations using SALK SNP viewer from Locus page



  - 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



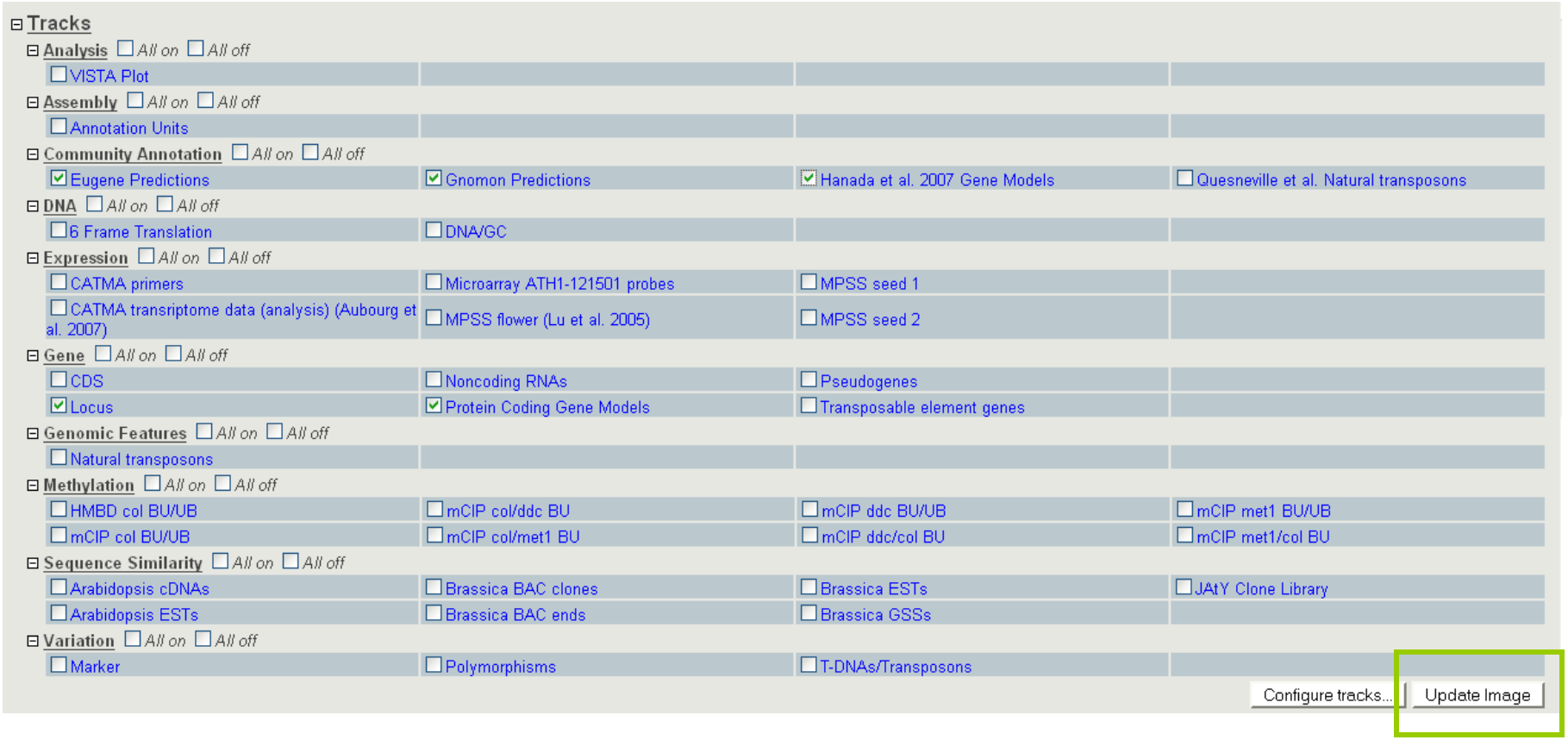
Locus: AT4G35790

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 phospolipase metabolism. Mutants are affected in hydrogen peroxide mediated cell death.
Other Gene Models	ATPLDDELTA AT4G35790.3 (splice variant) AT4G35790.2 (splice variant)

  - 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



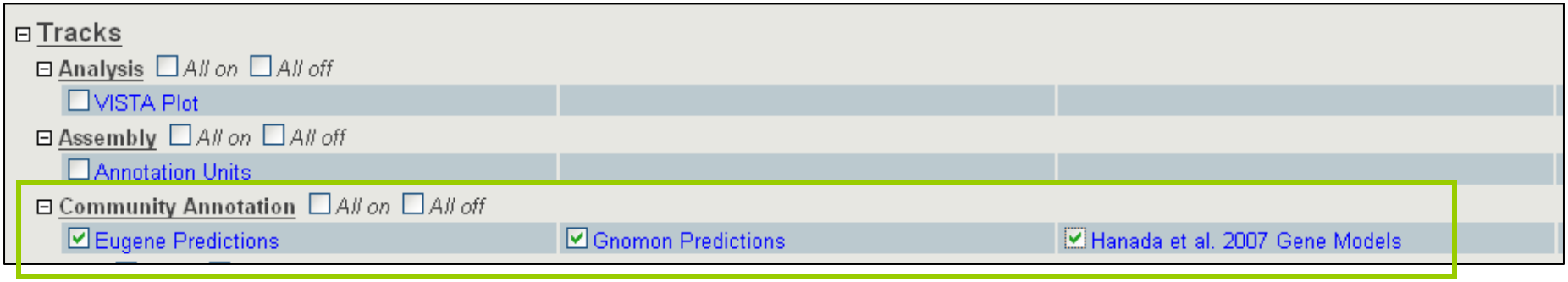
Tracks

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

Configure tracks... | Update Image

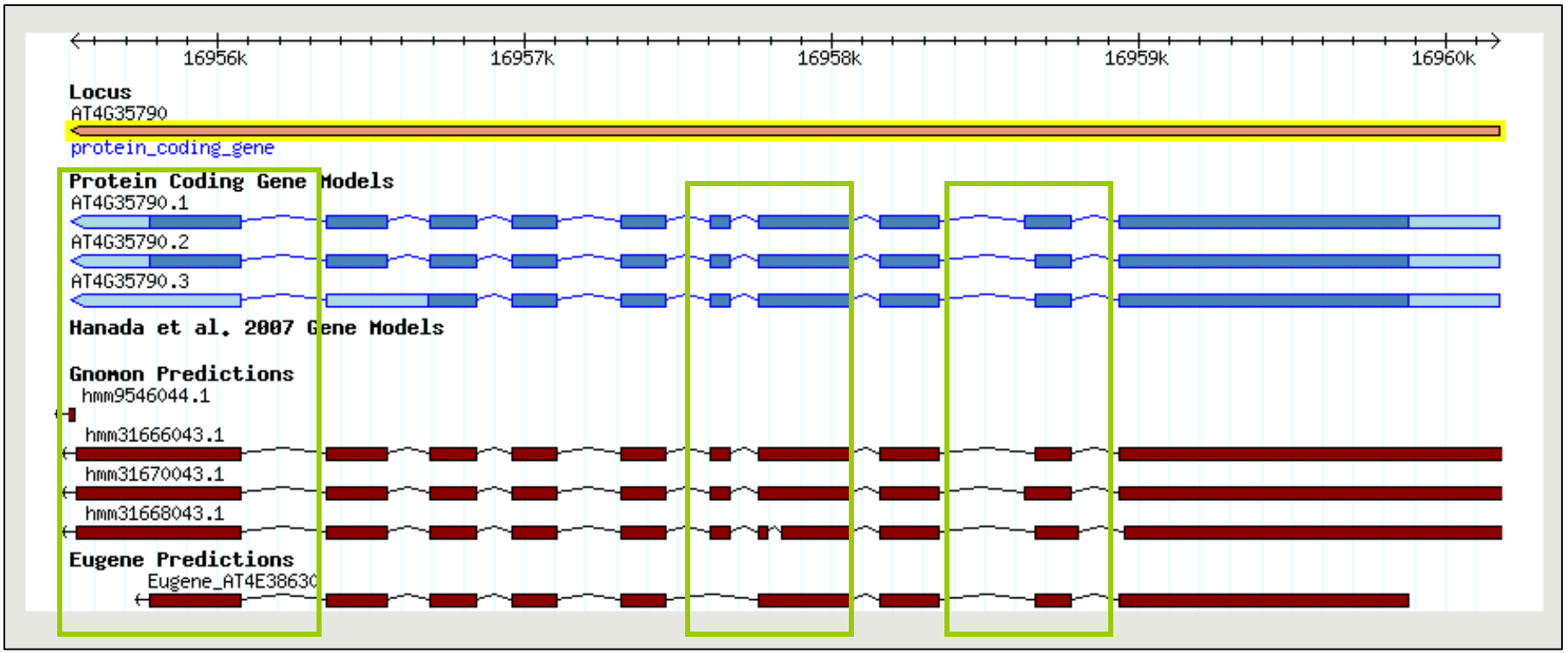
Confirming the candidate gene(s)

- Generate constructs for complementation
  - Check other gene prediction programs



Confirming the candidate gene(s)

- Generate constructs for complementation
  - Check other gene prediction programs

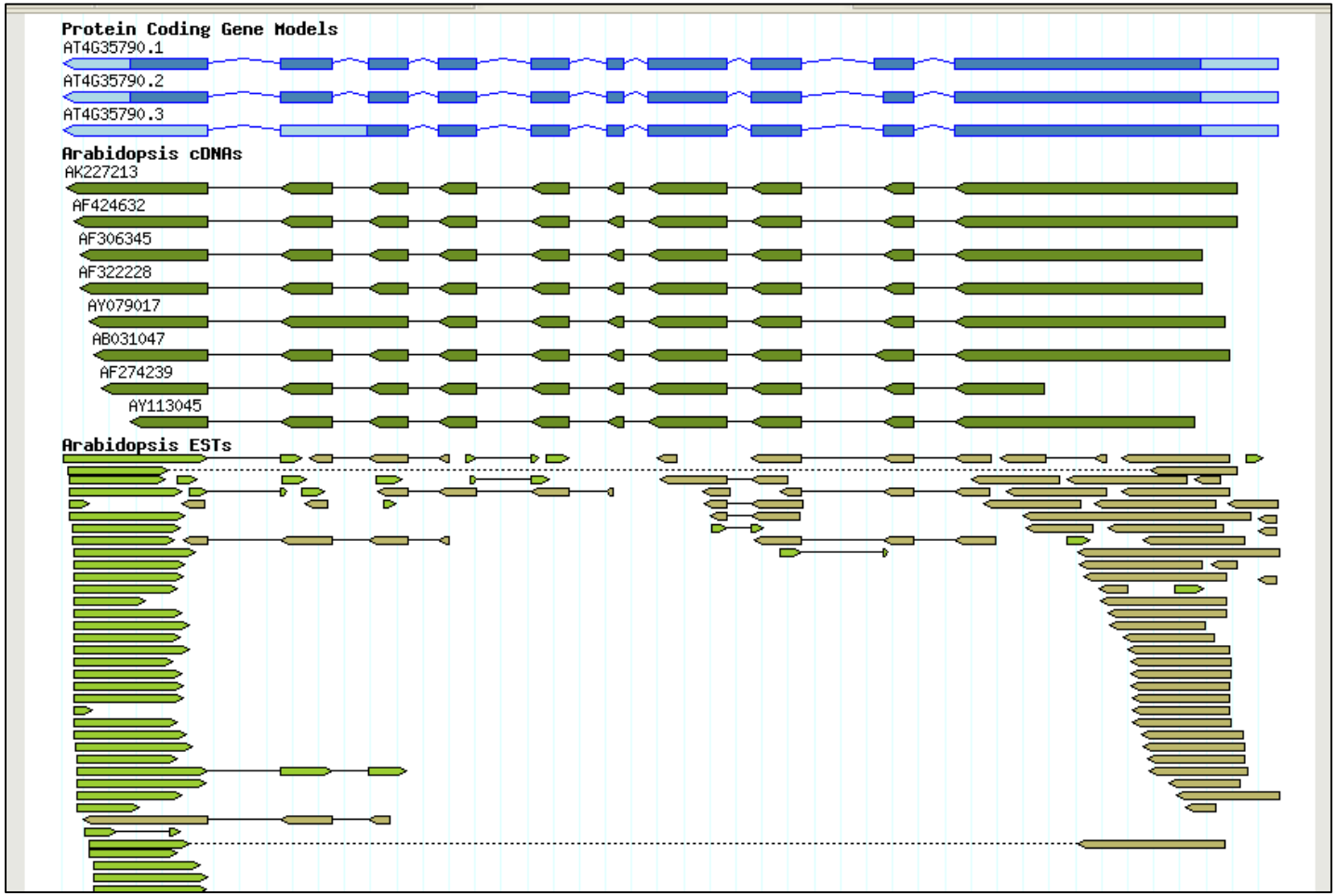


Confirming the candidate gene(s)

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

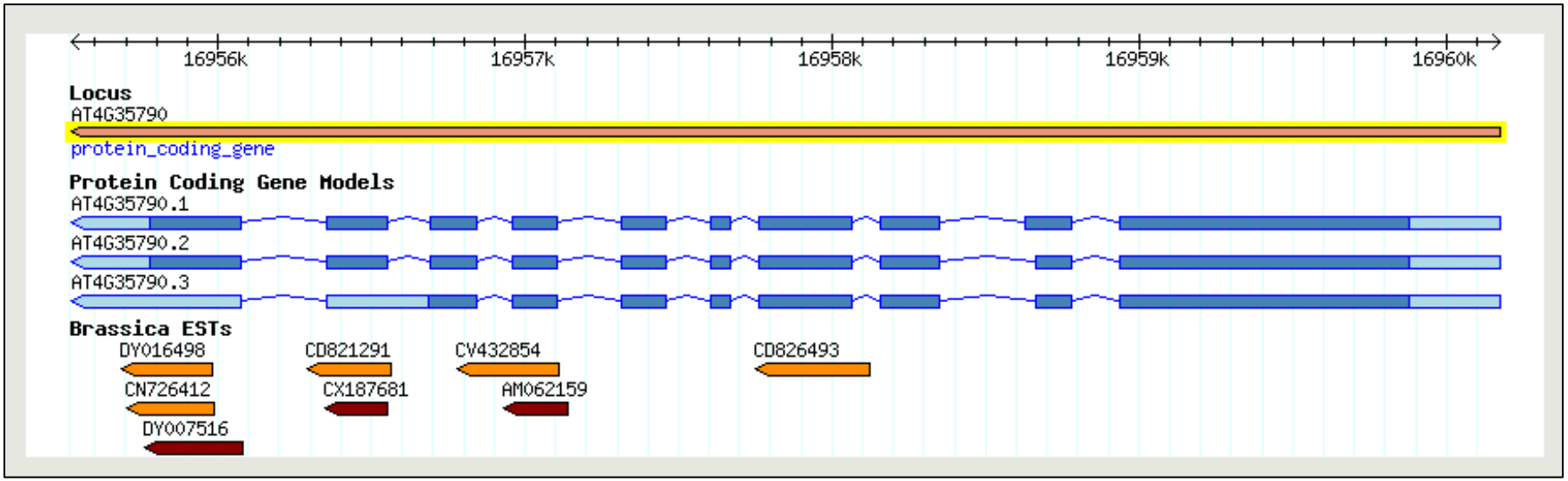
cDNAs

ESTs



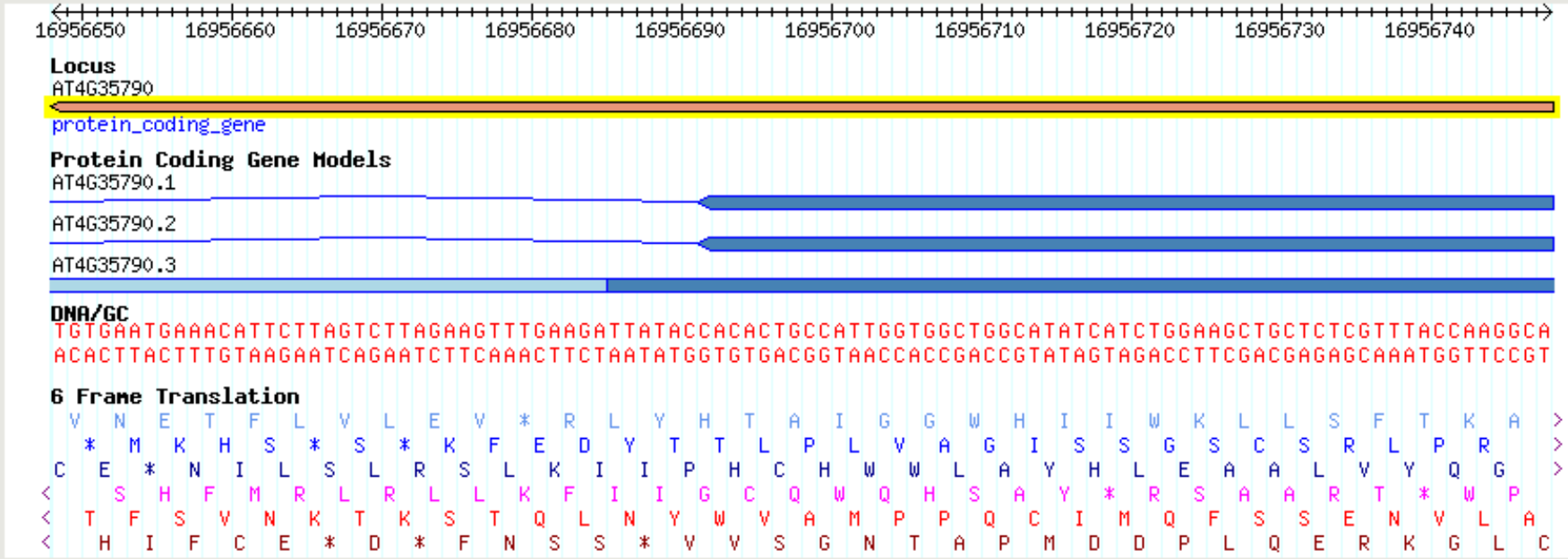
Confirming the candidate gene(s)

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



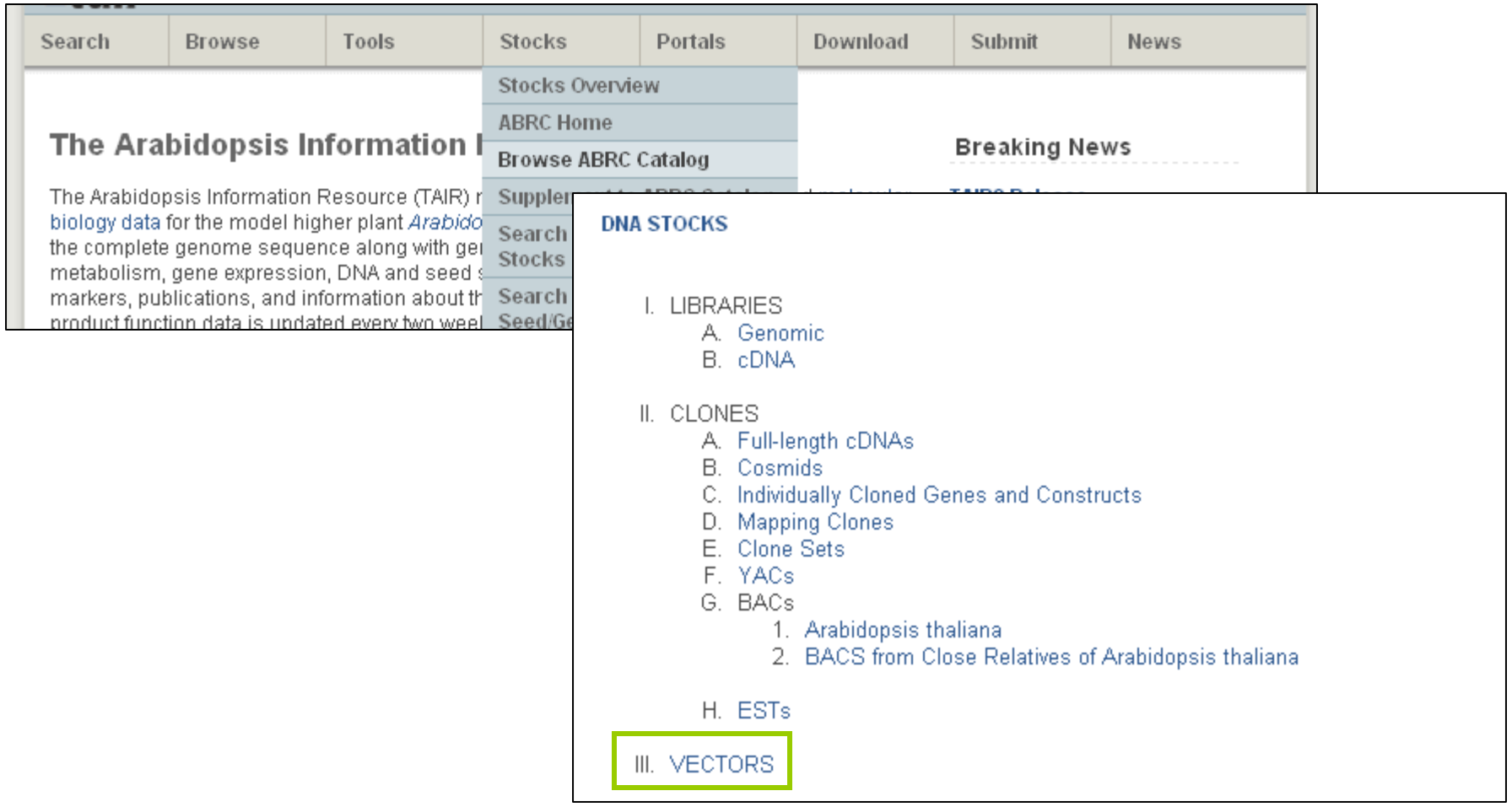
Confirming the candidate gene(s)

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



Confirming the candidate gene(s)

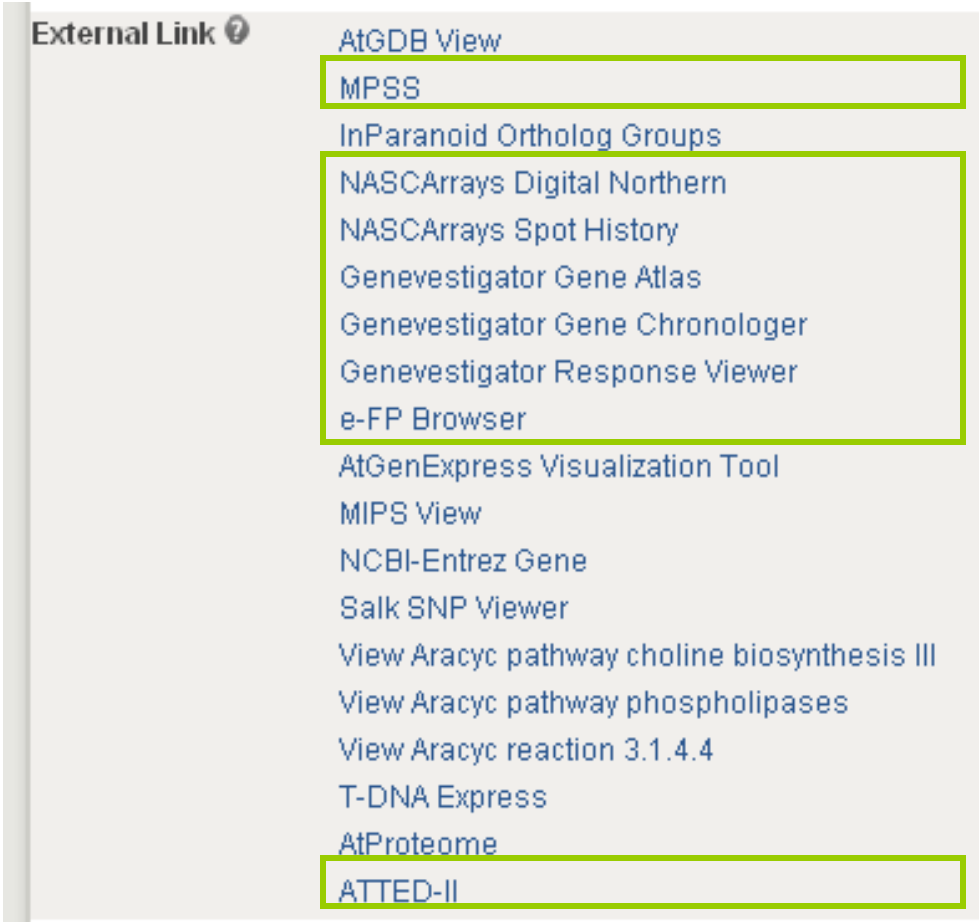
- Generate constructs for complementation
  - Browse ABRC catalog for expression vectors



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 Piriformospora indica in Arabidopsis thaliana.	THE PLANT JOURNAL	AT4G35790.1 AT4G35790.3 AT4G35790.2	2007
Early PLD{alpha}-mediated events in response to progressive drought stress in Arabidopsis: a transcriptome analysis.	JOURNAL OF EXPERIMENTAL BOTANY	ATPLDDELTA	2007
Mapping the Arabidopsis 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 Arabidopsis 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	ATPLDDELTA	2003
The Oleate-Stimulated Phospholipase D, PLD{delta}, and Phosphatidic Acid Decrease H 2O2-Induced Cell Death in Arabidopsis.	THE PLANT CELL	ATPLDDELTA	2003
Kinetic analysis of Arabidopsis phospholipase Ddelta: Substrate preference and mechanism of activation by Ca2+ and phosphatidylinositol 4, 5-bisphosphate.	JOURNAL OF BIOLOGICAL CHEMISTRY	ATPLDDELTA	2002
Molecular diversity of phospholipase D in angiosperms.	BMC GENOMICS	ATPLDDELTA	2002
A novel phospholipase d of Arabidopsis that is activated by oleic Acid and associated with the plasma membrane.	PLANT PHYSIOLOGY	ATPLDDELTA	2001
Involvement of a novel Arabidopsis phospholipase D, AtPLDdelta, in dehydration-inducible accumulation of phosphatidic acid in stress signalling.	THE PLANT JOURNAL	ATPLDDELTA	2001

View Complete List (10 of 11 displayed)

Characterizing the successful candidate

- Test predicted molecular function

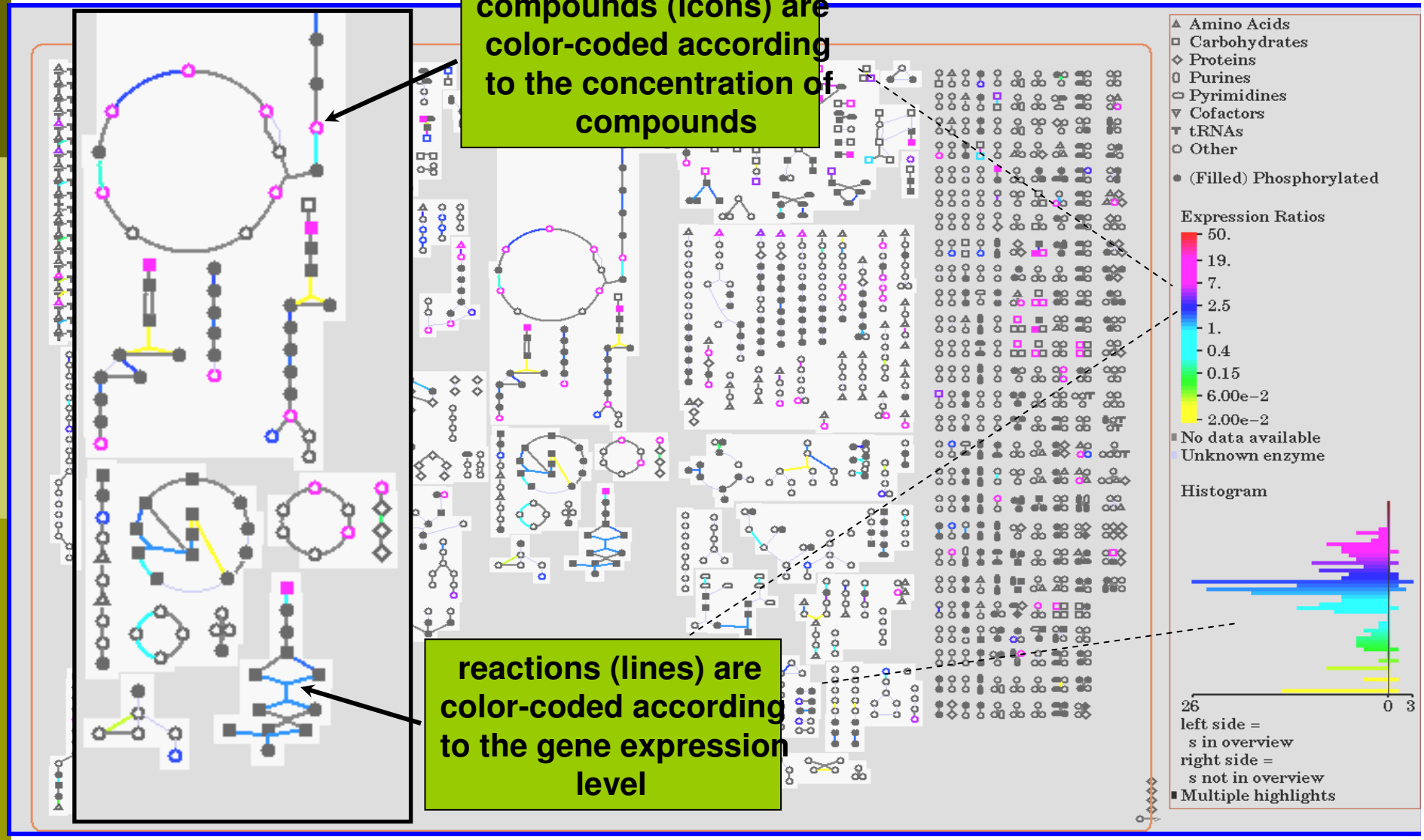
molecular function	has	phospholipase D activity	ATPLDDELTA	*inferred from sequence or structural similarity*: Sequence similarity (homologue of/most closely related to): none: Elias M, et al. (2002)

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

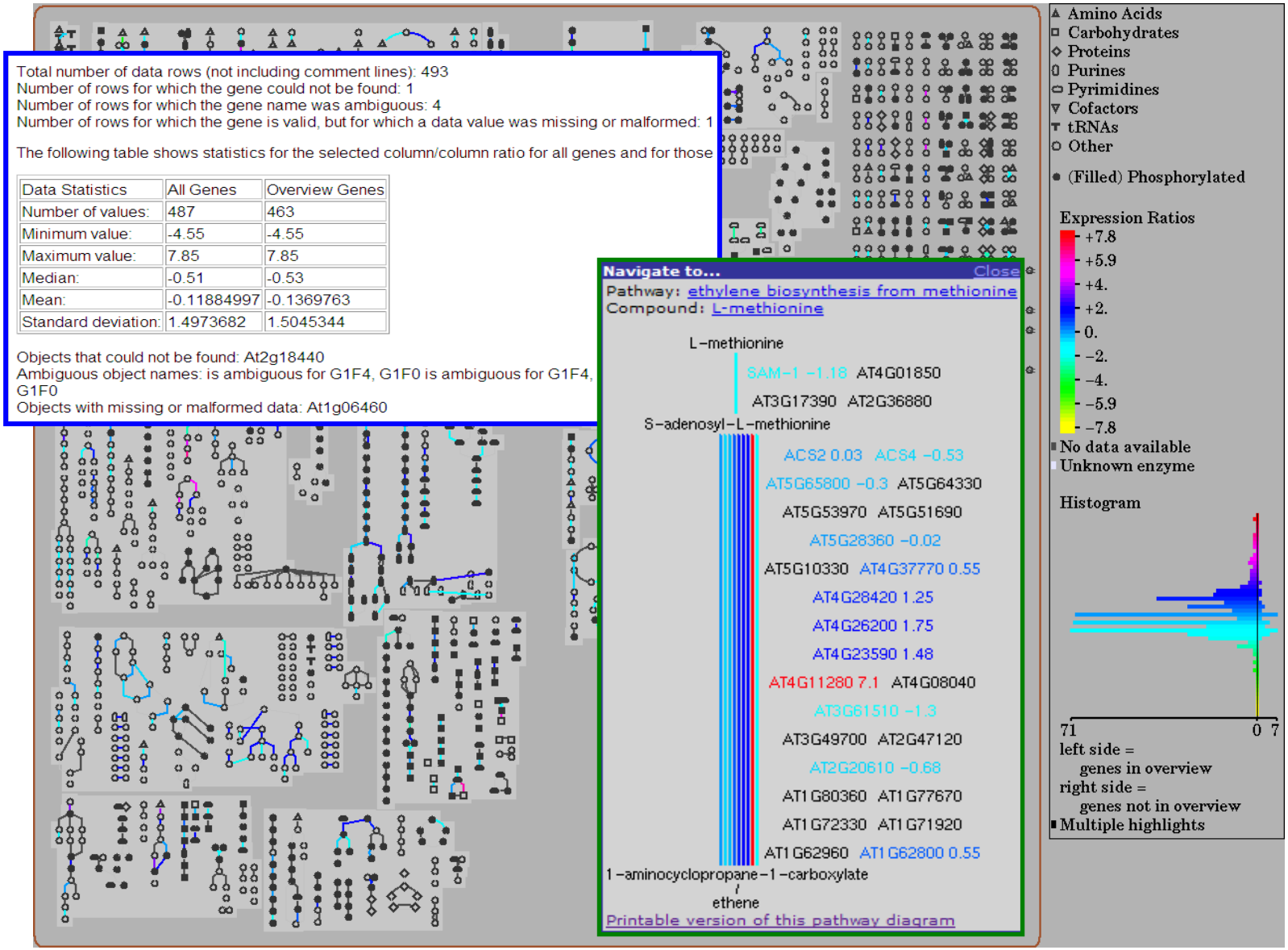
Characterizing the successful candidate

- 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

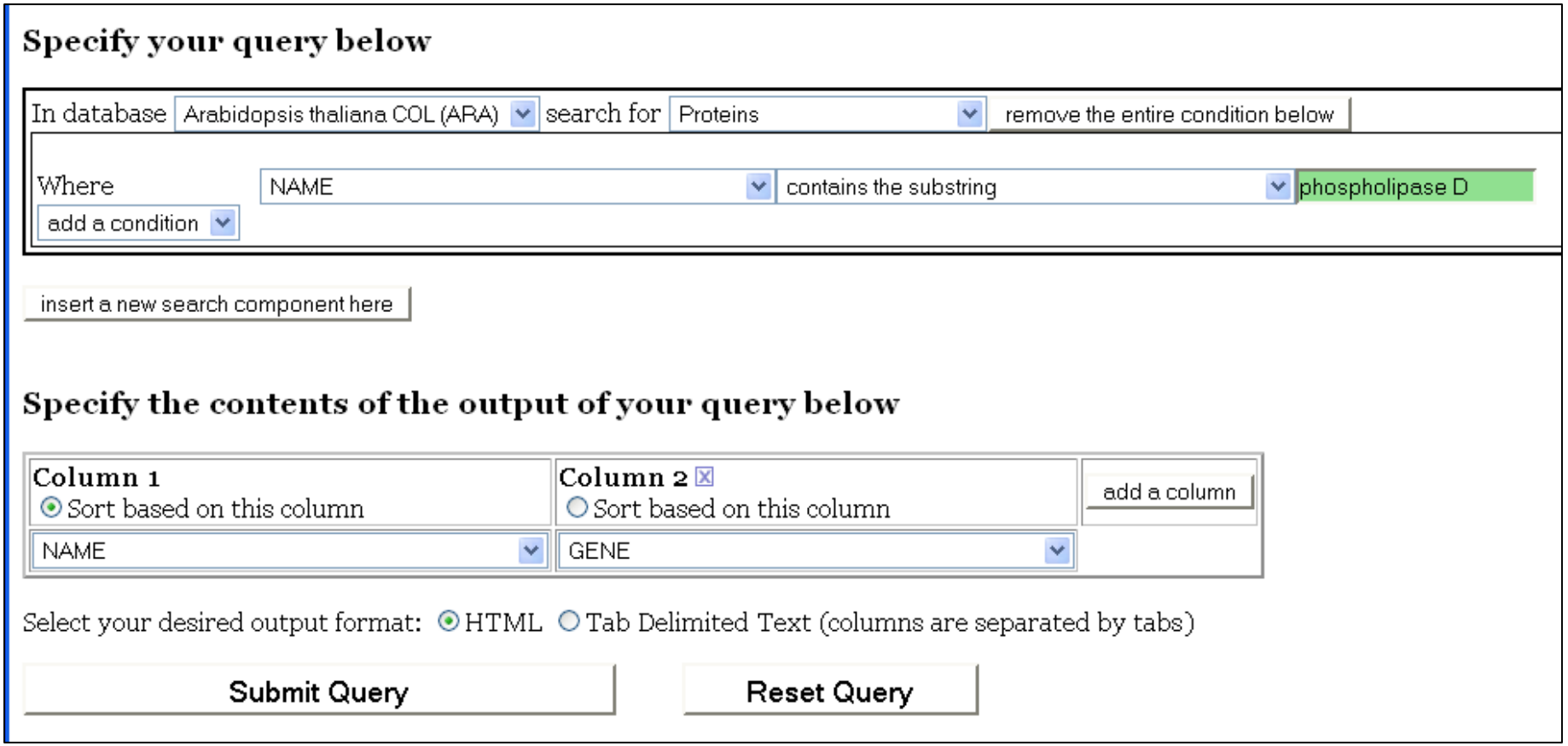


Characterizing the successful candidate

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

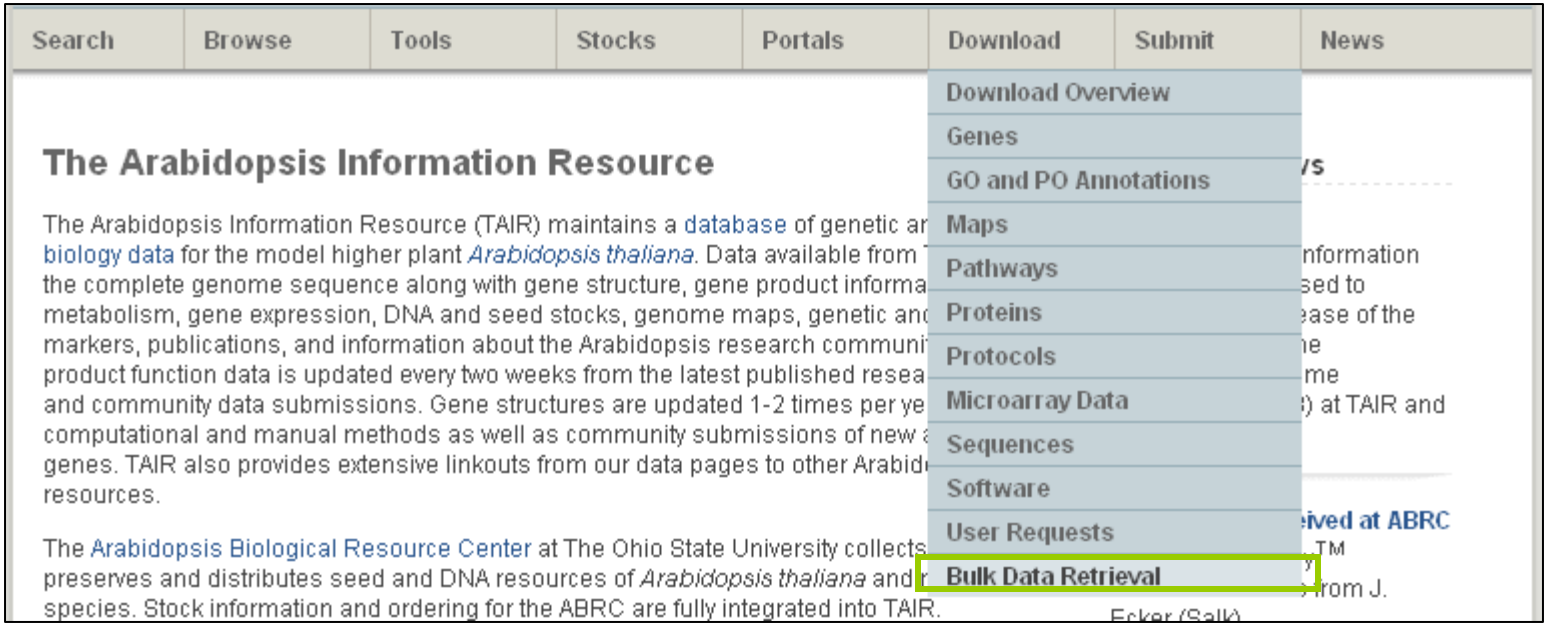
Characterizing the successful candidate

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



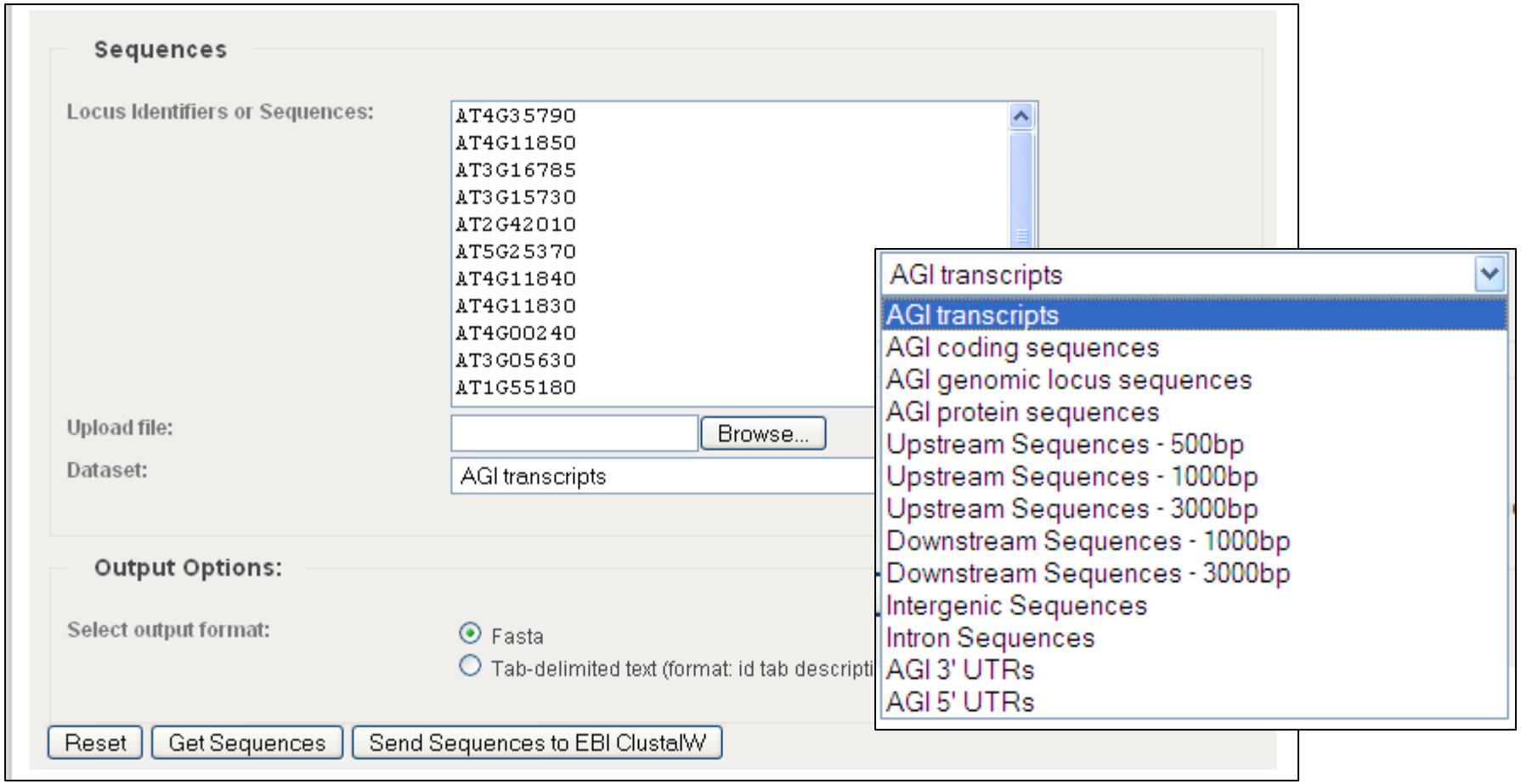
Characterizing the successful candidate

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



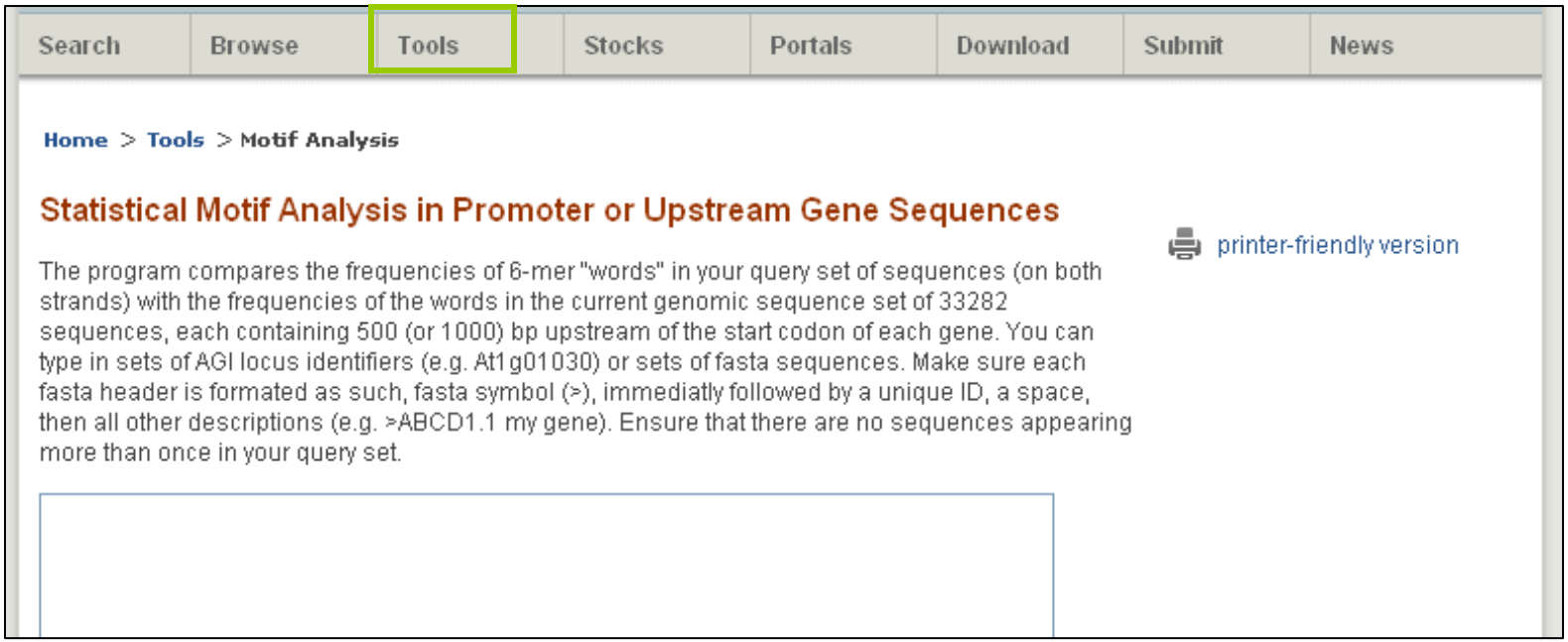
Characterizing the successful candidate

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



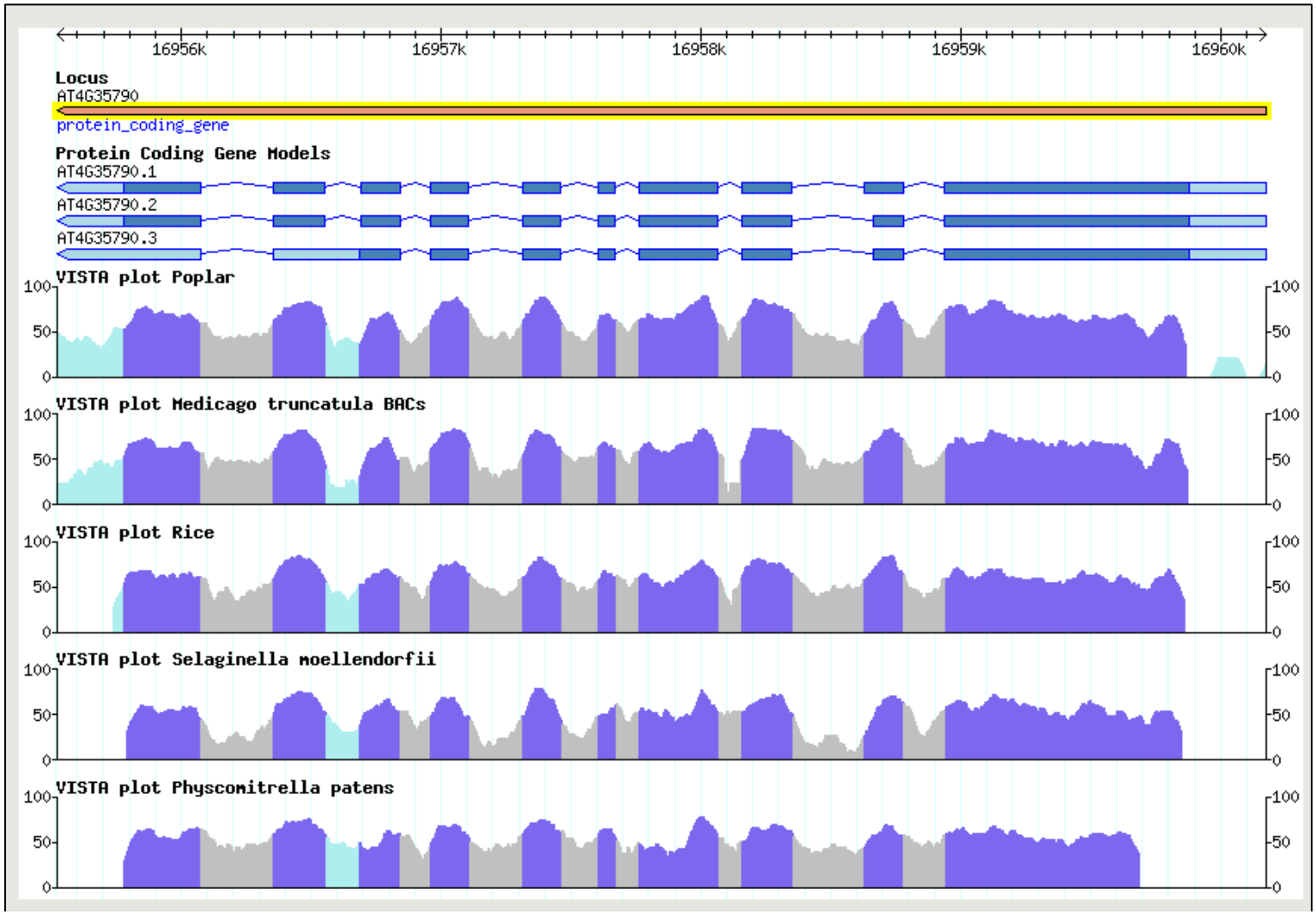
Characterizing the successful candidate

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



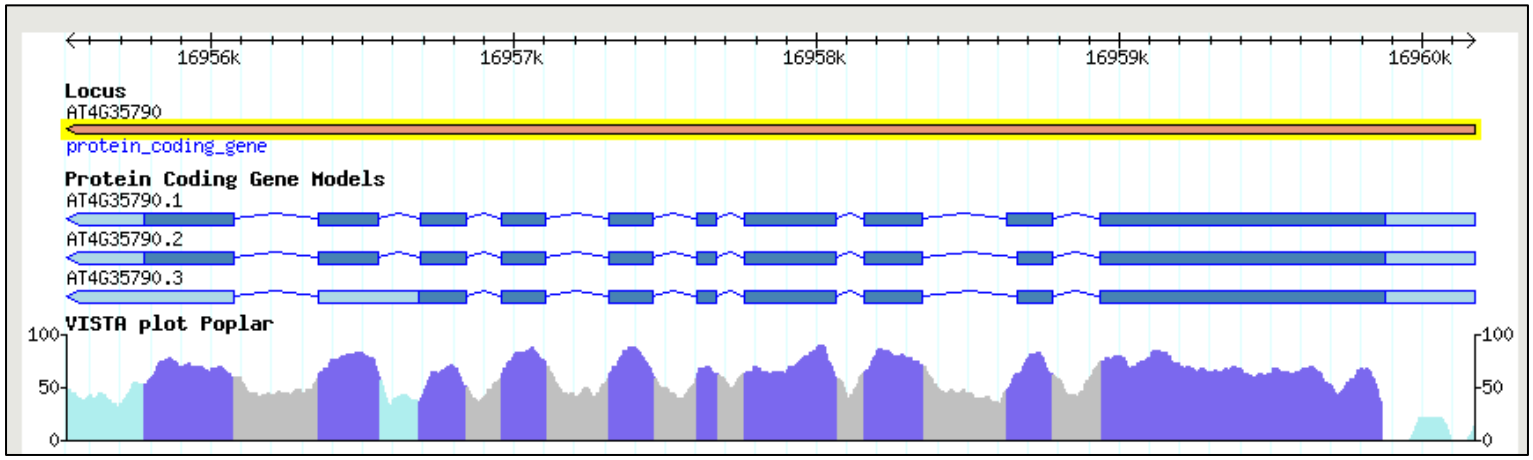
Characterizing the successful candidate

- View orthologs using VISTA tracks in GBrowse



Characterizing the successful candidate

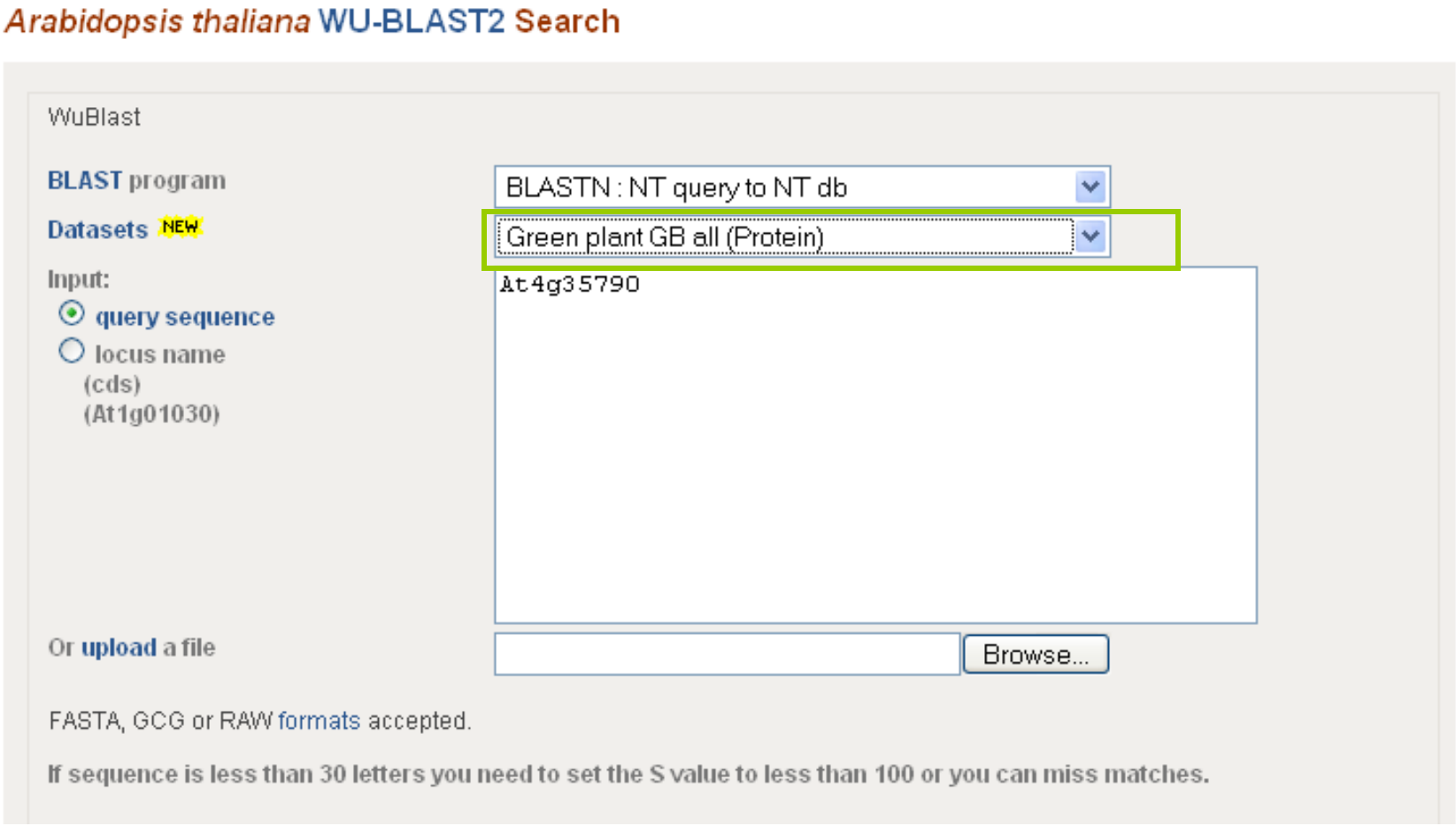
- View orthologs using VISTA tracks in GBrowse



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: 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: 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



Characterizing the successful candidate

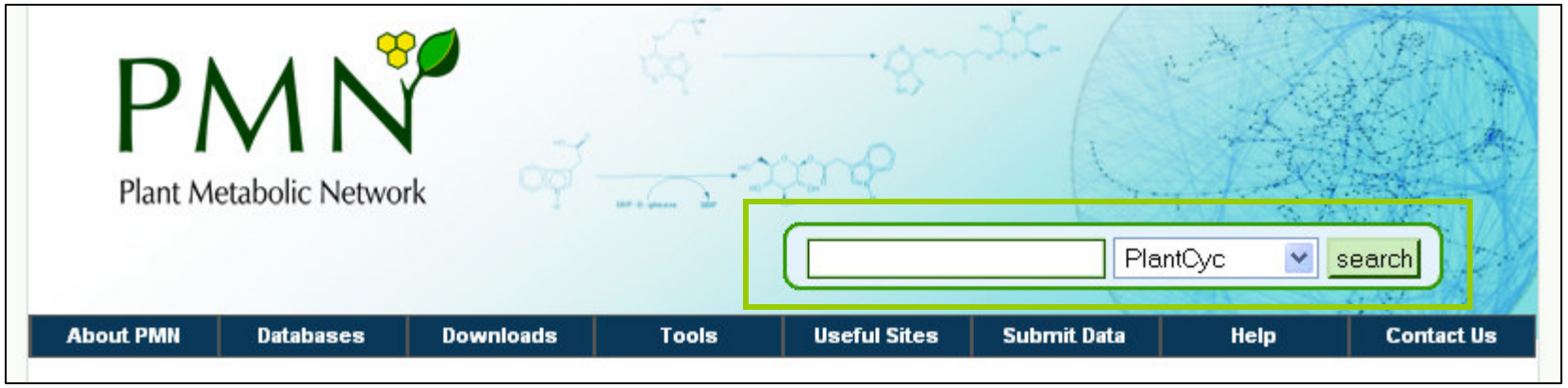
- Make connections to proteins in other species using GO terms on the Locus page

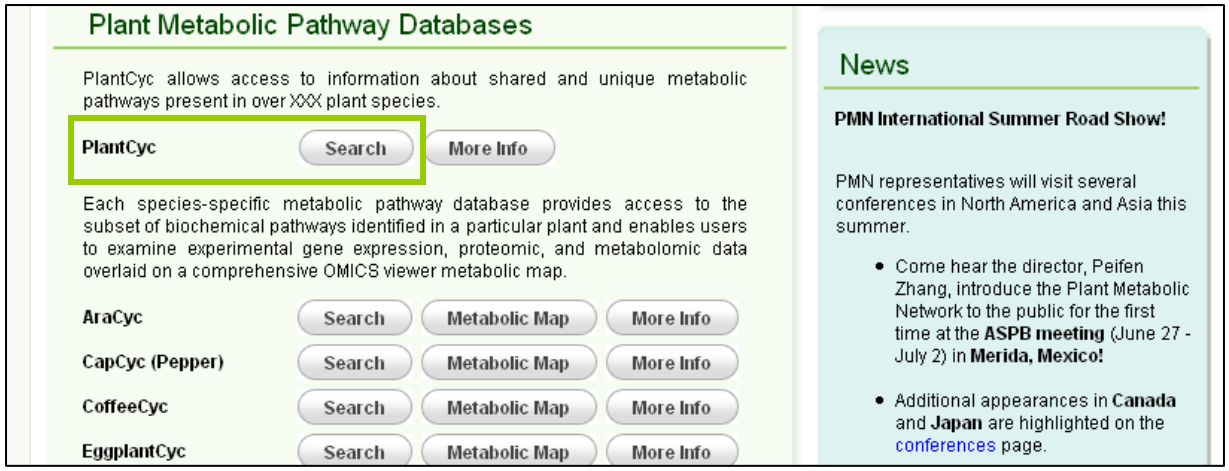
Annotations	Category	Relationship Type	Keyword
	GO Biological Process	involved in	response to cold, programmed cell death, phosphatidic acid metabolic process
	GO Cellular Component	colocalizes with	microtubule cytoskeleton, plasma membrane
		located in	plasma membrane
	GO Molecular Function	has	phospholipase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

Annotation Detail

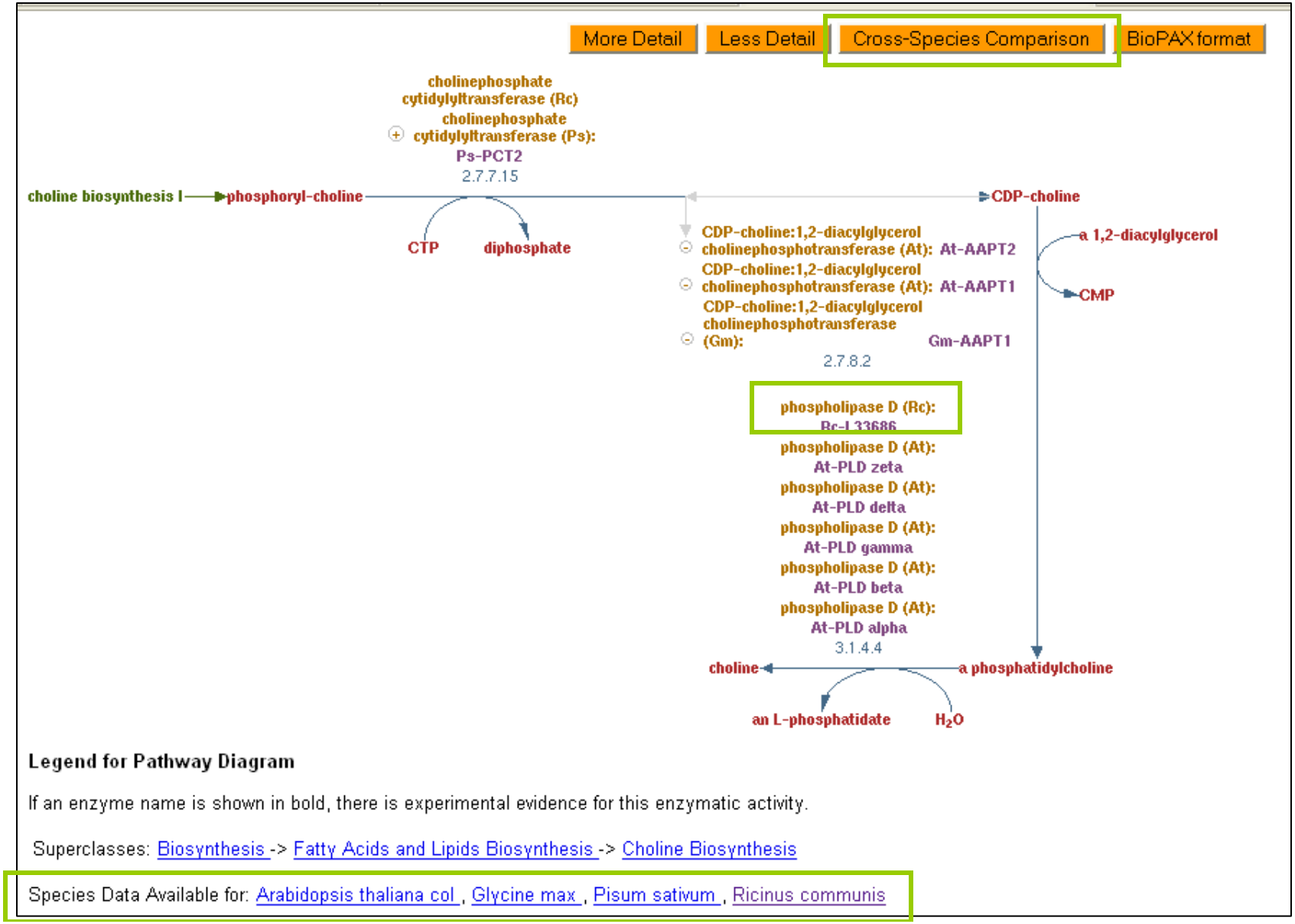
Characterizing the successful candidate

- Make connections to proteins in other species using PlantCyc





Characterizing the successful candidate



Generating more “candidates” at the PMN

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

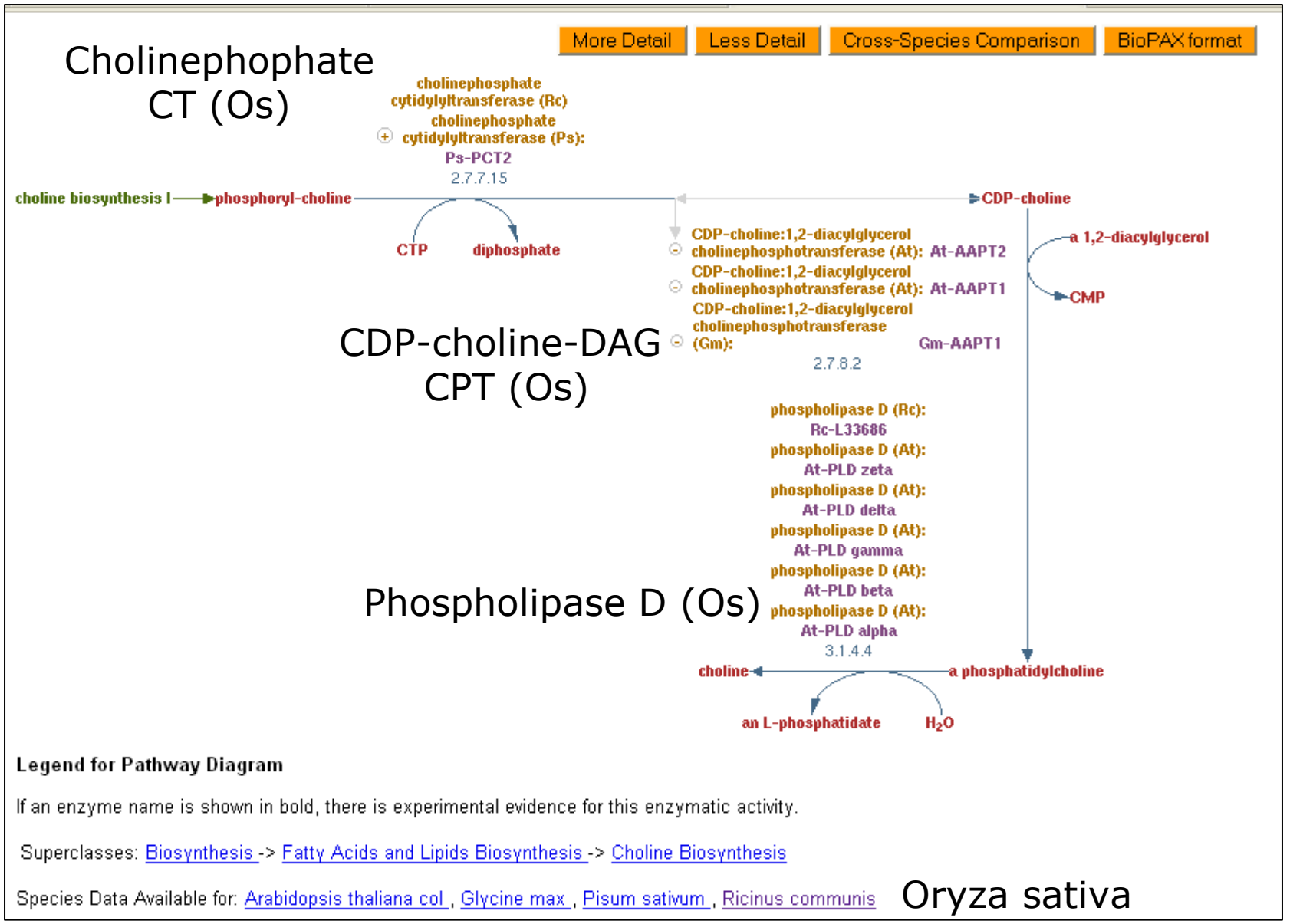
Generating more "candidates" at the PMN

- 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



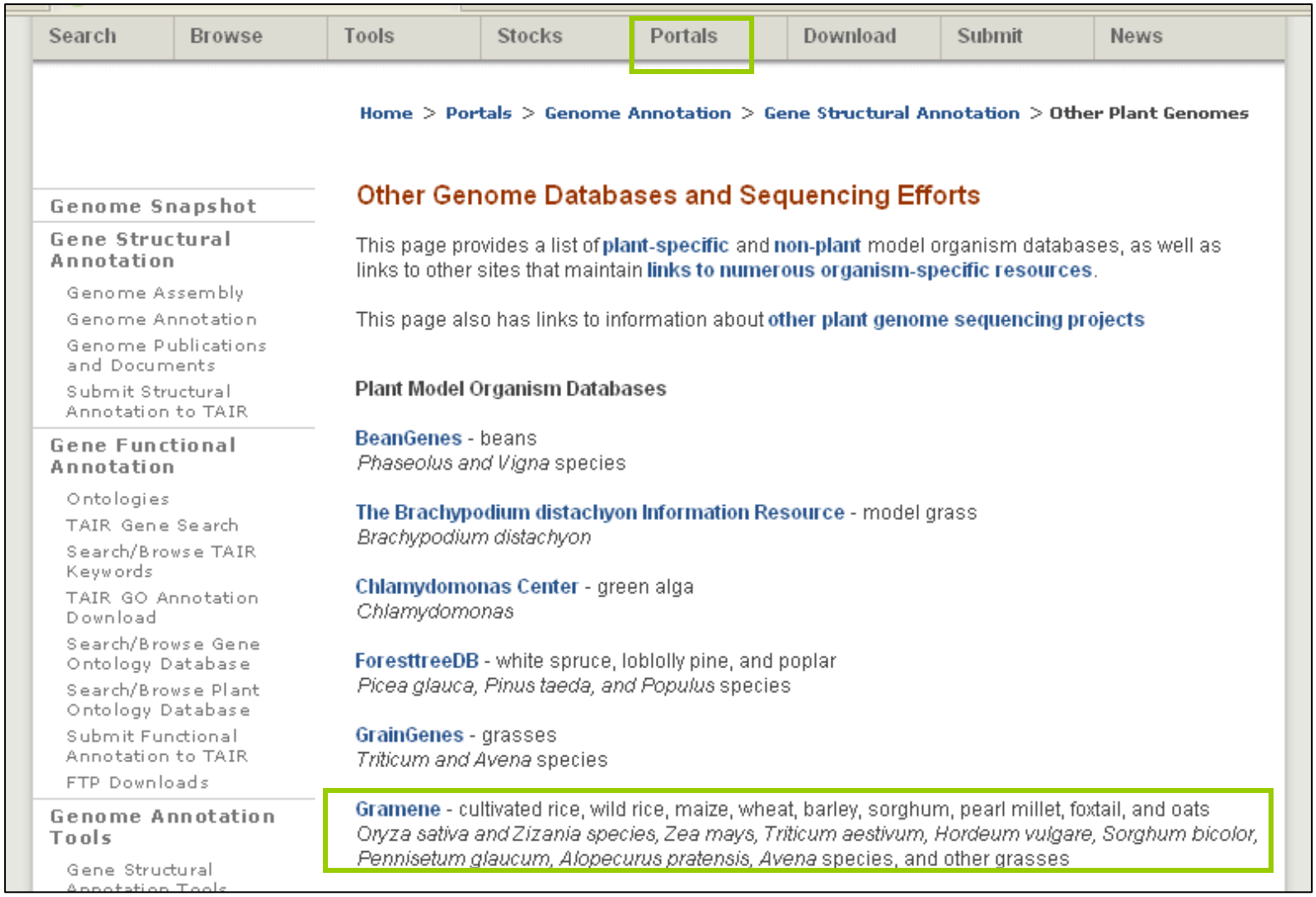
Characterizing the successful candidate

- 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

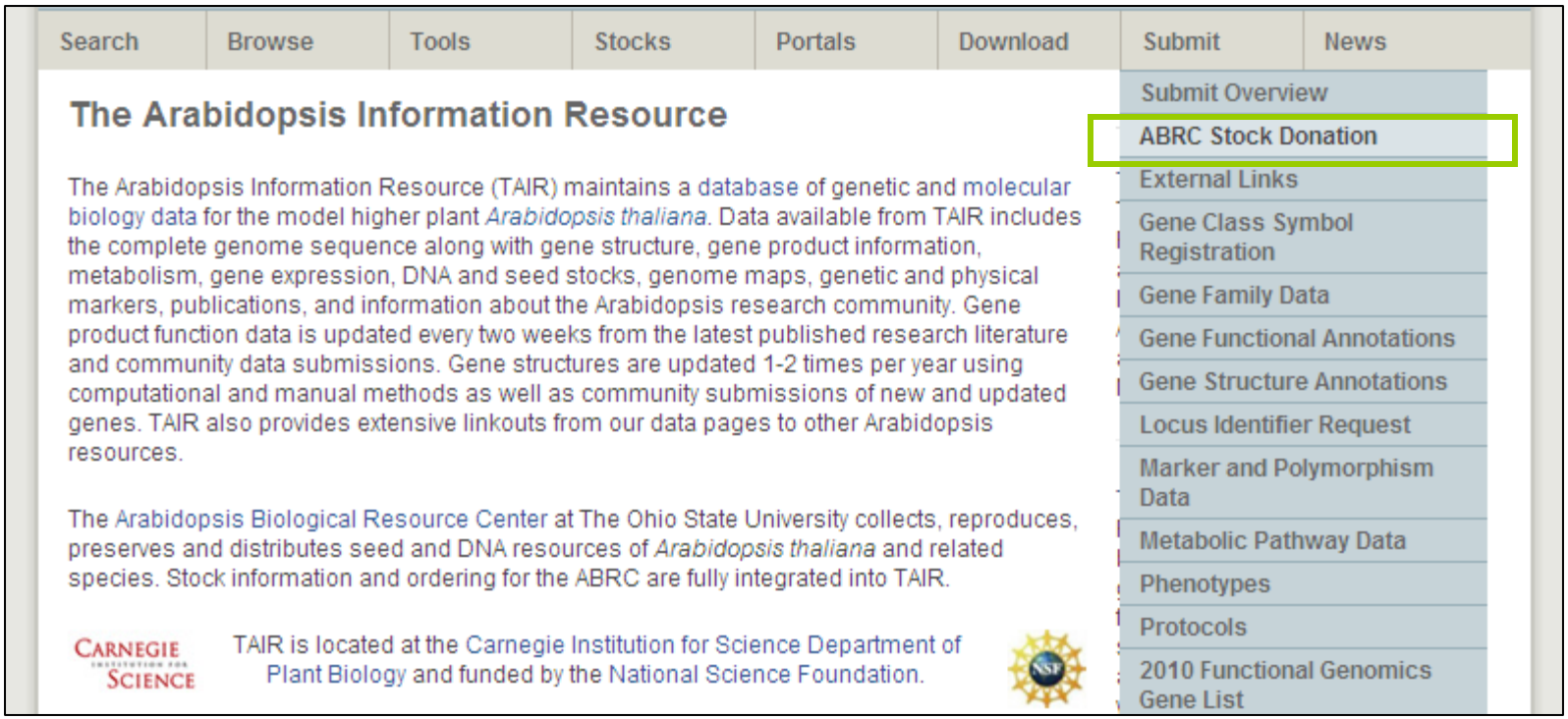
- Gain access to rice resources through MODs portal page



Sharing your knowledge with everyone . . .

Sharing your resources and tools . . .

- ABRC Submission
  - Seed stocks
  - DNA stocks



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

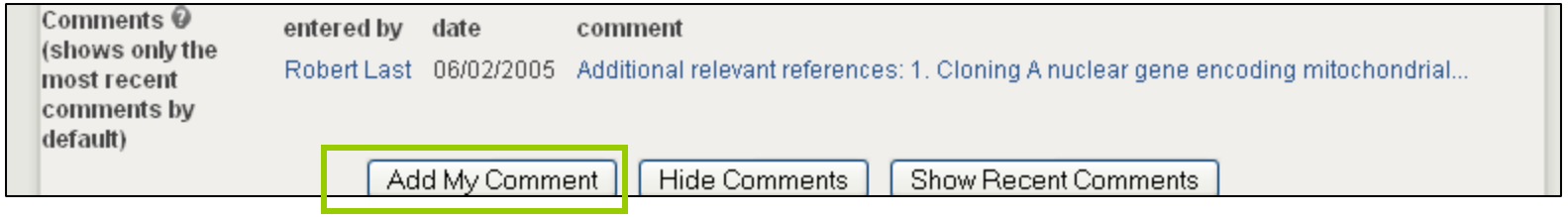
Publishing your results . . .

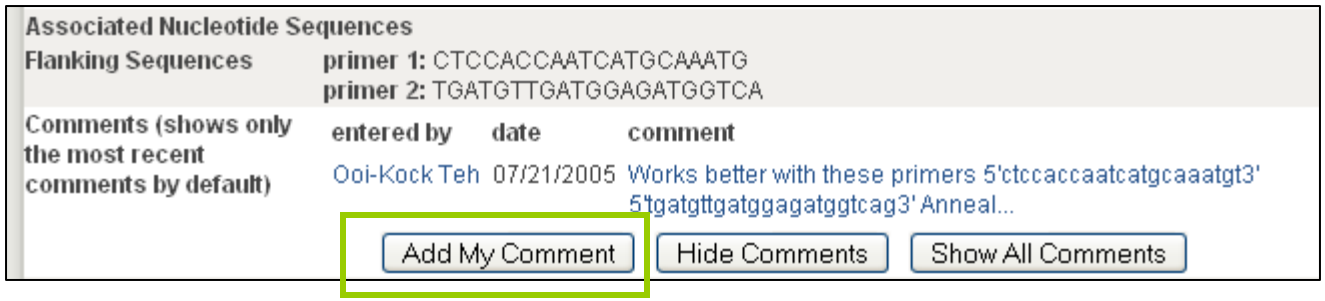
- TAIR-Journal collaboration
  - Authors submit functional annotations at time of paper acceptance
  - Plant Physiology requires participation as of March 2008
  - Other journals may join in the future

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 →





← 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.

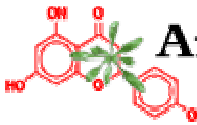
tair

PMN
Plant Metabolic Network

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!

tair



PMN
Plant Metabolic Network

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