How to effectively use the tools and resources at TAIR to enhance your research

Philippe Lamesch Donghui Li A.S. Karthikeyan

The Arabidopsis Information Resource www.arabidopsis.org

contact us: curator@arabidopsis.org

Model Organism Databases











TAIR: The Arabidopsis Information Resource

- The Arabidopsis Information Resource
- collect, curate and distribute information on Arabidopsis
- information freely available from arabidopsis.org

etair	Home Help	Contact About U≤	: Login/Register			Gene	Search
Search	Browse	Tools	Stocks	Portals	Download	Submit	News

The Arabidopsis Information Resource

The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. 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.

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

Breaking News

TAIR Curator position, Stanford, CA

We're seeking a genome annotation and literature curator beginning Jan-Mar 2008. Biology Ph.D and strong research background are required.

AraCyc 4.1 release

23 pathways were significantly updated in the last release in October. More **details.**

Outline

- Gene structure Philippe
- Gene function Donghui
- Metabolic pathway Karthik

Goal of this workshop

- highlight important data sets and tools
- demonstrate search strategies

Slides available from TAIR

Search	Browse	Tools	Stocks	Portais	Download	Submit	News.		
				Portais Ove	rview				
		Home > Portals > Education	Ciones/DNA Resources		ons(ppt)				
		10.00			Education and Outreach				
		ICAR WOR	ICAR Workshop 2009		Gene Expression Resources				
About Ara	bidopsis	TAIR gave a v	TAIR gave a workshop at the 20		Genome Annotation		pidopsis Research.		
Search Arabidopsis Protocols and Manuals Science Education and Outreach Programs		Putting TAIR	Putting TAIR to work for you - • Date: July 2, 200 • Time: 4:30 - 6:00 • Location: Kitsyth	MASC/Fund	tional Genomica	ig and advanced users			
				Mutant and Mapping Resources Nomenclature					
reaching i	cesources #ducettes	•	Conference site:	Proteomics	Resources				
and Outre	ach	Part I:	Tips and strateg	Metabolomi	cs Resources				
Plant Biolo Education Newsgrou	rgy ps	1. Gen 2. TAR 3. Gen	e searches and th R9, GBrowse, and rerating and worki	e locus page confidence so ng with data s	(presented by Eva cores (presented by its (presented by A	Huala) r Dave Swarbrec (ale Dreher)	N2		
TAIR Presentati	ions(ppt)		TAIR ICAR Work	shop Resource	e Guide				
TAIR Help		Part II (prese	Hands-on exerc nters and participe	ises and indi	vidual help				
			TAIR ICAR Work	shop Practice	Questions				
			TAIR workshop - (Microartay / OM	pert 2 - data f ICs viewer da	te file for Question	#1)			
			1			a line and			

TAIR is used worldwide

Visits per month (source: Google Analytics)



What we do: (1) Arabidopsis genome annotation



What we do: (2) manual literature curation

• Controlled vocabulary annotations

Gene Ontology (GO) http://www.geneontology.org/ Plant Ontology (PO) http://www.plantontology.org/

- Gene name, symbol
- Allele, phenotype
- Summary statement composition

What we do: (3) metabolic pathway curation

AraCyc

A metabolic pathway database for Arabidopsis thaliana that contains information about both *predicted* and *experimentally* determined pathways, reactions, compounds, genes and enzymes.

PlantCyc and PMN (Plant Metabolic Network)

What we do: (4) work with ABRC to distribute research material

Arabidopsis Biological Resource Center



- AMC Huma Page
- C Search Seed Disciss

ABRC Home

- C Search DhA Slovina
- O Browne Galating
- O Ohanging Soviere
- O Hear Int Make Payments
- O Order Stocks
- C Search Driler Halory
- O Stock Genetiset, Forma
- O DAA Slock Handling
- O Flani and Seni Handing
- Stock Acquisition/ Descipation Policy
- C ABIC Personnel
- C Alvisory Committee
- Seed and DNI Facilities
- ORDE
- C Links to other thesources
- O POMI Huma Fage
- C Christianis UK

Arabidopsis Biological Resource Center

The Anthiotopsis Biological Resource Center (ABRC) was established at The Chip Spate University in September, 1991. Primary support for the ABRC is provided by a National Science. Foundation prent. The mission of the ABRC is to acquire, preserve and distribute seed and DNA resources that are useful to the Anthiotopsis research community.

Dr. Randy Schor is the Director of the Center, and Associate Professor in the Department of Plant Century and Morecular Biology

New seed and DAA error ABRC through donation by national and international researchers. An identifying number is assigned to each stock. The stock is then evaluated, documented and made available for distribution. The preservation of seed and DNA stocks is conducted according to strict, defined protocols. Backups of stocks are maintained to ensure preservation.

The ABRC's holdings and stock distribution takes have increased rapidly as that presently hundress of thousands of stocks are available. More than 100,000 stocks are shopped annually to researchers in more than 60 countries, and modeal tees for stocks are charged. How do TAIR curators annotate gene structures? Where can gene structure related data be found at TAIR?

Part I: Tips for Accessing Gene Structure Data

Philippe Lamesch

Gene structure Annotation

- Arabidopsis genome sequenced almost 9 years ago
- High quality sequence with few gaps
- TIGR did initial genome annotation
- TAIR took over responsibility in 2005
- Current stats:

27,379 protein coding genes4827 pseudogenes or transposable elements1312 ncRNAs





Add novel genes



Delete wrongly predicted genes



Update mispredicted exon-intron structure



Update mispredicted exon-intron structure



Annotate splice-variants



Annotate 'atypical' gene classes



Define gene type



- ✓ Protein-coding
- ✓ tRNA
- ✓ snRNA
- ✓ snoRNA
- ✓ rRNA

•••

Gene structure annotation in Arabidopsis





















Manual annotation at different MODs



Manual annotation at different MODs



Nucleotide sequence Short peptides Protein similarity Alternative predictions

...

Apollo (Arabidopsis, Fly) Aceview (Worm) Zmap/Otterlace (Human) Artemis (Pathogen Project)

...

Exon size Intron size Number of UTRs Coding/Non-coding ratio Splice-junctions

• • •

Manual annotation at TAIR: Apollo





Examples of large-scale community datasets recently integrated into the Arabidopsis annotation

- Transposable elements (Quesneville Lab)
- Pseudogenes (Gerstein Lab)
- Short MS peptides (Baerenfaller et al, Castellana et al)
- Short genes (Hanada et al)

Where can you find gene structure data on TAIR?

- ON GENE MODEL PAGE
- Graphic of exon-intron structure
- Coordinates of each exon
- ON GBROWSE
- Graphic display of structure and overlapping evidence data
- ON FTP SITE
- GFF files with exact structures of each gene model
- Files with gene confidence ranking information

Gene Locus Page

tair	Hores Help Contact	About Un Logen/Register		Gere	+ Searc
Search I	Irowse Tools	Stocks Portals	Bownload	Submit	News
Locus: AT1	G07780				
late last nodified	2007-04-17				
AIR Accession	Locus 2026509				
lepresentative Sene Model 9	AT1G07780.3				
Jene Model Type	profeim_coding				
ther names:	AT1G07780 1, F2489 1 POTENTIAL 6, TRP6	1, F2489_11, PAI1, PHOSPHORIBO	SYLANTHRANLATE ISO	DMERASE 1, TR	ANSIENT RECEPTOR
escription 0	Encodes phosphoribosy	vianthrankate isomerase which cats	wzes the third step of th	he tryptophan bi	osynthetic pathway
	Member of gene family.		12	100 B	8
Ither Gene Models Ø	Member of gene family AT1007780.4 AT100 (splice variant)	07780.1 AT1007786.2 (aplice varient)			
Other Gene Models Ø	Menter of gere family AT1007780.4 AT104 (splice variant)	07760.1 AT1067786.2 (aplice variant) 24535		28125	
Other Gene Models ♥	Menber of gene family AT1007780.4 AT100 (splice variant) Frotein Coding	24123. (aplice variant) 24123. Gene Modela		· 2425 + +	
Ither Gene Indels Ø	Member of gene family AT1007780.4 AT10 (splice variant) Protein Coding sT1007780.1	ATTODITION ATTODITION 2 (aplice variant) 24534 Gene Hudels		* 2425 * * *	· · · · · · · · · · · · · · · · · · ·
Ither Gene Models Ø Map Detail mage	Member of gene family AT1007780.4 AT100 (splice variant) Frotein Coding stite(7700.1 at1b07700.2	arrad 1 ArtiOsrrae 2 (aplice variant) 2455A Gene Modelle		- 24120	· · · · · · · · · · · · · · · · · · ·
Ither Gene Models @ Map Detail mage	Member of gene family AT1:007780.4 AT1G (splice variant) Frotein Coding #T1507780.1 #T1507780.2 #T1507780.3	27780.1 AT1087786.2 (aplice variant) 2NESA. Gene Models		28125	
Ither Gene Models Ø Map Detail mage	Member of gene family AT1007780.4 AT10 (splice variant) Protein Coding sT1507780.1 sT1507780.2 sT1507780.3 sT1507780.4	2NEIX		* 2423 * * *	
Ap Detail mage	Member of gene family AT1007780.4 AT104 (splice variant) Frotein Coding stitler7780.1 at1007780.3 at1007780.3 at1007780.4	2455A		*2428 * * *	
Ither Gene Map Detail mage	Member of gene family AT1007780.4 AT100 (splice variant) Protein Coding #T1007780.3 #T1007780.3 #T1007780.4 Category 00 Between Sec.	ATTOOTTBO 2 (aplice variant) 24123. Gene Hodels Relationship Type @	Keyword Ø	*28125	
Ither Gene Models @ Map Detail mage	Member of gene family AT1007780.4 AT10 (splice variant) Protein Coding st1007780.1 st1007780.2 st1007780.3 st1007780.4 Category GO Biological Process	ATTOD 1 ATTOD 7780.2 (aplice variant) 2453A Gene Hodele Relationship Type @ involved in	Keyword Ø tryptophan biosynt	tetic process	
Ither Gene Aodels Ø	Member of gene family AT1007780.4 AT10 (splice variant) Frotein Coding st1507780.1 st1507780.2 st1507780.3 st1507780.4 Category GO Biological Process GO Celular Component	2453A Gene Models Relationship Type @ involved in located in located in	Keyword Ø bygtophan biosynt chloroplest	tetic process	
Ither Gene Models @ Map Detail mage	Member of gene family AT1:007785.4 AT1:0 (splice variant) Protein Coding #T1007780.3 #T1007780.3 #T1007780.3 #T1007780.4 Category 00 Biological Process 00 Celular Component 00 Molecular Function	2NE2A. Gene Hodels Relationship Type @ involved in located in has	Keyword Ø tryptophan biosynt chloroptest phosphoribosytent	tetic process	nse schvity
Gene Model Page



Where can you find gene structure data on TAIR?

- ON GENE MODEL PAGE
- Graphic of exon-intron structure
- Coordinates of each exon
- ON GBROWSE
- Graphic display of structure and overlapping evidence data
- ON FTP SITE
- GFF files with exact structures of each gene model
- Files with gene confidence ranking information

Gbrowse



GBrowse

Header	 manituring and state and state and a state	Dates FOCUDA, Carcinative admin (Divergence) Tel hand divegence, disk "configure" and to Bayeria & Analysis Download Decom Scontilizeers ((()))	a "Tepert & Endorte: Manufact Visco Port", Else (dat Real FAGTA-File & Condona Goo I [Brear 2014] x -1 - 2020 [File	n Trođiga e Tan Ser aleka gloga bo dingdog fan 13 al ba.
Main Browser Window	Interference Interference <td></td> <td></td> <td></td>			
Track Menu	 The second		International Control of Control on Section 2014	Setting of a surger 4.2 Mills and a subject to the Comparison of the Conference of the Conference of the Conference of the Conference of the Conference of the Setting of the

Where can you find gene structure data on TAIR?

- ON GENE MODEL PAGE
- Graphic of exon-intron structure
- Coordinates of each exon
- ON GBROWSE
- Graphic display of structure and overlapping evidence data
- ON FTP SITE
- GFF files with exact structures of each gene model
- Files with gene confidence ranking information

FTP site

Search	Browse	Tools	Stocks	Portala	Download	Submit	News
					Download Ov	erview	
The Ar	hidonele le	formation	Persures		Genes		
THE AL	abidopaia ir	normation	Resource	2 322 522 543	GO and PO A	resplations	
The Arabid	opsis information	Resource (TAR	maintains a datai	base of genetic and a	Mape		ALTAIR
the complet	e genome sequer	ce along with p	ene structure, gen	e product information	Metabolic Pa	divice/s	comparing
pène expre	salon, DRA and s	eed slocks, gen	ome maps, genetic	c and physical marks	Polymorphis	PER.	etween A
date is upd	ated every two w	eeks from the la	dest published res	earch iterature and	Proteins	dded soon.	
data subrei	salona. Gene stru	ctures are upda	ded 1-2 times per y	rear using computate	Protocols		
provides a	denaive lickouts 1	form our data pa	ges to other Arabi	dopais resources.	Microarray D	ofie:	
The Arabid	nosis Riplopical R	Hource Center	at The Onio State	University collects, re	Sequences		lelease
preserves	and distributes se	ed and DNA res	ources of Arabida	pois thaliana and re	Softmare) release is
Stock infor	nation and orders	of for the ABRC	are fully integrate	d web TAR,	User Reques	E8.	AR and NCBI,
SCIINC	CADADON TAR is located at the Carnegie Institution for Science Department of F Science Biology and funded by the National Science Foundation.					new spice var	
T Worksh How to your res Worksho	AIR at t op I: effectively use earch: July 2 op II:	he 2009 the tools a 0 12:20-2p	9 ASPB (nd resources m Room 318	Conference at TAIR to enhan	:е юн	Tips for sear stocks inclus and amiRBA of 2009j Are you search and vectors av ABRC? Here a	ching DNA ling vectors clones (May 20, hing for clones railable from re some tips for

Annual Meeting of the American Society of Plant Biologists & the Phycological Society of America

pages. Specific instructions for

finding amRNA clones and multifunctional vectors are in the state of

FTP site

Index of ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/TAIR9_gff3/

🖺 Up to higher level directory

Name	Size	Last M	odified
Assembly_GFF		6/19/2009	10:19:00 PM
Community_annotation_GFF		6/19/2009	10.17:00 PM
Expression_GFF		6/19/2009	11:14:00 PM
Genomic_features_GFF		6/19/2009	11:24:00 PM
Methylation_Phosphorylation_GFF		6/20/2009	8:41:00 PM
Conholog_GFF		6/19/2009	11:18:00 PM
Sequence_similarity_GFF		6/20/2009	9:37:00 PM
TAIR9. GFF3. genes.gft	39298 KB	3/26/2009	12:12:00 AM
TAIR9_GFF3_genes_transposons.gff	44833 KB	6/19/2009	11:06:00 PM
Variation_GFF		6/25/2009	6:34:00 PM
arabidopsis.conf	121 KB	6/29/2009	5:28:00 PM

FTP site

(The set	TATES	-		14	204024			Therefore 1 - Manager 7h = 1
Cont A	TATES	CHARGE ST	2622		0014.70			PROPERTY AND A PROPER
ing1	TALKS	gene	3631	2033	*	22		ID-AIIO01010/Note-protein_coding_generName-Aii001010
nr1	TAIKS	BOOLA	3631	2233		*		1D-ATISOIDIO.1/Parent-ATISOIDIOJSame-ATISOIDIO.1/Index-1
.hr1	TAIRS	procein	3760	5630				ID-Aligololo.1-Fretein/Mame-Aligololo.1/Derives_frem-Aligololo.1
chr1	TAIRS	exon	3631	3913	1	*		Parent=AT1G01010-1
hrl	TAIRS	five_pr	ime_UTR	3631	3759	4		. Parent=ATIG01010.1
ihrl	TAIRS	CDS	3760	. 3913		* C	0	Farent=AT1901010.1,AT1901010.1-Protein;
Thr1	TAIRS	6805	3996	4276	(4)	*		Farent=AT1G01010.1
nx1	TAIRS	CDS	3996	4276	*		- 2	Parent=AT1901010.1,AT1901010.1-Proteins
hr1	TAIR9	0300	4486	4605	4	+	1.1	Farent=AT1901010.1
hr1	TAIRS	CDS	4406	4605		*	0	Farenc=AT1901010.1,AT1901010.1-Protein)
Thr1	TAIRS	exon	4706	5095		+		Farent=AT1G01010.1
ngi	TAIRS	CDS	4706	5095		+	0	Farent=AT1001010.1,AT1001010.1-Protein;
Thr1	TAIR9	0300	5174	5326		+		Farent=AT1G01010.1
Chr1	TAIRS	CDS	5174	5326		+	0	Farent=AT1001010.1, AT1001010.1-Protein:
Thr1	TAIR9	excn	5439	5599		+		Farent=AT1G01010.1
ing1	TAIRS	CDS	5439	5630		+	0	Farent=AT1001010.1,AT1001010.1-Protein/
Chrl	TAIRS	three p	rime UTR	5631	5599		+	. Farent=AT1G01010.1
Chri	TAIRS	gene	5928	8737		÷3		ID=AT1001020/Note=protein_coding_gene/Name=AT1001020
Thrl	TAIRS	mENA	5928	8737				ID=AT1G01020.1; Parent=AT1G01020; Name=AT1G01020.1; Index=1
Chg1	TAIRS	protein	6915	8666	÷	-		ID=AT1001020.1-Protein/Mame=AT1001020.1/Derives_from=AT1001020.1
Inr1	TAIRS	five pr	ime UTR	8667	6737		-	. Farent=AT1G01020.1
the1	TAIRS	CDS	8571	8666		-	0	Farent=AT1001020.1,AT1001020.1-Protein;
Chrl	TAIRS	excn	8571	8737		-		Farent=AT1G01020.1
Chri	TAIRS	CDS	8417	8464	÷.	-	0	Farent=AT1001020.1,AT1001020.1-Protein;
Thrl	TAIRP	excn	8417	2464		-		Farent=AT1601020.1
Chri	TAIRS	CDS	8236	8325		-	0	Farent=AT1001020.1,AT1001020.1-Protein/
Chrl	TAIRS	excn	0236	0325		-		Parent=AT1G01020.1
Chrl	TAIRS	CD8	7942	7987	12 C	-	0	Farens=AT1001020.1.AT1001020.1-Proteins
Chr1	TAIRS	exon	7942	7987		-		Farent=AT1G01020.1
Chrl	TAIRS	CDS	7762	7835	2	-	2	Farent=AT1001020.1.AT1001020.1-Protein/
Chrl	TAIRS	##CD	7762	7835		-		Farent=AT1G01020.1
Chrl	TAIRS	CDS	7864	7649	2	-	0	Parent=AT1001020.1.AT1001020.1-Protein/
-					1.1		1257	

Where can you find gene structure data on TAIR?

- ON GENE MODEL PAGE
- Graphic of exon-intron structure
- Coordinates of each exon
- ON GBROWSE
- Graphic display of structure and overlapping evidence data
- ON FTP SITE
- GFF files with exact structures of each gene model
- Files with gene confidence ranking information

Gene Confidence Rank

 Attributes confidence scores to all exons and gene models based on different types of experimental and computational evidence

TYPES OF EXPERIMENTAL & COMPUTATIONAL DATA USED

We used 4 types of experimental evidence to generate confidence scores:

- 1) Arabidopsis transcript data (ESTs, cDNAs, fragmented mRNAs from Ecker lab: Lister et al Cell 2008)
- 2) Proteomics data (Castellana et al PNAS 2008; Baerenfaller et al Science 2008)

3) X-species alignments (Brassica, Radish, + 5 other dicot (Glycine max, Solanum tuberosum, Malus x domestica, Gossypium hirsutum and Vitis vinifera) and 5 other monocot (Triticum aestivum, Zea mays, Oryza sativa, Saccharum officinarum and Hordeum vulgare) sequences from JCV1 http://plantta.jcvi.org/ index.shtml)

Vista genomic conservation (Shah et al Bioinformatics 2004)

Assigning A Confidence Rank





그는 영양은 영양이 이야지 않는 것이 같이 많이 많이 했다.

How to effectively use the tools and resources at TAIR to enhance your research

Part II: Tips for Accessing Gene Function Data

Donghui Li



Focus on how to use the tools, not on how we make functional annotation

- 1. Finding a specific gene of interest in TAIR
- 2. Getting lots of data about that gene of interest
- 3. Getting data sets on multiple genes
- 4. "Analyzing" your data sets

1. Finding a gene/protein

Case 1: You know exactly what gene you want

- You know the AGI locus code (e.g. At2g46990)
- You know the gene symbol (e.g. PHYA)

1. Finding a gene/protein: by name/AGI locus code

Use the quick search box!

80								
tair	None Auto	Cortest . Hour	the LogivPager	w.	PROA	Gene	2	dearth
Seath	Denarts	Tools	989c810	Portale	Donational	Soleyl.	Heat	

Check All	Uncheck All Description P Transcription factor interacting with photoreceptors phyA and phyB. Formal elementy complex in vitro with 0-box element of th	Oene Modello Ø ATTO09530.1	Other Names P F14J9,18 F14J9,19 FNP3 PHOTOCURRENT1 PHOTOCURRENT1 PHOTOCURRENT1 PHOTOCURRENT1 PHOTOCURRENT3	Expression @ 4 anthesis, C globular stage, D bilateral stage, DNA binding, E expanded totyledon stage, F mature embryo stage, UF D4 four leaves visible, carpet, cauline leaf, circadian methan, de-etiolation, embryo, fower, pibberelic acid mediated signaling, hypocotyl, inforescence mensitem, leaf, leaf apez, leaf lamina base, likef whori, nucleus, pedicel, petal, petal offerentiation and expansion stage, positive
		A71008530.2	PHYTOCHROME- ASDOCIATED PROTEIN 3 P#F3 PIOC1	regulation of anthocyanin metabolic process, protein binding, red or far red light signaling pathway, response to red or far red light, seed, sepal, shoot, shoot apex, signal hansduction, stern, hanocription factor activity, transcription regulator activity
	Light-rabile intoptaismes rediffar-red SgM photoexceptor envolved as the regulation of photoexceptropeneesis. If autalis in two L	A710685781	ELONOATED HYPOCOTIE_8 F14J9_23 F14J9_23 FAR RED ELONOATED 1 FAR RED ELONOATED HYPOCOTIE_2 FHY2 FRE1 HY8 PHYA PHYTOCHROME A	A anthessa, 4 leaf seriescence stage, C globular stage, D bilateral stage, E expanded cotyledox stage, F instance embryo stage, 0-protein rougred photoneceptor activity, LP /02 two leaves visible, LP /04 four leaves visible, LP /05 six leaves visible, LP /08 eight beaves visible, LP /05 six leaves visible, LP /09 eight beaves visible, LP /01 len beaves visible, LP /03 twelve leaves visible, CP /05 exclosers visible, LP /03 twelve leaves visible, CP /05 exclosers visible, LP /03 twelve leaves visible, carbel, caution teat, cotytedox, rytoplasm, embryo, flower, gravitopson, typocoti, inforescence membries, leaf, leaf apex, leaf tamona base, leaf whort, maie gametophyte, nuclear bods, nucleus, pedicel, petal, petal differentiation and expansion stage, petiole, phototopism, protein histicine kinase activity, red light signaling pathway, red or far-red light photoeceptor activity, regulation of transcription, DNA-dependent, response to far- ning income to continuous far red light stimutus to the night income response system, response to far- red light photoeceptor schools to far- red instage response to the red light stimutus to the

1. Finding a gene/protein: Exact name search



1. Finding a gene/protein: by sequence

Case 2: You have a gene / protein sequence and want to find the most similar gene/protein(s) in TAIR

- You have the nucleotide sequence of a gene from another organism and want to find a homolog
- You have the amino acid sequence of one Arabidopsis protein family member and want to find related family members

Use a sequence similarity search tool BLAST

1. Finding a gene/protein: by sequence

| Dyenois: Mapin | Mapin (METRICHING ANIMOPEPTI. ...

1,049900138

BLAST query on TAIR9 Genes (+introns, +UTRs) (DNA) sequences Query performed by the The Arabidopsis Information Beauros (TAIP) for full BLAST options and parameters, right to the MCBIBLAST Discoversistion Your comments and suggestions are requested Send a Message to TATE BLASTE 2.2.17 [Aug-26-2067] Beferences allechal, Stephen F., Thomas L. Badden, Aleyandro J. Schaffer, Jinghun Thang, Theng Thang, Webb #111er, and Bavid J. Lipsen (IPP7). "Gapped SLAFT and PSI-SLAFT: a new generation of protein database search proprise?, Nuclear Acids Les. 25:3300-3402. party- stor-schwitted sequence (SECH Letters) Batabase: SEIRS Genes (cintrons, offici) (DEA) 39,640 seguences: 87,027,097 total letters Doug . н Sequences producing significant alignments; (DIDE) Value ations/10.1 | Dyshulat PETA, PETI, PETI, STR | PETA (PETTOC.... 185 +-165 stionsto,1 | Symbols; FETA, FET2, FEEL, SYE | PETA (PETTOC 180 6-165 #2%G1#530.1 1 59Mb01#1 PRTE 1 PRTE (PRTTOCHDORE DEPECTIVE K.... 32 74-05 sTE042640.1 i Symbols: i sind finger (CINS type) family pr. ... 48 0.005

44

0.004

2. Getting data from locus page

- locus
- other names
- description
- map detail
- annotations
- RNA data
- Nucleotide sequence
- Protein data

- polymorphisms
- germplasms
- phenotype
- external links
- publication

2. Getting data from locus page: description

- curated description
- computational description

Locus: AT	IG09570
Date last modified	2003-05-02
TAIR Accession	Locus 2012300
Representative Gene Model Ø	AT \$500528.3
Gene Model Type	protein_coding
Other takenet:	ELONGATED HIPOCOTYL 8, F14.96.33, F14.96_23, FAR RED ELONGATED 1, FAR RED ELONGATED HIPOCOTYL 2, FHY2, FRE1, HV8, PHY3, PHY3DCHROME A.
Description Ø	Light-table cytoplasmic red/lar-red light-photoreceptor involved in the regulation of photomorphogenesis. It exists in two inter-convertible forms: Pr and Ptr (active) and functions as a dirver The N terminus carries a single tetrapersie chromophore, and the C terminus is involved in dimerization. It is the sole photoreceptor mediating the FR high imadiance response (HIR). Major regulator in red-light induction of phototropic enhancement, involved in the regulation of de-etislation. Involved in gravitropism and phototropism. Requires FW11 for nuclear accumutation.

2. Getting data from locus page: description

• computational description

Representative Gene Model © Gene Model Type Other names:	ATTIGAUSAILT protein_coding MPCUID_E_MPCUID_E
Description 9	enc finger (C2H2 type) femily prosen; FUNCTIONE IN: transcription factor activity, pric on omding, ruceac acid tending; MVOLVED IN: regulation of transcription; UOCATED IN: intracelulus; CONTAINED InterPric DOMAINE: Zhic finger, C2H2-ska (InterPric PRO15880); Zhic finger, C2H2-type (InterPric IPR0215987); BEET Avatistopers Italiana prosen match is zinc Tinger (C2H2 type) family protein (TAIR AT2015743; T); Has 18 Baal hits to 18 proteins in 4 aperiter: Archae - 0; Becteria - 3; Metazoa - 4; Pung - 3; Planta - 14; Vinueet - 0; Other Eukerystes - 5 (Insurse: NC8) 86,740;

Locus: AT	1G09570									
Date last modified	2003-05-02									
TAIR Accession	Locus 2012300									
Representative Gene Model Ø	AT \$509528.1	19699578.1								
Geras Model Type	protein_coding		free text annotations							
Other manager:	ELONOATED HIP FH/2, FRE1, H/R	DOOTYL 8, F14,8533, F1 PHYA, PHYTOCHROME	AUR_22, FAR RED EL. COTTL 2,							
Description Ø	Light-table cytopla inter-convertible for chromophore, and imadiance respon- de-etiplation, invol-	emic red/lanced light pho mu: Pr and Pft (active) a the C deminus is involve te (HIP). Major regulator web in praethopism and p	oforeceptor involved in the regulation of photomorphogenesis. It exists in two nd functions as a diriver The N terminus carries a single tetrapersie ed in dimerization. It is the sole photoreceptor mediating the FR high in red-light induction of phototropic entrancement. Involved in the regulation of indufropism. Requires FVICt for nuclear accumutation.							
Other Gene Models Ø	AT1009570.2 jtplcs varients									
Map Detail Image	Protein Co without a	Soles. Sol Ling Gree Models	controlled vocabulary annotations							
Anesotations @	Calendar	Relationship Taxe 0	Newsyst 9							
	00 Diological Process	involved in	gravitropiern, phototropiern, regulation of transcription, DNA- dependent, response to arounic, red light signaling pathway, response to very low flyance red light stimulus, response to continuous far red light stimulus to the high-mediance response system							
		Helabert 15	response to tar red light							
	Component	tocated to	rytoplanty, nullear body, nucleys							
	00 Nolecular Function	tes.	D-problem coupled photomereptor activity, protein institutione kinaster activity, signal transducer activity, red or far-red light photoseceptor activity							
	Developmental Developmental Blages	expressed during	petal differentiation and expansion stage, 4 anthesis, 4 last senescence stage, C globular stage, E expanded cotytedon stage, F reakare endered stage, UP 02 heo leaves visible, LP 04 four reaves visible, UP 38 as leaves visible, UP 08 eight leaves multir, UP 10 ten leaves visible, UP 12 levelve leaves visible							
		equipadin	D bilateral stage							
	Pierl thuckers	equesed in	male pametophete, seed, embryo, cobrection, topocobil, root, shoot, Rower, stamen, carpel, sepal, petal, inforescence mentitem, pedicel, test, teal aper, petiole, dem, cauline test, shoot spez, teal temina base, leaf whom							

2. Getting data: GO / PO annotations

Antorizations	TAIR Annotation Search (Hea) Your query for annotations based on the locus ATT008570 resulted in 54 records.									
	Distanting 1 - 25-6634 records on page 1 of 3 pages.									
	Licen	Anasoluted Genus of Locies	Postationschap Type:	Riegwood	Report Category	Eindence Code R Eddence Description & Eddence Wills Eddensce Ø	Associated Dy- Date Lost Medited			
	ATTORNE	EAT-000578	index (angenera.	Brolopt al process	chlamed thors mutant privericijour, amarystis of visitier trads none; Comrill MJ, et al. (2001).	The Arebiduares Instantiation Resources 2005-06-29			
	A710096	NATIONALIS	model a	Problem of	process.	Inferred thirts mailent phenolyphe amatysis of visible that home connel, all al (2001)	The Arabiditions Information Pressure of 2005-00-28			
	A710890	147138801	involved m	liegulaten d' transionption, Chuk-dependent:	anti-plat groupse	indexed Rom adquesce (in attactually revealed Tricky computational analysis) indexProj RN001204 Trick Analytics annotation team (2005-02-17)	The involute for Genores: Palee auchi 2003-09-19			
	APTOONS	NATIONISTI	anned a	Internation in accesso	process:	Internet tions multient Internetionet analysis of phytoologic at rengement tionet burg, et al. (2007)	The Arabidopeix extension Theodorest 2007-04-00			
	A210696	NATION CITATI	itedead m	mappines to very low fluence and light attors	tudiogii et lun primera	indened from mutanz phenotypes analyzes of physiotops at responses some: transportes, at at, ct att7)	The Arabidopsis Estornation Pressurer 2003-11-14			

2. Getting data: getting the plant lines you want



2. Getting data: expression

What is known about this gene's expression pattern?

Associations @	Category	Relationship Type @	Keywood @				
	00 Dological Process	moled in	gravitropism, phototopism, regulation of transcription, DNA-dependent, response to arounic, red light signaling pathway, response to very low fluence red light stimulus, response to continuous far red light stimulus i the high-mediance response system.				
		related to	responde to tar red light				
	00 Cettular Component	tocated in	rytoplanty, mattear bods, Huckeys				
	00 Molecular Function	hai	D-problem coupled photoreceptor activity, protein instidute kinase activity, signal transducer activity, red or far-red light photoseceptor activity				
	Overvells and Developmental Stages	expressed during	petal differentiation and expansion stage, 4 anthesis, 4 last seneticence stage, C globular stage, E expanded ophredon stage, F mature enderso stage, UP 02 two leaves insible, UP 04 four reaves insible, UP 38 ais leaves insible, UP 58 eight leaves mobilit, UP 18 ten leaves insible, UP 12 Swetve teaves visible				
		equipadia	D bilateral steps				
	Pieri shuchan	equinaed in	male pamelophete, seed, embryo, colyredon, topocolel, root, shoot, flower, stamen, carpel, sepal, petal, inforescence mension, pedicel, leaf, leaf apex, peticle, dem, cauline leaf, shoot spear, leaf lamina trace, leaf when				
			Atmotiation Detail				
and the second							

2. Getting data: more expression data via external links

What else is known about this gene's expression pattern?



2. Getting data: more expression data via external links



Arabidopsis eFP Browser expression data

2. Getting data: gene model page

Locus: AT	1G09570			
Date last modified	2003-05-02			
TAIR Accession	Locus 2012300			
Representative Viewe Model Ø	4710309578.1			
Geran Model Type	protein_coding			
Other manager:	ELONIGATED HIT FHY2, FRE1, HIT	POCOTYL 8, F14,833, F1 PHYA, PHYTOCHROME	UIL_23, FAR RED ELONGATED 1, FAR RED A	ELONGATED HIPOCOTYL 2,
Description @	Light-table cytopla inter-convertise fo chromophore, and imadiance respon- de-eticlation. Invol	ermic ved/fail-red light pho muc. Pr and Pth (active) as the C demninus is involve to (HIR). Major regulator wed in gravitropism and p	toreceptor involvent in the regulation of photon of functions as a direct. The Niterminus carrie d in dimerization, it is the sole photoreceptor nired-light induction of phototropic enhancem hototropism. Requires FH/11 for nuclear accu	norphogenesis it exists in two is a single tetrapyrole mediating the FR high ent. Involved in the regulation o mutation.
Officer Games Modella Ø	AT1009570.3 Joplice variants			
Map Detail Image	Protein Co stilloutine.1	ing Gene Rodels	N 3098 3099	30.00k
Ausotations 9	Category	Felationship Type Ø	Repared 9	
	00 Diological Process	insted in	gravitopiern, phototopiern, regulation of tra- response to arcenic, red light signaling path fluence red light stimulus, response to conti- the high-intediance response protein	varingBon, DNA-dependent, wax, response to very low numer. Far red light sterrulus by

2. Getting data: model-specific data

Protein: AT10	Surgeon a	The set of the second
Properties Data last avoidfied 0 Table Accession 0 Extended Do 0 Properties	Send to WURLAST	 MININESTUDI ENGINEERA RELACTIVILA REAGINETESI DISCUTTIVIE MININESCE PRESIDENTITI URBERGINE CONTELLAD EXTENSION MININESCE PRESIDENTITI URBERGINE CONTENTS CONTENTS MININESCE PRESIDENTI URBERGINE CONTENTS MININESCE PRESIDENTI URBERGINE MININESCE PRESIDENTI MININES
	Associated Genes	Long Model O. Locost O. Suppressed Bull.
	Extremal Lank	The Subcellular Location of Proteins in Arabitopett Cutatives (SUSA) - Arabidopole Protein Phraphorylation Site Cutativese (PhraphAb)
	More Community @ tideness only the specific record community by default)	Add My Commant Hide Commants Show All Commants

Sometimes, one gene isn't enough

- Scientists often want to work with more than one gene or protein that are related through some common feature
- TAIR offer some basic tools to create and/or enhance these customized data sets

3. Customized data sets: case studies

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

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

- Get a list of all the genes in the mapping interval involved in "flower development"
- Download all the associated GO and PO terms for the candidate genes

3. Customized data sets: Search:Genes

Get a list of all the c	Search by Name or Phenot	Nb+ 0		
	Gene nome	charts with 💌		
		(barring the input hos blank self reference art article)		
69	Becaulo els coloristica ganaria			
tair	Search by Associated Key	word @		
	Report Territ	containe 🛩 flower development		
Search Interne Tools	GO PO ID Instact mantch redyn			
Sealch Oversee	Reyword Type	GÓ Molecular Function		
Cond. Change		GO Cellular Component		
Destaura	Eddard O	Asstory		
- Lo de genere	NES9036	interved from genetic interaction		
Genera		intered from material planets pa interestion		
40 Appendiations			-	
Britywood dis	Restrict by Features @			
Locas Motory	General Mondeil Trainer O	Iransposable element game		
Manage		protein coding		
Bitronari av Chennet		ribosomel me	-	
The second se	Advanced	gene structure predicted	m.	
		is sequenced		
THE POWER BY CODE OF LESS	Data Decisiona D	is not sequenced	<u>M</u>	
Pergentate	Inter Production . W	C only search test 2 months		
Polynovaplanes, Allenes	Restrict by Map Locations	0		
Province				
Physics of the	Martine Q	1		
Page and a	Range Q	heteren w	-	Contex 1
Frank, denter		and the second s		- Contraction -

pment"

3. Customized data sets: Search:Genes

TAR Gens 1	Search Results	<u> </u>					
new second. New game search		download	download all download all results of		download checked the boxes below and download results		
You TAIR dese dese dese Dise Gene dese Gene dese dese dese dese dese dese dese d	Accession (2007145 (2007709 (200864 (2010345 (2011455 (2011455 (201765) (201765) (201765) (201765) (201765) (2019942 (2019942) (2029942) (2020411 (2024097 (202609) (2027000)	Locus Gene aT1049490 aT1049270 aT1012870 aT1012870 aT1052740 aT1052740 aT1052740 aT1051640 aT1051640 aT1051640 aT1064520 aT1010670 aT1010670 aT1053990 aT1070510 aT1062360	Hodal Gen aT1049490.5 aT1049270.5 aT1012820.5 aT1012820.5 aT1052740.5 aT1052740.5 aT1054400.5 aT1051660.5 aT1051660.5 aT102920.5 aT1004400.5 aT1064520.5 aT1064520.5 aT1069990.5 aT1062360.5	e Type peot peot peot peot peot peot peot pe	Seaccipt tein_coding	ion Other Name (Type) Encodes a senser of the NAC encodes a protein similar to dery1-tHNA synthetase targe AFES (ATIIN SIGNALING F-BOS Encodes HTAS, a histone HIA ubiquitin carrier protein Encodes SUF4 (SUPPRESSOR of Encodes a sitogen-activated AUSI encodes a protein simil Flue light receptor mediati FNN12A (RECULATORY FARTICLE One of the three genes enco Encodes a ubiquitin-specifi Encodes a ubiquitin-specifi Encodes a tigediation of the Amember of class I knotted Class I knotted-like homeou	Keywords transcription fo o G-coupled recep red to chioropian (); excin bindla ();
	(TARAT2	ils Balana) 046873				ANNOT SWOT	

3. Customized data sets: Search:Genes

Gene search

Function: identify a list of genes (that have certain function) (located in a specific region) (additional parameters)
3. Customized data sets

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

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

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

Q: What are the potential target genes that might be regulated by this TF?

- Find all genes with similar DNA sequences in their upstream regions
- Identify all domains found in the encoded proteins

3. Customized data sets: Download:Bulk Data Retrieval:Sequences

Get upstream "promoter" sequences

Sequences Local ModTern or Seguences	#74035790 #74013850	AGI transcripts AGI coding sequences AGI genomic locus sequences AGI protein sequences Upsteam Sequences - 500bp	
SATISSIAND Chritilizeeneele CCOTCOTCORRODADOCTOSOTTAGOTOS GCARRAAAGCOTTRATTGTTCATCTACO GCARRAAAGCOTTRATTGTTCATCTACO TOTOSTATTTTAAATOTTTTAGGTRAAAAT AATARACOTRAATGRAAACRAAGABOTOST CGARTTTCTTATTOGCAARGCATTTRAATT CCATCCAARRAGARGTATTTTGAAGGTTTC TCCCRAATTTGTAATCCTTTOGCATTTG AATTTATCATTTGTAATCCTTTOGCAARGCAC TGACGTTTTGGCAARGTATTTGAARACAC ATTTATCATTTGCAARGCCTTCACARGCAC SATGSCAGTRCACOCRACOGCTGARGCAAR SATGSCAGTRCACOCRACOGCTGARGCAAR SATGSCAGTRCACOCRACOGCTGARGCAAR SATGSCAGTRCACOCRACOGCTGARGCAAR SATGSCAGTRCACOCRACOGCTGARGCAAR SATGSCAGTRCACOCRACOGCTGARGCAAR SATGSCAGTRCACOCRACOGCTGARGCAAR SATGSCAGTACARGCATGTGCATRATTRAATT TAAGGTCCARACTGGATCTGTTATTTGAC ACCARGCTTRGTGATCTGTTATATACGTT ACARTCTTCTTATTGTGACCCCATTTAC CTAARACGTTRGTGATCTGTTATATACGTT AAARGCTTRGTGATCTGTTGATATTCGTTAARGTT ACARTCTTCTTATTTCCACARCCCRTTTAC	658963 FORWARD AOTTTRICTGALACATC CTTTREATCTGALACATC CTTTREATCTGGTATEGA SOTTRACTGGTATEGA SAGARTTARSGATTCTT TATTREAGTGGTTARGA RAAGCTACGATTCTTT TATACCATACTGCALACT CATOCITGCTATCTTTA SACARCALGECALATTO CATOCITGGTATCTCTC AGATTTGAGGATTCTC AGATTTGAGGATTCACT ACCARCATGCACTTETT TCGAGGALAGECGATCCA ATOTRARGCCRGAGGATCCA CTAGCACGAGGGTCCACG	TCTGATACGAGOTCTCATETACAAAGGAAGATGATG ATAAAAGGGATTTTGGGTTTTGATGTTAATTTTTT TTTBGGCABGACTATGGGTTTTGGTGTGAAGTGTGG GCACCTTAGGCTATGAGGATGAGGTGGCACAAATGAT TTGGTATGGTTATGAGGATGAGGTGGCACAAATGAT GGTAGGATGTTATGAGGATGAGGTAGCACAAATGAT GGTAGGATCTACGAAGGTGAGCAACAAATGAT GATBACACBCTATGGAAAGCAACAACAAAAGCAAC AATBACACBCTATTGGAAACCAATTTTGGTBGATCAAC GTCCTTCTGTGGAAAATTATOACTTGGCBGTGGATTAA AATBACACBCTATTGGAAACTATTTGGTBGATCAAC GTCCTTCAGGAATATTTCTTCGAAAGGAACAAAAACCAA CABCTTTTTTTACAAGTAGGAACGACTAATTTACTAC TCCTCTCAGGAATATTTCTTCGAAGGATGAACAAAAACCAA CABCTTTTTTACAAGTAGGAACGACTAATTTACTAC TACOGTTCBAGAAAATTATGGTGCAGGCCCAGCAACAAAACCAA CABCTTTTTTACAAGTTGGTGCAGGCCCAGCAACAAAACCAA CABCTTTTTTACAAGTTGGTGCAGGCCCAGCAACAAAACCAA CABCTTTTTTACAAGTTGGTGCAGGCCCAGCAACAAAACCAA CABCTTTTTTACAAGTTGGTGCAGGCCCAGCAACAAAACCAA CABCTTTTTTACAAGTTTGGTGCAGGCCCCAGCAAGAACAAACCAA CABCTTCCCAGGTGAACTTCGTTGAGATATTTGAAGAT TTCATATATAACACCACACAAAAAGTTATTTTATCATCTTT GTGTCATACCACTTGGTTTGGCTTATTTATCATCTCCCAAAAAAAA	AATOTTGACCOATOCTT TATTTTATTTTAACAAG AARACGTBARCTGTCAA CATRTOOTTTAGRCCAA GARCTTCTBAATTCTAA CATRGAGCCTCATTROA CATRGAGCCTCATTROA DCTBAAATOGAGCTGTT ACCAACACATTTTERTT TGGCACTGGTGGTAABCAA TCTATCGGTGTAAGGTAGG TERTTCGGTGGTGGGAAA TGTTCCACTTGTGGGAAA CGCGCCCCTTACACGCGT AGTGAAARCTAACACAA TCTRTCTCTCTCTCTCCAC

3. Customized data sets: Tools:Motif Analysis

Search for over-represented or prevalent DNA sequences in "promoters"
 Use the Motif Analyzer in TAIR to identify common 6-mers

tair -	sligeffer Absolute number of this colgoffer in goery set Absolute number in genomic set Fundet of sequences in query set containing sligeffer Funder of sequences just of 31281 in genumic set; containing sligeffer p-value form bihomial distribution						(M) Annua (Annua)	
Statistical Mo	dreed, and	Parts -	C#4 2084	Tainik:	t this oldge	Mei .		
The program comp strands) with the to require as, each topic in sets is AOI	CATOCC	70	9574	17/12	4995/3282	7.35e-04	ATTO482TO ATTO12820 ATTO52T40 ATTO14400 ATTO4400 ATTO1710 ATTO5980 ATTO45050 ATTO28250 ATTO43850 ATTO25358 ATTO65400 ATTO68000 ATTO7830 ATTO81045 ATT079000 ATTO68100	percept engine
kots hander is for han al other desc nore than once in	00CAT0	20	9574	17/62	499533292	7.316-04	AT10482T0 AT1012829 AT1052T48 AT1014408 AT1054800 AT1017110 AT1062990 AT1045850 AT1028280 AT1043950 AT1025258 AT1045880 AT1069900 AT1091839 AT1081046 AT1079800 AT1068100	
AT1G69490 AT1G48270 AT1G11870 AT1G12820	DATCAA	20	27687	1552	\$7238/33282	9,51+-54	AT1068490 AT1051880 AT1022925 AT1018870 AT1017110 AT1062990 AT1070518 AT1026280 AT1043950 AT1047260 AT1059640 AT1078000 AT1015716 AT1026630 AT1012408 AT1088180	
	TTOATC	22	21657	18/62	17230/33282	0.51e-04	ATTIG68490 ATTIG51680 ATTIG22820 ATTIG18670 ATTIG17110 ATTIG51680 ATTIG70515 ATTIG28260 ATTIG42850 ATTIG47260 ATTIG59640 ATTIG78000 ATTIG18710 ATTIG26830 ATTIG13400 ATTIG68160	

3. Customized data sets: Tools:Patmatch

Channes & Rame	anne Database Isl	link and hadd to see
Al justic Arabitrary	to industrial cards to	bund within these daties of
TAPB Loci Up the	eam Sequences - 10	(DP44) 😸
START PATTER	IN SEALCH I AN	ranged Acets
- aller to Provide	and a second second	The second s
PLEASE WAR FOR	EACHIPEOUEST 10-01	INPLETE DEPOPE SUBME
Mare Ontions	1	
Maximum Life 767	00 00	
a Drie liband bot	h uttanicle.	-
Munistry 0 N		
Manual Tree Co.	incention El receture	
the second se		the second se
Meximum Hits par fo	etients 1	-
Meximum Hits per D	equerce 1	-
Meximum Hits per 5 Maximum Hits per 5	equence 1	
Minimum Hits per D Maximum Hits per D	equence 1	-
Minimum Hits per D Maximum Hits per D Supported Path	equerce 1	amples:
Morrow Hits per D Maxmum Hits per D Supported Path	ern Syntax and Ex	amples:
Meximum Hits per D Maximum Hits per D Supported Path	ern Syntax and Ex Character EVLXWAGGYP TSHEDDRER	Lamples: Discretion
Morrow Hits per D Maxmum Hits per D Supported Path	ern Syntax and Ex Character PVLVWAGCVP TSHEDOPER J anset	Exert match Ny feetablic resist (7+C-MM-OCT)
Monmum Hits per D Maximum Hits per D Supported Path Supported Path Supported Path Supported Path	equence 100 errs Systan and Ex Character PVLVMMOGNP TSHEDOHKR J J J	Exert match Exert match Providences (Fricing Strategy Str
Meximum Hits per D Meximum Hits per D Supported Path Search Type Deptide Searches	ere Syntax and Ex Character PVLVMAGCVP TSHEDOPER J asses 0 peet	Exert match Exert match Proy technologics (Frick Medicing) (Frick Medicing
Monmum Hits per D Maximum Hits per D Supported Path Reach Type Poptide Searches	equence 1 equence 100 errs Syntan and Ex Character PVLSWANGCVP TSHEDGHAR J asses 0 asses 8 asses 2 asses	Exert match Exert match Proy technological (Frick Medicine) (Frick Medicin
Meximum Hits per D Maximum Hits per D Supported Path Tearchits Tearches	I UNION I UNIONI U	Amples: Moving Eventmatch Structure of the Structure of t

-	trund	[404							
Deq.	and also with this		490						
549	ances parties	10948							
10,044	searched .	10018254							
P after	r#.		HICVTCOUTNA		a a caracterization				
Cata	oet sawy hez.		TADRE LOUI VEIDE	eath Sequi	n(#8 - 300	0.10 (244)			
			Next Results +						
	10000000	1.57	198	Matching	Peattors	100 Seeperation			
	* Segence same 1	* of Mrs	pattern	start	end				
			THACCONTUT	401	603				
	A) FOILINES	1	TOADOODITA.	121	885	(séduence			
	- Martine		ACCTCOOFTA	401	640	000000			
	*1992039	1	TAACCOADOT	771	78d .	addonu a			
5	AL6003180	ÿ	TCOTCOOTIA	\$1T	821	14queres			
£	x71003610	ŧ	TOADCOURDA	431	459	eequeru e			
F.	x55000120	1	ACOTCOOTIA	801	829	1044-0010			
8	KT1000000	4	TECTEOOTIA	338	349	00000			
9	#THOUSERS	3	ADATOODTIJA	76¥	368	sédowne			
11	x11000179	t	ACOTCOOT7A	368	388	0000000			
11	xf100e1e0	1	TARCORACOT	198	1.8.8	-			
11	ATHO01989	+	Tectootea	201	378	140,000			

3. Customized data sets: Download:Bulk Data Retrieval:Proteins

Iden Ball	ik Protein Se	earch pain	Č.						
6.	wch proteins acco Reset	reding to pres	ticted pro	tein pros	serties, hi	rle: All datase	fs. have bee	Get Protein Diete	Home Belesse (April 20
tair -	Locus	Protein Sequilewer	NW [Dal	14	110 Domuites	Structural Class	Domatos		
hearch In	AT1002640.1	Sequence SV	62984.3	8.4214	1. C	interspersed alphatiets	PR902172 Divohvd	100F52279 ne_3_C098-6293	
Bulk Data Retrieval	6					32-12-12	#19902172	PP01015	
Gene Descriptions Download gene description							PR001764 0U-MDRL PR001764	A8E3(115-134)	
GO Associations							0LH0DRL PR001764	ABE3(202-218) PR00133	
genes and view the res							2990176	PF00833	
Segurices							0308A3.4	2.50.1706(365-628)	
Download nucleotide a sequences to Clustelli							0306A33	(RO-391) 0.20.200(RO-366)	
incrittute	AT10026383	sequence .	20345.8	9,4304			PTHP:033	(25F3(3-3#5)	
Proteins.		200					PR002249	PTHR10202	
Gearch profeins accord class prediction.							PRIOZ251 DERENTR	_manippi_>.340) PR01130 RelPRT(78-101)	
Locate History							@99002248 DEMENTE	PR01930	
Find out how a locus to locus name.							PRIDOZIST DERIENTR	MEPRO1130 MEPRO1303-318	
Mici namay Dements							DERIDNITE	PR21335 INSPR1321-345	
Dearch array allements							PW507258	PP01733	

3. Customized data sets: summary

bulk data retrieval:sequence	list of AGI codes	upstream sequence
motif analyzer	list of AGI codes	DNA motif in promoter
Patmatch	motif (DNA or protein)	list of sequences
Bulk data retrieval:protein	list of AGI codes	protein features

4. Analyze data sets: Functional categorization

Classify your list of the second s	GO annotation search, function	onal categorization and download own Oxtology at TxPr [Help] Th into the leaders and press loss of the submit before before. The identifiers have to
60	the deparated by table, commas, startlage Chicking on Oet all OO winetations will d categorization will group the genes into t	refums or spaces. Alternatively, you can upload a file, same forwalting as for the technol, isptay in detail all the OO envolutions done to your set of genes. Citcling on Functional road functional categories based on the trigh tevel terms in OO hierarchy.
	You may download the whole genome 0	O annotations from TNR FTP site
teach mours by	Whole Genome Categoriza	ation
Segistr Generation	- GRE	Whole Genome Categorization
EMAX Doors	0.0 Annotations	
Genera	Locan Montflier E	AT1049490 AT1048270
GO Apportations		ATUGIZBUO
Bergawat da		AT1652740 AT1634400
Local Metery		AT1630970 AT1651680
Marters		AT1421920 AT1604400
Microsoft og Slemmet	and the second se	AT1064820
Microsoft ay Copyrighted	right and time.	Browse.
Differences by Cope we also		
Propertate	output options:	
Polynerspiritures Alleren	Select onlynd hype:	Serting, O'test Process total Part Frances Than 1070 into any entered, carbon and be conser-
Province		
Weekseeds	Reset Get all GO Annotations	Functional Categorization
Publication	Fage b	ast modified on August 2, 2005 and the data in updated weekly

and a local state of the second second second second	which had been been and seeming and other a break is more than the
aste torus identifiers couch as Artigo e separated by tatric commas, carria ficking on Oet all 010 annotations wit allegorization will group the genes into	10.000 into the foldbox and press one of the submit buttons below. The sterithers have pereturins or spaces. Attenditively, you can uplead a file, came formating as for the b strapility is detail all the OO annotations done to your set of genesi. Clicking on Function stroad functional categories based on the high level terms in OO hierarchy.
oo mie download the whole genome	00 annultations from TAIR FTP site.
Whole Genome Categori	cation
102	
Ove	Whole Genome Categorization
GO Annotations	
Locus Montéliers:	
Upload Nec	Browne.
Output Options:	
Select output type:	@HTML O Text (Phease hole that if more than 1000 loci are entered, only text output will be p

How to effectively use the tools and resources at TAIR to enhance your research

Part III: Tools for analyzing the results of experiments

on an Arabidopsis metabolic map

A.S.Karthikeyan

Everyone is studying metabolism

- Many biological processes connect to metabolism
 - Drought tolerance changes in osmolyte concentrations
 - Hormone signaling biosynthesis and degradation of hormones
 - Photosynthesis chlorophyll production and ROS scavenging
 - Translation amino acid biosynthesis and riboswitching
 - Plant defense phytoalexin synthesis
 - Your favorite process . . .
 - AraCyc can help you find these connections!
 - <u>Ara</u>bidopsis Metabolic En<u>Cyc</u>lopedia
 - <u>Cyc</u>lopedia
 - Database of metabolic pathways found in Arabidopsis
 - www.arabidopsis.org/biocyc/ (TAIR)
 - www.plantcyc.org/ARA (Plant Metabolic Network)



How to Access AraCyc

- AraCyc <u>Ara</u>bidopsis Metabolic En<u>Cyc</u>lopedia
 - Database of metabolic pathways found in Arabidopsis



Accessible from:



- TAIR The Arabidopsis Information Resource
 - www.arabidopsis.org

tair	from the	Daniel Alexand	The page-theory	-		Metabolic R	sharpy Will save
iners its	formers.	2008	1002.00	Partais.	freemant :	- Bulletet	Terres
		Taxas Descent		-			
		deferrance.					
		Despinent.					
		Bigt month			sale		
racyc	100	Articlyr, Manu	abovio: Patrimaion		00		
Linear Array	con .	BLAST :					
Manhallandine A		100.01.0.07					

How to Access AraCyc

AraCyc – <u>Ara</u>bidopsis Metabolic En<u>Cyc</u>lopedia

Database of metabolic pathways found in Arabidopsis



Accessible from:



PMN – Plant Metabolic Network

www.plantcyc.org

PAR	AN abolic Netwo	P		191	000		
About FWN	Databases Databases	Inertach	Turis	Darful Stea	Laborat Data	649	Teelbark
Introduc		on the	Antonia	wiiCys. Arelickopein)	Item of t	he Month	
The Place Lin	Athenes	ed Search	she propert a	nong databasas	The PMM Iau	nches PlantCy	c NB

AraCyc 5.0

Pathways	332
Compounds	2312
Reactions	1928
Genes (in pathways)	1970
Citations	2543

AraCyc Pathway pages

All items can be clicked on to obtain more information



+ Additional curated information

AraCyc Pathway pages

Superclaimer: Descetterin > Eats Acids and Lands > O	atte Bezzitteri	Classification		
General Information: Choirs is a fundamental metabol non-plactic plant membranes [Min(Q_) Choirse is also is accumulated and confers also tolerance to salinity, do	itte in plants bacausa of its contr a precursor for the formation of gl rought, and other environmental of	Button to the synthesis of the mend yoine betaine (glycline ketalos bi n teoses. In ediblicon choline has been	rane phospholipid phosphatidylch cynthesia III (plannia) in cartain recognizad as an essential rubs	olone, which accounts for 40 to 60% of lipids in plants such as spinach, where this correspondentant and for humans (<u>McNat01</u>).
The choice biosynthetic pathway enables plants to decr to environmental conditions where large and variable and	ouple choine synthesis from laid ounts of choine are beneficial for	metabolism (Herinedy pathway - 15 survivel PantersC1_]	erafishcored biorgethesis.) and	provides them with the metabolic flexibility to adapt
Pathway internation: The first step in choice bioryst Ethenolamine is widely recognized as the entrance com-	hesis is the direct decarboxylation pound to choline bioxynthesis	s of serine to ethenologicne (<u>Rocker</u>	(1), which is catalyzed by a set	ne decarboxylare unique to plants <u>RomenS2.</u>]
The pathway variant displayed (nucleotide pathway) repr where three consecutive N-methylation steps are carried [] [Ma81.]	esents the biosynthetic rode as I I sul either on hee-bases [Prode	found in dwarke plant families. The s commet(2), phospho-bases (Naccid)	ysthesis of choine from ethanola []], phospheliclyi-bases [McMail	mine may take place at three parallel pathways, 21 or a minime of the latter <u>Darkottin Darkottin</u>
The synthesis of intermediates on both the plucipho-bes out that the synthesis of phosphatolytethansterime and assumed that the reactions embedded in the nucleotide kinney/011	e and phosphatidyl-base level no phosphatidyl-bolie is characteri and methylation pathways may b	lodes the nucleotide pathway via CD and by a high degree of interaction a setwo characteristics of one overall (P-phospheamsoulicithel and the nd funcation in the rannos levels hosphoamnealcohol pathway for	methylation pathway. However, It has been pointed of arising intermediates. Consequently, it has been the formation of phosphatidylcholese in plants [
The release of choice from the different pathway levels in into phosphaticlytchoine with the subsequent minese of bean [<u>Wanght</u>]. Athough a well-defined physiological distinct cellular functions [<u>Vol.09</u>]. The remaining oncy substrates. This may be beneficial to process the hetero	a also apecies-specific. Phospho choline, as in tobacco (McMal0) use of phosphologics (O(LD) still mes methed in this pathway, pho operatous mixture of possible sub-	choime can either be directly depton 1. The latter reaction has been sheen awaid further research, progress ha sphoarminicational cyticity/therefera sheaten but it also indicates that the	phorylated to reference choline as in to be specifically catalyzed by a been made to actage some me as and COP aminoalcohol phosph pathway flux is probably controlle	observed in spinach [Summer(Q)] or incorporated phrspholpase D (phrspholpases) in cartor obses of the haterogeneous tarsity of PLD's to obtanetistase, cover a troader spectrum of d more upstream [Veney(Q)]
Experiation in a substitution of choice becauthers	Superpathv	ways		Summary
Undication Links: MataCys (NWY 1981) Legend for Pathway Diagram	Pathway var	riants		
If an analysise name is shown in bold, there is experimental w	sidence for this enzymatic activity	n)		
References				
Datkett, Datio Art, Mold Sci. (1995). "Estemps of about-	ndykholme, earthwais in Lenna,	atabase, and carst." Plant Physiol	113001.00.1330-1540.	References
Datistica Date Art, Mult Str. (1981) "Phosphatelytcholog	e systhesis. Offering pidlems in s	andram and camp." Plant Physical o	1988.00.854.001	
Hits81 Hits WD, Shodes D, Hansen AD, 119675 "Redetrace 68, 814-822	a widence implicating phosphory?	and phosphaticit bases as interna-	Sales in betane synthesis by wi	the streated batter leaves." Flast Physiol. (1981),

OMICS Viewer:

Analyzing the results of experiments

on an Arabidopsis metabolic map

OMICS Viewer to the rescue . . .

- Overlay experimental data on a metabolic map
 - Transcript data (for enzymes)
 - Proteomic data (for enzymes)
 - Metabolomic data
- Use multiple data sources
 - Your data
 - Publicly available data
 - Gene Expression Omnibus
 - NASC Proteomics Database
 - NSF2010 Metabolomics
 - Many more . . .
- Generate new testable hypotheses about your favorite
 - gene
 - metabolite
 - biological process
 - etc.

Addressing common research conundrums . . .

- Compare transcript levels / protein levels / metabolite levels in wildtype and mutant plants using the OMICS viewer
- Look for "hidden" perturbations in metabolism
- Use the OMICS viewer to compare data from WT and mutant plants
- Download publicly available data sets related to the biological process
- Scan for affected areas of metabolism
- Use the OMICS Viewer to quickly identify metabolic <u>pathways</u> that are up- or down-regulated
- And many more . . .

OMICS Viewer Overview



OMICS Viewer Overview



Viewing gene/protein families



Getting more information



AraCyc Pathway pages

All items can be clicked on to obtain more information



+ Additional curated information

AraCyc Pathway pages

Superclasses: Descetterin > Eats Acels and Lasts > D Summary:	toble Beconterin	Classification		
General Information: Choirs is a fundamental metal- non-plactic plant membranes [Most2]. Chaine is also to accumulated and confers also tolerance to salinity, d	otte in plants because of its contri- a precursor for the formation of giv rought, and other environmental ob	button to the synthesis of the mend cone betaine (glycine ketalne bir roses in edition choine has been	rane phospholipid phosphabelych synthesis III (plants) in certain recognized as an essential rubs	olese, which accounts for 40 to 60% of lipids in plants such as spinach, where this corroprotectant ent for humans [<u>McNat01</u>]
The choline biotyrithetic pathway enables plants to dec to environmental conditions where large and variable am	pupie choine synthesis from lipid ounts of choine we beneficial for	matabolion (Hannedy pathway - Ma symbol (<u>RontanG1 </u>	erafishcorrol biosynthesis.) and	provides them with the methodic flexibility to adapt
Pathway information: The first step in choine bioryst Ethenolemine is widely recognized as the entrance com	here is the direct decarboxylation pound to choire bioxyrthese.	of serine to athendiamine Eastern	[1], which is catalyzed by a serv	ne decueboxylare unique to plants <u>Romanil2</u> .1
The pathway variant displayed (nucleotide pathway) rep where three consecutive N-methylation steps are carried [] <u>Midd1_</u>]	resents the biosynthetic route as f f out either on hee-bases [<u>Procho</u>	ound in diverse plant families. The s mmetry 1, phospho-beams (Naccod)	ysthesis of choline from ethancian []. phospheticlyl-bases [<u>Mr.Mark</u>	nine may take place at three parallel pathways, <u>[1]</u> or a minimum of the latter [<u>Darkoffin]</u> [<u>Darkoffi</u>]
The synthesis of internediates on both the phospho-bis out that the synthesis of phosphatidylethanularise and assumed that the machine embedded in the nucleotide <u>kinew/00</u>].	ie and phosphatidyl-base level incl phosphatidyl/thdine is characteria and metholation pathenys may b	lodes the nucleotide pathway via CD and by a high degree of interaction a a two characteristics of one overall (P phospheameroalcohol and the r nd furcation on the various levels o phosphoamenalcohol pathway for	methylation pathway However, it has been pointed of arising intermediates. Consequently, it has been the formation of phosphatidylcholine in plante [
The release of choline from the different pathway levels into phosphatidylcholine with the subsequent release of bean [Wanght] Although a well-defined physiological distinct cellular functions [<u>Virk09</u>]. The remaining onzy substrates. This may be beneficial to process the heter	a also species specific. Phosphoc choline, as in 505ecco [<u>McMar00</u> role of phospholipase D (PUD) still mes miched in this pathway, pho- ogeneous micture of possible subo	holime can either be directly deptor]. The latter reaction has been shor await further research, progress ha uphoaminosicohol cytickylyttransfera metes but it also indicates that the	phonylated to release choline as o in to be specifically catalyzed by a been made to assign some mer se and COP-aminoalcohol phosph pathwey flux is probably controlle	observed in spinach Summer(Q) or incorporated phrsphologiese D (phrsphologieses) in captor obsers of the hateroperature barries of PLD's to obtanetiesase, cover a broader spectrum of d more upstream [<u>Hence</u> (Q)]
Superpatheous assessments of choice bosontheous - Variants: chaine bosontheous R, chaine bosontheous (Superpathv	vays		Summarv
Unification Links: MataCys (WWY 3881) Legend for Pathway Diagram	Pathway var	iants		
If an analysise name is shown in bold, there is experimental e	vidence for this enzymatic activity	2		
References				
Daskelli, Darko Art, Mudd Sit, (1995), "Estamen, of about	endylcholme, sunthexis in Lenne, i	Index, and cash." Place Physiol	.11900.00.1300.1340.	References
Datistia Curve AH, Medd SH, (1989) "Phesphatelytcheler	e systhesis. Offering pidlems in a	rubsies and carsis." Plant Physical of	79885,80,854.061	
Hits81 Hits WD, Phodes D, Harsen AD, 11905 'Redotnes 68, 814-822	n widence implicating phosphory?	and phosphatical bases as interne-	fates in between synthesis by wat	to streated balley leaves " Flast Physics, (1981),

OMICS Viewer Statistics



OMICS Viewer in action

- Multiple data sets can be entered using the same input file
- An animation can show changes in data sets:
 - wild type /mutant a / mutant b / mutant c / . . .
 - time points 0, 1, 2, 3, . . .
 - compound concentration x, y, z, . . .

Suberin Biosynthesis

Wild type



mutant A

mutant B





OMICS Viewer data inputs

- Input •
 - tab-
 - pre —
 - Idei _
 - •

 - ٠
 - Dat
 - Sar _____

filo	Bildermoft Lacal -	presentation.to	1				
me	1 10 BR 300	Accessifi Texit presentations tot 2 Dec Dec					
dolimit	10.00 0 0 0 0						
-demme	00000	ALC: NO	Magat Without	10000	tig management		1000
-analvz	- M	& sGana			-		
	A CONTRACTOR	0	C	0	e.		0
ntifiers	1 Exciting Metabo	olic Detall					
Canaa	2 Sectorers 0	Course 1	Column 2	Column 3	Column 4	Courses	Course &
Genes	4 Cons	WT	material	Ciretum .	WT + telt	matant1 + salt	Not + Creature
– AGI	6 A72624270	1.20	2.40	1.68	4.00	9.60	6.72
-	6 AT1042970	8.20	0.10	0.07	1.24	0.62	0.43
Enzyme	7 AT1032440	0.30	0.39	0.27	0.54	0.70	0.49
– AGI	8 AT1012960	3.60	9.72	6.00	14.40	17.44	12.25
	9 AT6062920	1.28	0.51	0.36	7.94	3.17	2.22
— EC r	15 AT1012980	4.07	8.34	5.64	7.51	15.01	10.61
Compoi	11 AT5060680	0.24	0.12	0.08	0.96	0.40	0.34
Compod	12 AT6068360	8.60	11.30	7.00	10.72	24.54	17.04
– Enzv	13.AT6008570	0.77	2.08	1.46	1.39	2.74	2.62
	14:A74028380	0.49	0.19	0.13	1.92	0.77	0.54
– Con	15 AT3042980	0.87	1.34	0.94	4.15	0.35	6.82
a in add	16 AT3088880	3.20	1.60	1.12	6.76	2.80	2.02
	17 AT304684.0	4.10	5.33	3.73	12.17	75.82	11.07
absolute	18 AT2049160	0.60	1.62	1.13	1.08	2.92	2.04
	19 AT0004050	0.30	0.16	0.11	1.52	0.61	0.43
	22 ATDOJ1980	0.42	0.04	0.59	2.00	5.21	3.95
mole file	21 AT2026680	0.00	0.05	0.03	0.16	0.00	0.00
npie ille	22 A12036530	0.17	0.22	0.15	1.05	1.37	0.36
	21 AT2030630	1.40	4.00	2.00	2.06	7.19	6.00
	the second se						

OMICS Viewer data upload

TAIR - www.arabidopsis.org

tair		-			_	Gene	W Ases
Seaton	illenets.	Tools	lineka.	Partielle	Deserved	Baterit -	fees .
The Ar	abidopsis in	Toorn Debr Glassese Sequences Management		ana ir parati	and motion ar-	Breaking N	iews stracts inatistie
the complet matabolism matabolism	to partome tregue to partome tregue to game expression utilizations, and in	AreCar No. ISANT	lakalı: Pathereys	product etco	mation, and photogal units Game	and hold your developy for the Montradi mont	meeting e spicionskij

AraCyc - www.arabidopsis.org/biocyc



OMICS Viewer data upload

PMN – www.plantcyc.org

P/ Plant M	MN etabolic Netwo	r ek			Pav	Ope ×	and the second se
firmed PRES	Databases	Buerdoade	Tools	Deefad Silves	Tudenti Data	nety.	Teedlash
Introduction		Netabolic Maps + OBDCS Viewer +		Analys (Analosis)			
The Plant Metabolic Network (PMRI) is a cultable		Pathway Tools Biorge Bundle		Tutari	-	1.0	
and biocher pathway dat plant bioch ittesture an	nists with a comm tabages. A control emical pathway of st computational	tion goal to build a il feature of the PM Database, containin analyses about the	Curated inter	ruton Nom The	On June 17, web debut, pathways with	2008, Plank containing th informatio	Cyc made its plochemical n from over

OMICS Viewer data upload

Pathway Tools Omics Viewer					
The Pathway Tools Omics Viewer (formerly the Pathway Tools Expression Viewer) paints data values from the user's high-throughput and other experiments onto the Metabolic Overview diagram for an organism. The Omics Viewer can be used for:	Select a dataset: File containing experimental data (NOT a URL): Do you want to display absolute or relative data values?	Arabidopsis thaliana COL 💌 C:\Documents and Setti Browse Relative 👻			

- Select species
- Upload data file
- Enter appropriate parameters
- GO!

New ideas to pursue . . .

Wild type



The transcript levels of several enzymes in the **Iysine biosynthesis pathway** are reduced ~4-fold!

Mutant with "no phenotype"

Mutant with a metabolic phenotype



New hypotheses and experiments



www.arabidopsis.org

curator@arabidopsis.org



www.arabidopsis.org/biocyc

curator@arabidopsis.org



www.plantcyc.org

curator@plantcyc.org

Acknowledgements

AraCyc and TAIR

<u>PI and Co-PI</u> Eva Huala Sue Rhee

Curators:

- Peifen Zhang
- Tanya Berardini
- David Swarbreck
- Kate Dreher
- Rajkumar Sasidharan

Tech Team :

- Bob Muller
- Larry Ploetz
- Raymond Chetty
- Anjo Chi
- Vanessa Kirkup
- Cynthia Lee
- Tom Meyer
- Shanker Singh
- Chris Wilks

Metabolic Pathway Software: - Peter Karp and SRI group





TAIR Workshop II

TAIR, PMN, SGN & Gramene Joint Workshop

When: Tonight at 7:00 PM

Where: 318 A & B