



模式植物阿拉伯芥資料庫(TAIR)與其夥伴們： 核酸分析

陳荷明

March 16, 2023

陳荷明

- 中央研究院 農生中心 副研究員
- 20年生物資訊相關經驗
 - 資料: Microarray, Small RNA, RNA degradome, RNA-Seq, Ribo-seq
 - 物種: 阿拉伯芥、香蕉、水稻、番茄、其它
- 喜好: 東野圭吾推理小說、謎幻樂團



基因體世代

- 生物資料富翁
- 善用公開的生物大數據資料
 - 可以少做實驗，少走冤枉路，可以快點畢業
 - 可以從不同視野，做生物學研究，超越老闆，成為次世代植物科學家

Next Generation of Training for Arabidopsis Researchers: Bioinformatics and Quantitative Biology

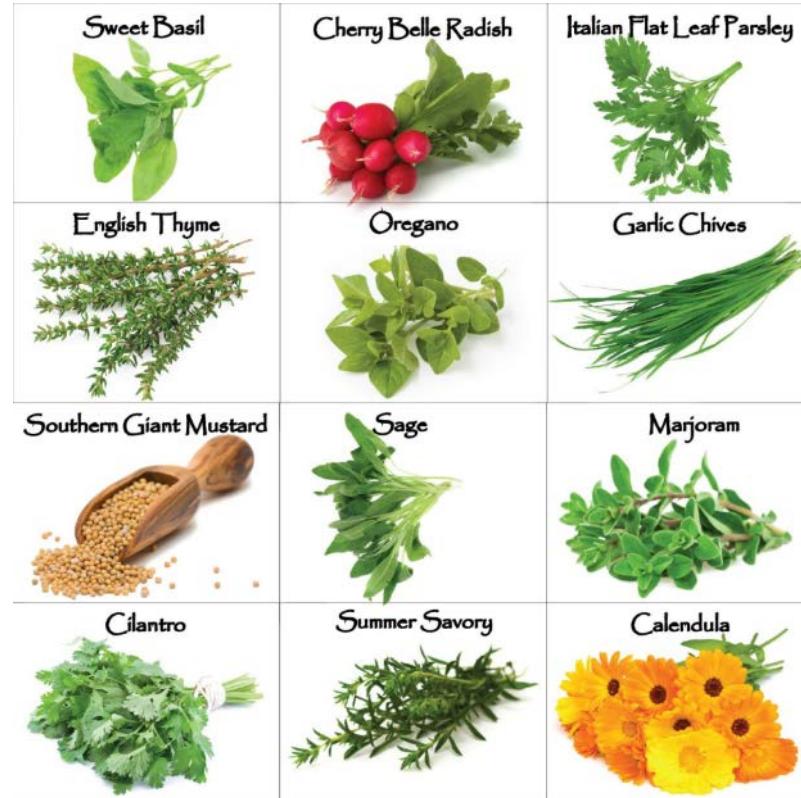
<http://www.plantphysiol.org/content/175/4/1499>

Why Arabidopsis?

Arabidopsis thaliana



Crops or herbs



A plant model organism with

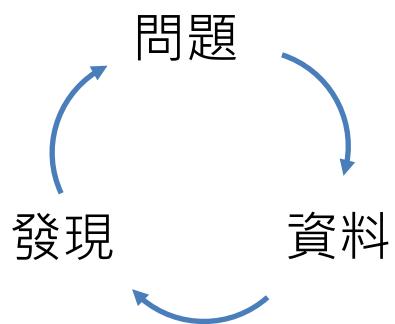
- Small plant size
- Short life cycle
- Small genome size

Either without genome or good annotation

<http://pumene.com/herbs/>

Questions for TAIR

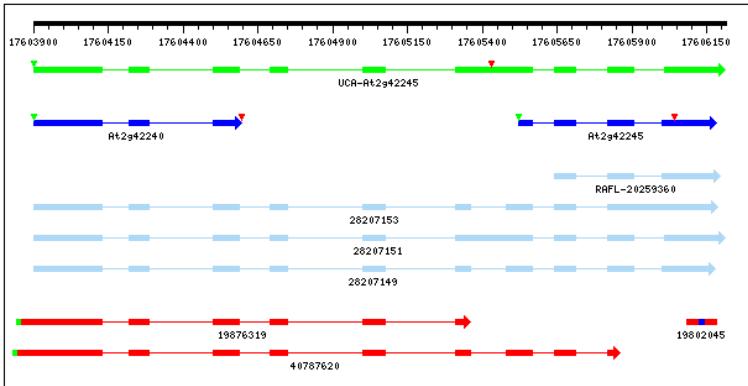
- **To what extent is known for the gene of your interest?**
 - DNA, RNA, protein sequences, isoforms
 - Functions and cellular localization (predicted, supported by experimental evidence)
 - Gene expression (spatial, temporal, response to stress, hormone ...)
 - Others: mutants, small RNAs, RNA degradation, ribosome footprints and DNA methylation
- **Are there other genes having similar features as the gene of your interest?**
 - Similar DNA, RNA, protein sequences
 - Similar functions
 - Similar expression pattern
 -



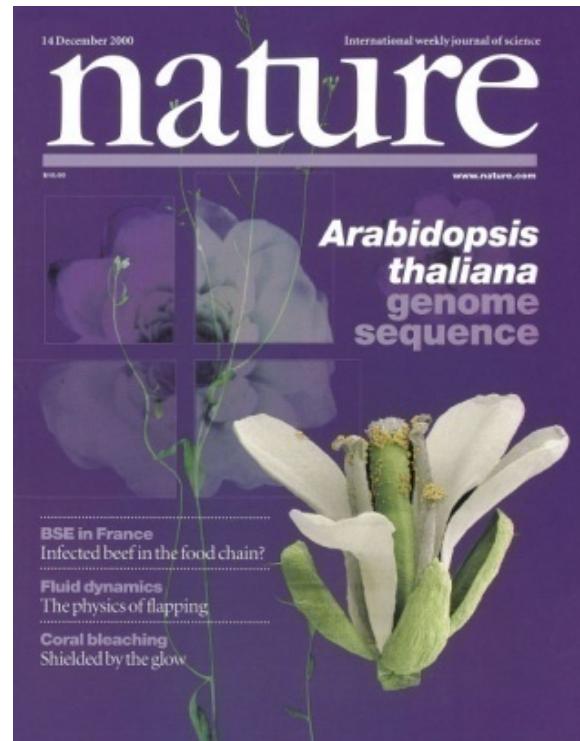
Arabidopsis genome annotation

- Arabidopsis genome sequencing was finished in 2000
- High quality sequence with few gaps

Genome annotation: The process of identifying the locations of genes and all of the coding regions in a genome and determining what those genes do.



<http://www.plantgdb.org/>

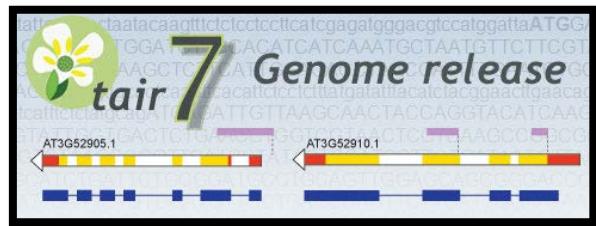


Arabidopsis gene structure annotation

TAIR10: Use new experimental data and new prediction tools to further improve gene structure predictions

TAIR10

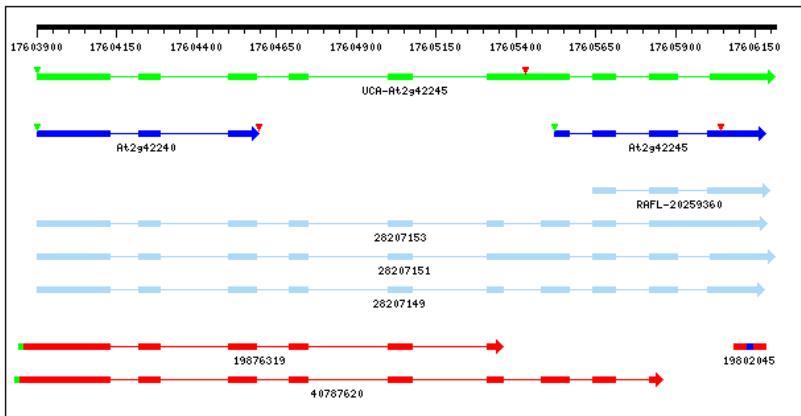
TAIR6-TAIR9: Use ESTs and cDNAs and a assembly tool called PASA to improve gene structures



TAIR10: using proteomics and RNA-seq data to improve genome annotation

4-step process:

1. Mapping RNA seq & Peptides
2. Assembly/Gene built
3. Manual review
4. Integration (genome release/Gbrowse)



<http://www.plantgdb.org/>



<http://mpss.udel.edu/>

All genome versions statistics

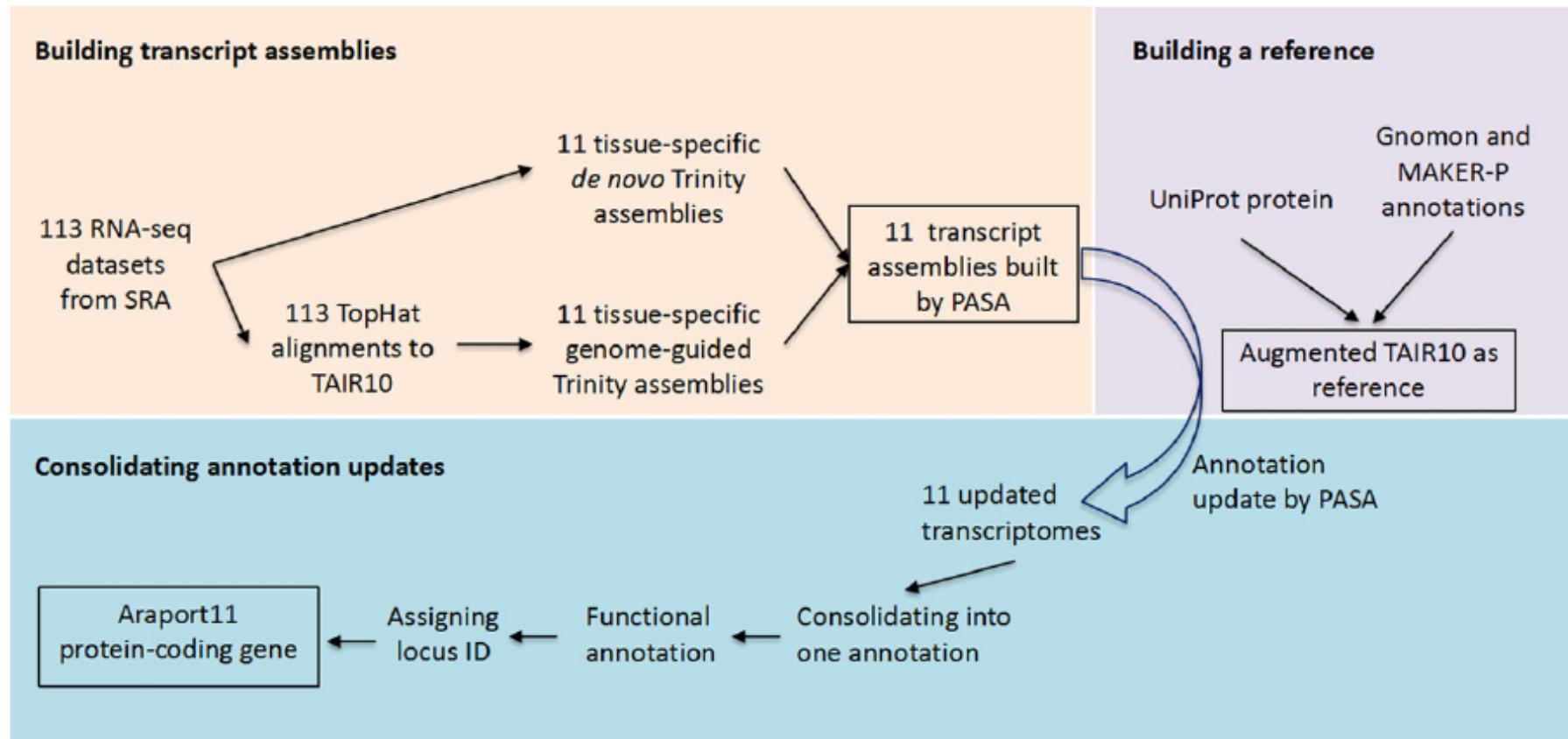
All Genome Versions Statistics							
	Protein Coding Genes	Transposons and pseudogenes	Alternatively spliced genes	Gene density (Kb/gene)	Avg. exons per gene	Avg. exon length	Avg. intron length
TAIR10 (11/10)	27,411	4,827	5,885	4.35	5.89	296	165
TAIR9 (6/09)	27,379	4,827	4,626	4.35	5.67	304	165
TAIR8 (4/08)	27,235	4,759	4,330	4.37	5.62	306	165
TAIR7 (4/07)	26,819	3,889	3,866	4.44	5.79	268	165
TAIR6 (11/05)	26,541	3,818	3,159	4.48	5.64	269	164
TIGR5 (1/04)	26,207	3,786	2,330	4.54	5.42	276	164
TIGR4 (4/03)	27,170	2,218	1,267	4.38	5.31	279	166
TIGR3 (8/02)	27,117	1,967	162	4.32	5.24	266	166
TIGR2 (1/02)	26,156	1,305	28	4.48	5.25	265	167
TIGR1 (8/01)	25,554	1,274	0	4.55	5.23	256	168
Nature (12/00)	25,498	NA	NA	4.50	5.20	250	168



Annotation pipeline for Araport11

The Arabidopsis Information Portal (ARAPORT) has taken over producing the genome releases

<https://apps.araport.org/thalemine/begin.do>



Comparison of TAIR10 and Araport11

Protein coding genes

Type	TAIR10	Araport11	Change
(A) Protein-coding genes			
Total number of loci	27 416	27 655 ^a	+239
Number of transcript isoforms	35 386	48 359	+12 973
Number of loci with two or more splice variants	5804	10 696	+4892
Number of loci with changes in CDS	–	–	+1158
Number of loci with changes in UTR(s)	–	–	+21 298
Upstream open reading frame (uORF)	58	84	+26

which one
should you use ?

Araport11 → Col-0 v12

A. thaliana Col-0 v12 reannotation effort



This site aims to collate information about the A. thaliana Col-0 v.12 reannotation effort. The broad project plan and timeline are shown below. Additional pages linked on the left cover online meeting minutes, in person meeting presentations, and other project updates. Updates are also disseminated through TAIR's Mastodon (@tair@genomic.social) and Twitter (@tair_news) accounts.

[Click for Latest Updates](#)

Do you want to get involved? Comments? Suggestions? Please email curator@arabidopsis.org.

Project goal: To produce an improved structural annotation of the A. thaliana Col-0 genome based on a 2022 community consensus genome assembly

Project phases:



Timeline: Roughly 14 months, starting Nov. 2022

TAIR home Page

<https://www.arabidopsis.org/>

Locus ID or Gene name (eg. At5g06100 or MYB33)

The screenshot shows the TAIR home page with a search bar at the top right containing "Gene". A red arrow points to this search bar. Below the search bar is a dropdown menu with options like Metabolic Pathways, Textpresso full text, and Exact name search. To the left of the search bar is a navigation menu with links for Home, Help, Contact, About Us, Subscribe, Login, Register, and Institution: Academia Sinica. A red arrow also points to the "Browse" link in this menu. The main content area features a large image of an Arabidopsis plant. On the left, there's a sidebar with links to various tools: GO Term Enrichment, JBrowse, GBrowse, Synteny Viewer, Seqviewer, Mapviewer, AraCyc Metabolic Pathways, Integrated Genome Browser, BLAST, Patmatch, Motif Analysis, VxInsight, Java Tree View, Bulk Data Retrieval, and Chromosome Map Tool. A red arrow points to the "GO Term Enrichment" link. On the right side, there's a "Breeding" section with links to various breeding-related resources, and a "Help us identify these 'unknown' genes" section with a message from Aug 7, 2019.

Locus page: name and structure

representative gene model: In TAIR, this is the reference gene model for the locus.
The sequences and structural features for the locus are derived from the representative gene model.

Locus: AT5G06100

Representative Gene Model [AT5G06100.2](#)

Gene Model Type: protein_coding

Other names: ATMYB33, MYB DOMAIN PROTEIN 33, MYB33

Description: Encodes a member of the myb family of transcription factors (MYB33), contains Pfam profile: PF00249 myb DNA-binding domain. Double mutants with MYB65 are male sterile- anthers are small, pollen development is defective. Spatial expression appears to be under the control of miR159, contains a target site for this micro RNA. A highly conserved RNA secondary structure abuts the miR159 binding site which facilitates its regulation by miR159. When the target site is mutated, expression is detected in leaves, roots, anther filament, pistil. The expression of a translational fusion is specific to anther locules in contrast to constructs lacking the miR159 target site. Phenotype is conditional and can be restored by lower temperature or higher light intensity.

What it is

Other Gene Models: AT5G06100.5 AT5G06100.4 AT5G06100.1 AT5G06100.3 (splice variant) (splice variant) (splice variant) different transcription start, end or splicing junction

Center on AT5G06100 | Full-screen view

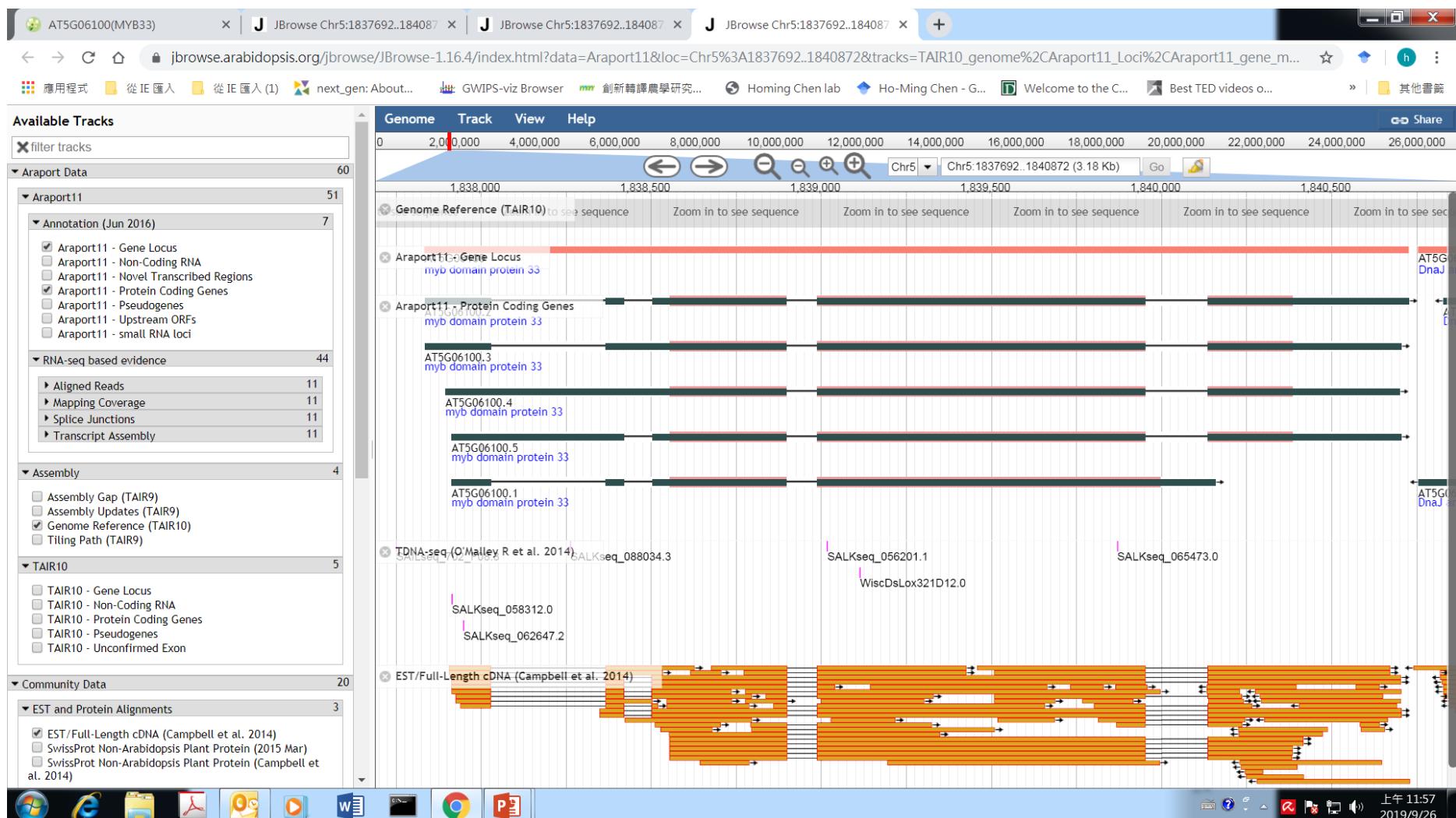
Araport 11

Map Detail Image

JBrowse

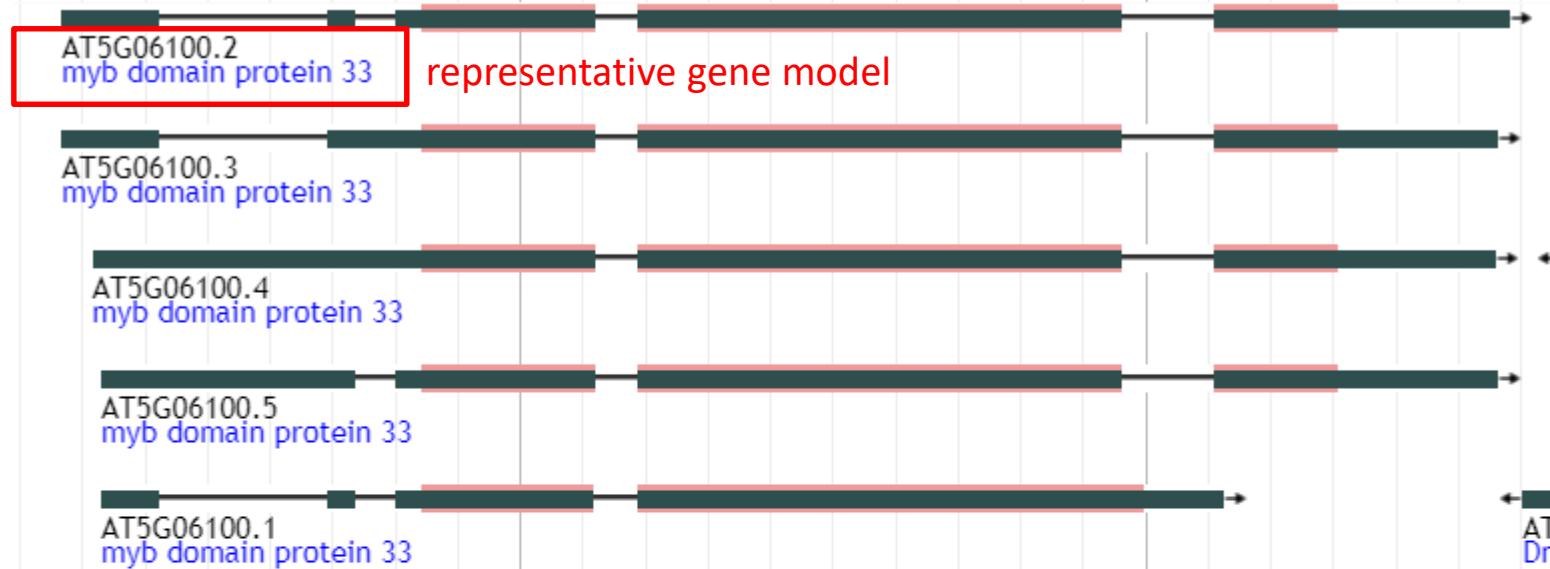
Gene structure: exon- blue box; red frame (coding region)
intron- thin line

JBrowse page: gene model, cDNA, RNA-seq, T-DNA...

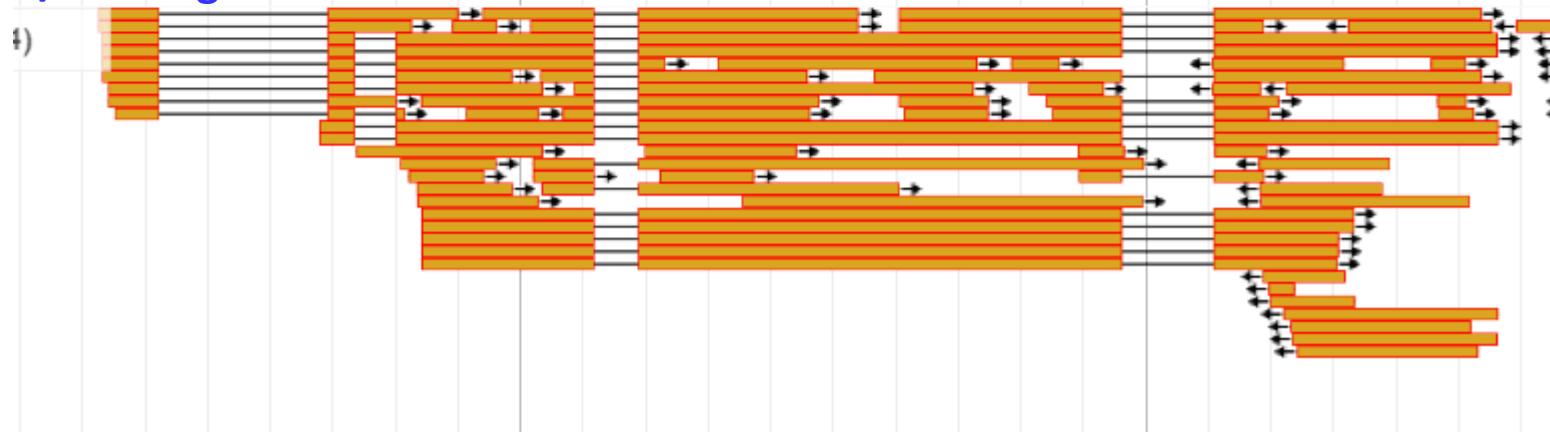


JBrowse page:

representative gene model supported by ESTs/full length cDNA?



ESTs/full length cDNA



Locus page: annotation

Annotations ?		
Gene Ontology (GO)	category	relationship type
	GO Biological Process	involved in
	GO Cellular Component	located in
	GO Molecular Function	functions in
	GO Molecular Function	has
Plant Ontology (PO)	Growth and Developmental Stages	expressed during <i>when</i>
	Plant structure	expressed in <i>where</i>
	Plant structure	expressed only in

Functional annotations with
controlled vocabulary terms :

Gene Ontology (GO) and **Plant Ontology (PO)**
to describe
Genes function , expression and **localization** of
the gene product.

Most gene annotations are made by curators
using the published literature as sources. Some
annotations are based on computational
methods. Detailed information including
references and supporting evidence can be
obtained by clicking on the link to view
Annotation Detail.

Locus page: annotation – Gene Ontology (GO)

◆ MYB33, a transcription factor

Annotations	category	relationship type	keyword
	GO Biological Process	involved in	anther development, anther wall tapetum morphogenesis, cell differentiation, gibberellic acid mediated signaling pathway, negative regulation of cell population proliferation, negative regulation of growth, pollen sperm cell differentiation, positive regulation of abscisic acid-activated signaling pathway, positive regulation of programmed cell death, positive regulation of transcription, DNA-templated, protein storage vacuole organization, response to cytokinin, response to ethylene, response to gibberellin
	GO Cellular Component	located in	nucleus
	GO Molecular Function	has	DNA-binding transcription factor activity, sequence-specific DNA binding, transcription regulatory region DNA binding

- Function in plant development and cellular processes
- Response to treatments
- Localization of the gene product (protein)
- Molecular action and activity of gene product (protein)

Locus page: annotation – Plant Ontology (PO)

◆ MYB33, a transcription factor

Growth and Developmental Stages	expressed during	L mature pollen stage, LP.02 two leaves visible stage, LP.04 four leaves visible stage, LP.06 six leaves visible stage, LP.08 eight leaves visible stage, LP.10 ten leaves visible stage, LP.12 twelve leaves visible stage, flowering stage, mature plant embryo stage, petal differentiation and expansion stage, plant embryo bilateral stage, plant embryo cotyledonary stage, plant embryo globular stage, vascular leaf senescent stage
Plant structure	expressed in	anther, carpel, caulin leaf, collective leaf structure, cotyledon, filament, flower, flower pedicel, guard cell, hypocotyl, inflorescence meristem, leaf apex, leaf lamina base, petal, petiole, plant embryo, pollen, root, seed, sepal, shoot apex, shoot system, stamen, stem, vascular leaf
Plant structure	expressed only in	anther

Annotation Detail



- Expression in developmental stages

Locus page: annotation detail

Locus/Gene Model	Gene Symbol/Full Name	Relationship Type	Keyword	Keyword Category	Evidence Code ⓘ: Evidence Description ⓘ: Evidence With: Reference ⓘ	Annotated By/ Date Last Modified
AT5G06100	MYB33/ MYB DOMAIN PROTEIN 33	involved in	regulation of transcription, DNA-templated	biological process	<i>inferred from sequence or structural similarity:</i> manually reviewed TIGR computational analysis: INTERPRO:IPR001005: TIGR Arabidopsis annotation team (2005-02-17)	The Institute for Genomic Research/ 2002-10-17
AT5G06100	MYB33/ MYB DOMAIN PROTEIN 33	involved in	transcription, DNA-templated	biological process	<i>inferred from electronic annotation:</i> none: UniProtKB-KW:KW-0804: analysis_reference : GO_REF:0000037: Gene Ontology	UniProtKB/ 2017-04-12
AT5G06100	MYB33/ MYB DOMAIN PROTEIN 33	involved in	anther wall tapetum morphogenesis	biological process	<i>inferred from mutant phenotype:</i> none: none: Millar AA, et al. (2005)	UniProtKB/ 2017-02-16
AT5G06100	MYB33/ MYB DOMAIN PROTEIN 33	involved in	response to gibberellin	biological process	<i>inferred from expression pattern:</i> none: none: Gocal, et al. (2001)	UniProtKB/ 2017-02-16



Information under GO term

Keyword:anther wall tapetum morphogenesis

Date last modified	2013-04-18
Keyword	anther wall tapetum morphogenesis
Synonyms	differentiation of tapetal layer; tapetum morphogenesis; tapetal layer morphogenesis
Definition	The process in which the anatomical structures of the anther wall tapetum are generated and organized. The anther wall tapetum is a layer of cells that provides a source of nutrition for the pollen grains as they mature.
Keyword ID	GO:0048655
Keyword Category	biological process
Data Associated To This Term	1 publications 3 annotations 3 loci
Data Associated To Children Terms	2 publications 3 annotations 3 loci
TreeView	TreeView
Similarly Annotated Genes From Other Organisms	GO Database

	Locus	Description
1	AT3G11440	Other names: myb domain protein 65;(source:Araport11)
2	AT5G06100	Other names: myb domain protein 33;(source:Araport11)
3	AT5G22260	Other names: MALE STERILITY 1, MS1 Sporophytic factor controlling anther and pollen degeneration occurs after microspore release and to PHD-finger motif transcription factors.

Enrichment of GO term

GO annotation search, functional categorization and download Gene Ontology at TAIR

[Help]

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

You may download the whole genome GO annotations from TAIR Downloads Site.

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

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

Whole Genome Categorization

Get:

Whole Genome Categorization

GO Annotations

Locus Identifiers:

Upload file:

未選擇任何檔案

Tool for GO Term Enrichment for Plants

Home > Tools > GO Term Enrichment

GO Term Enrichment for Plants

Statistical Over/Under Representation (powered by PANTHER)

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

Enter a list of valid identifiers, separated by newline. [Try a sample gene list](#)

Choose Organism

Arabidopsis thaliana ▾

Choose GO aspect

biological process ▾

biological process

molecular function

cellular component

OR use Advanced settings at PANTHER

Result of GO term enrichment

	Arabidopsis thaliana (REF)	upload 1 (▼ Hierarchy NEW! ?)				
	#	#	expected	Fold Enrichment	+/-	P value
GO biological process complete						
cellular response to hypoxia	239	119	21.19	5.62	+	7.23E-39
↳ cellular response to stress	1215	231	107.72	2.14	+	3.49E-20
↳ cellular response to stimulus	3873	669	343.39	1.95	+	2.60E-56
↳ cellular process	14852	1594	1316.81	1.21	+	2.64E-24
↳ response to stimulus	9304	1387	824.91	1.68	+	3.91E-106
↳ response to stress	5286	958	468.67	2.04	+	1.09E-101
↳ response to hypoxia	323	140	28.64	4.89	+	9.66E-41
↳ response to decreased oxygen levels	331	141	29.35	4.80	+	2.28E-40
↳ response to oxygen levels	333	141	29.52	4.78	+	3.94E-40
↳ response to abiotic stimulus	4114	690	364.76	1.89	+	3.61E-54
↳ cellular response to decreased oxygen levels	241	120	21.37	5.62	+	3.24E-39
↳ cellular response to oxygen levels	242	120	21.46	5.59	+	4.47E-39
↳ cellular response to chemical stimulus	2047	468	181.49	2.58	+	1.28E-66
↳ response to chemical	5053	880	448.01	1.96	+	2.60E-82
response to chitin	308	153	27.31	5.60	+	1.17E-50
↳ response to oxygen-containing compound	3087	605	273.70	2.21	+	4.83E-67
↳ response to organonitrogen compound	648	202	57.45	3.52	+	7.20E-42
↳ response to organic substance	3506	648	310.85	2.08	+	5.23E-64
↳ response to nitrogen compound	790	229	70.04	3.27	+	1.51E-43
regulation of jasmonic acid mediated signaling pathway	45	20	3.99	5.01	+	3.67E-04
↳ regulation of signal transduction	424	79	37.59	2.10	+	4.57E-05

Locus page: DNA and protein sequences

Sequence ?	full length genomic	full length CDS	full length cDNA	protein
Intron		CDS	5'UTR	Protein sequence of
5'UTR			CDS	the representative
CDS			3' UTR	form
3' UTR				

different

- Some gene models don't have 5' UTR or 3' UTR
- Some gene models have incorrect 5' UTR or 3 'UTR
- If 5' UTR or 3' UTR is important to you, compare the annotation with ESTs/full length cDNA or RNA-Seq data in JBrowse

Locus page : RNA data

RNA Data			
Two-channel Arrays	array element name 	avg. log ratio (std. error)	avg. intensity (std. error)
	G2E4	-0.047 (0.038)	2811.288 (106.995)
One-channel Arrays	array element name 	avg. signal intensity (std. error)	avg. signal percentile (std. error)
	18125_AT	4.646 (1.558)	17.63 (0.993)
	250710_AT	67.877 (0.818)	46.334 (0.191)
	18746_F_AT	12.832 (1.061)	27.312 (0.806)
Associated Transcripts 	type	number associated	
	EST	(21)	
	cDNA	(6)	

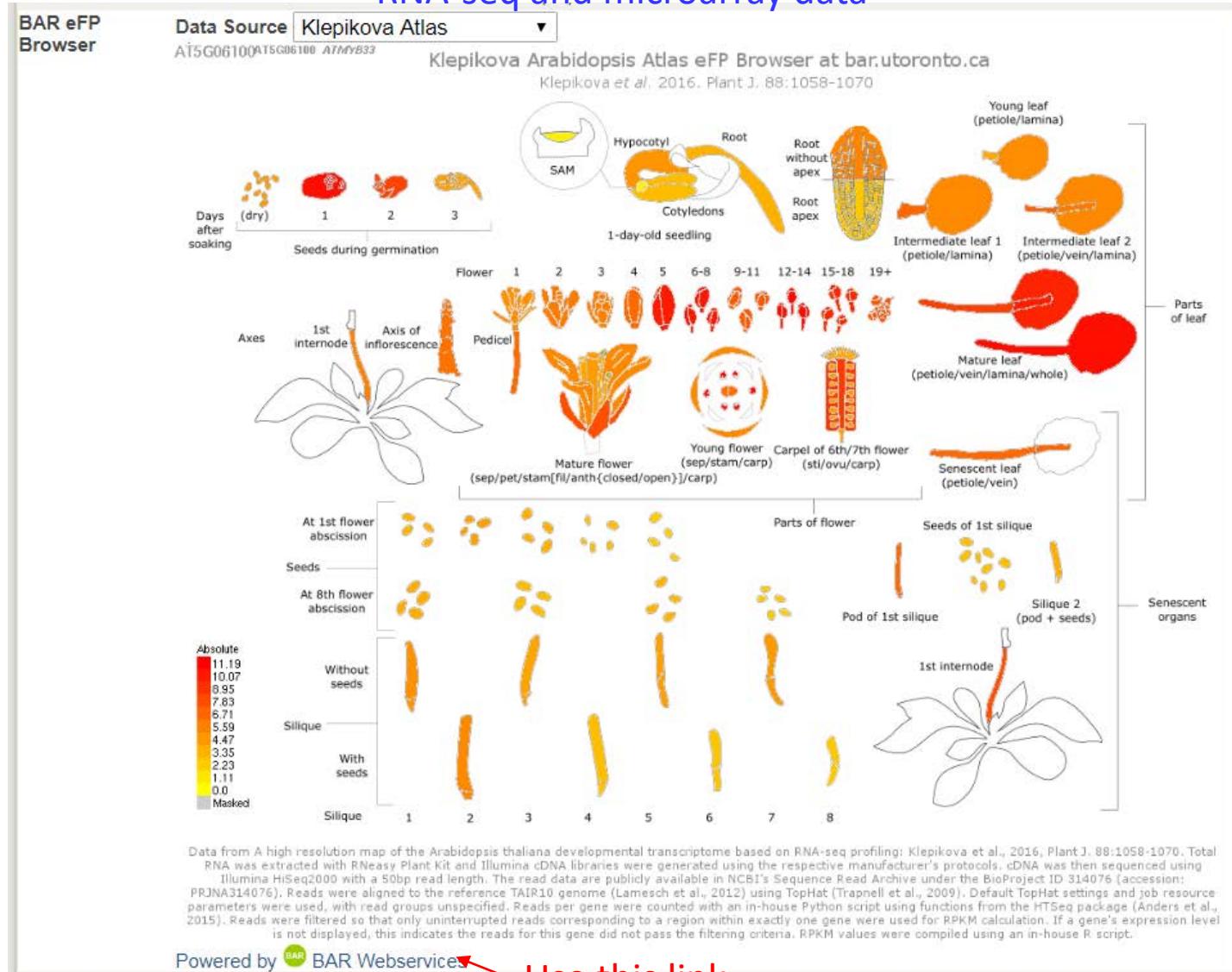
Microarray



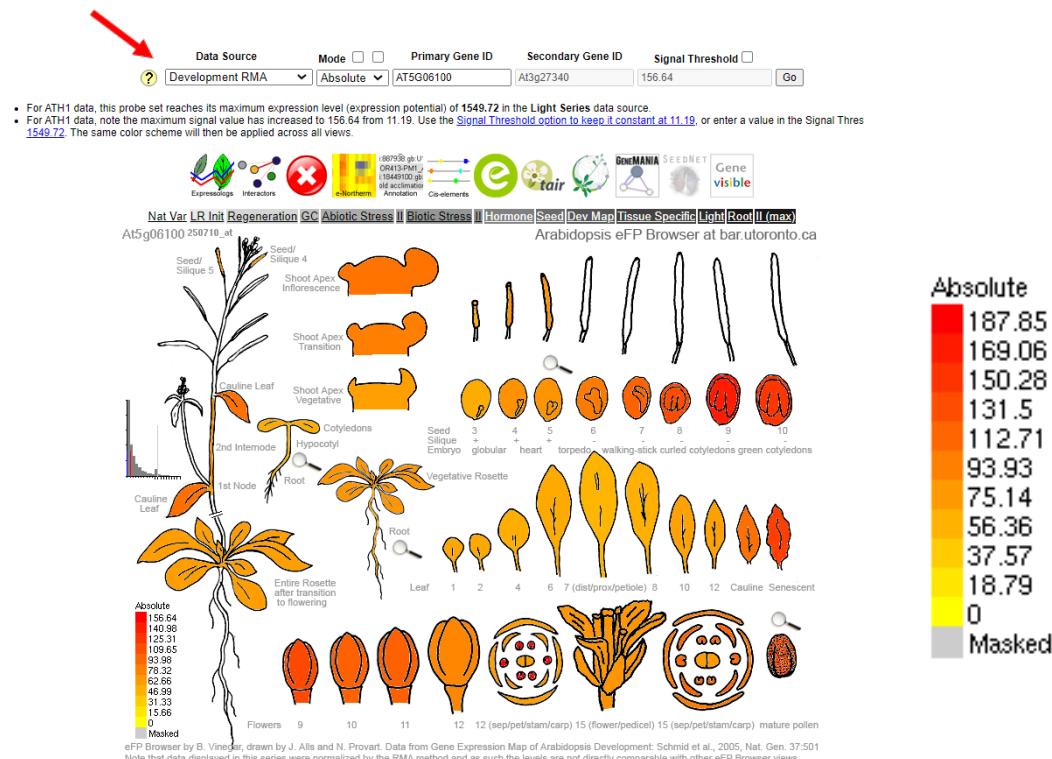
If no EST or cDNA, the gene is often computationally predicted, a pseudogene, of low abundance or expressed in limited tissues or conditions

Gene expression: BAR eFP Browser@TAIR (slow)

RNA-seq and microarray data



Gene expression: eFP browser



Gene expression: eFP browser

Abiotic stress: time course



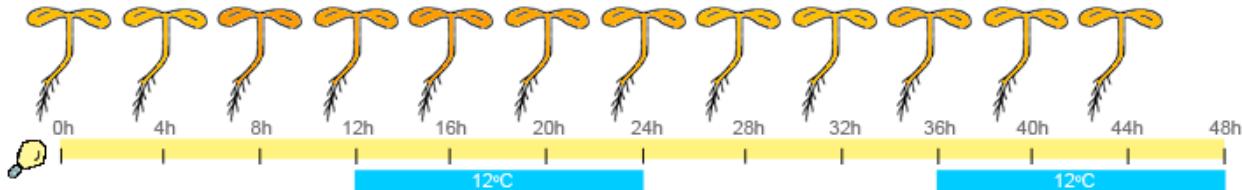
Gene expression: eFP browser

Light series: time course Diurnal series and circadian series

Diurnal series Yellow and black bars indicate lighting regimes at the stated light intensity. Blue bars indicate that temperature dips to 12°C from 22°C.

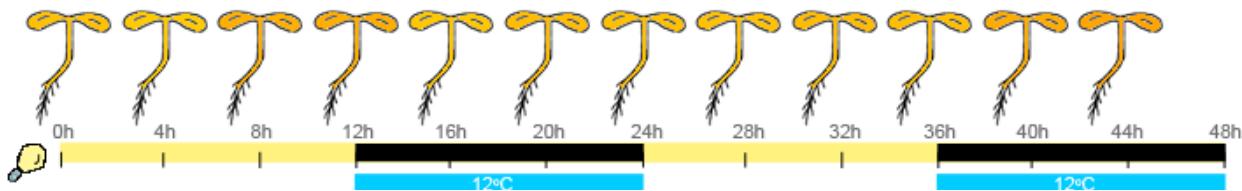
LLHC

Michael et al., 2008
7d old Col-0 seedlings
agar grown, no sucrose
continuous light, 100 μ E
22°C/12°C cycles



LDHC

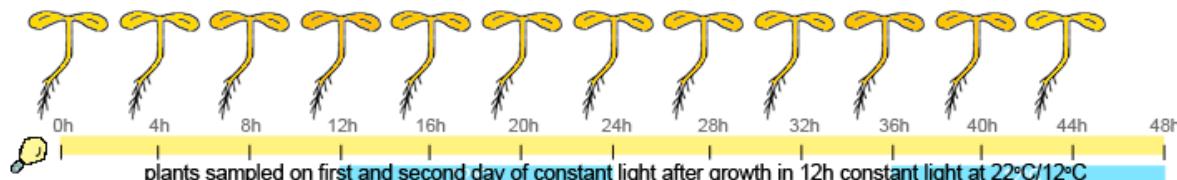
Michael et al., 2008
7d old Col-0 seedlings
agar grown, no sucrose
12h light, 100 μ E
22°C/12°C cycles



Circadian series Yellow and grey bars indicate lighting regimes prior to constant light or dark. Blue bars indicate the entrainment temperature dipped to 12

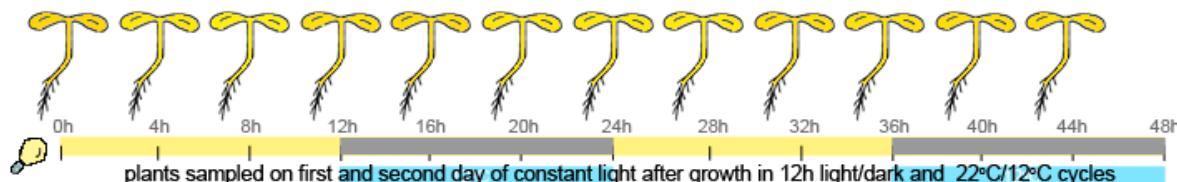
LL_LLHC

Michael et al., 2008
9d old Col-0 seedlings
agar grown, no sucrose
Constant light, 100 μ E
22°C



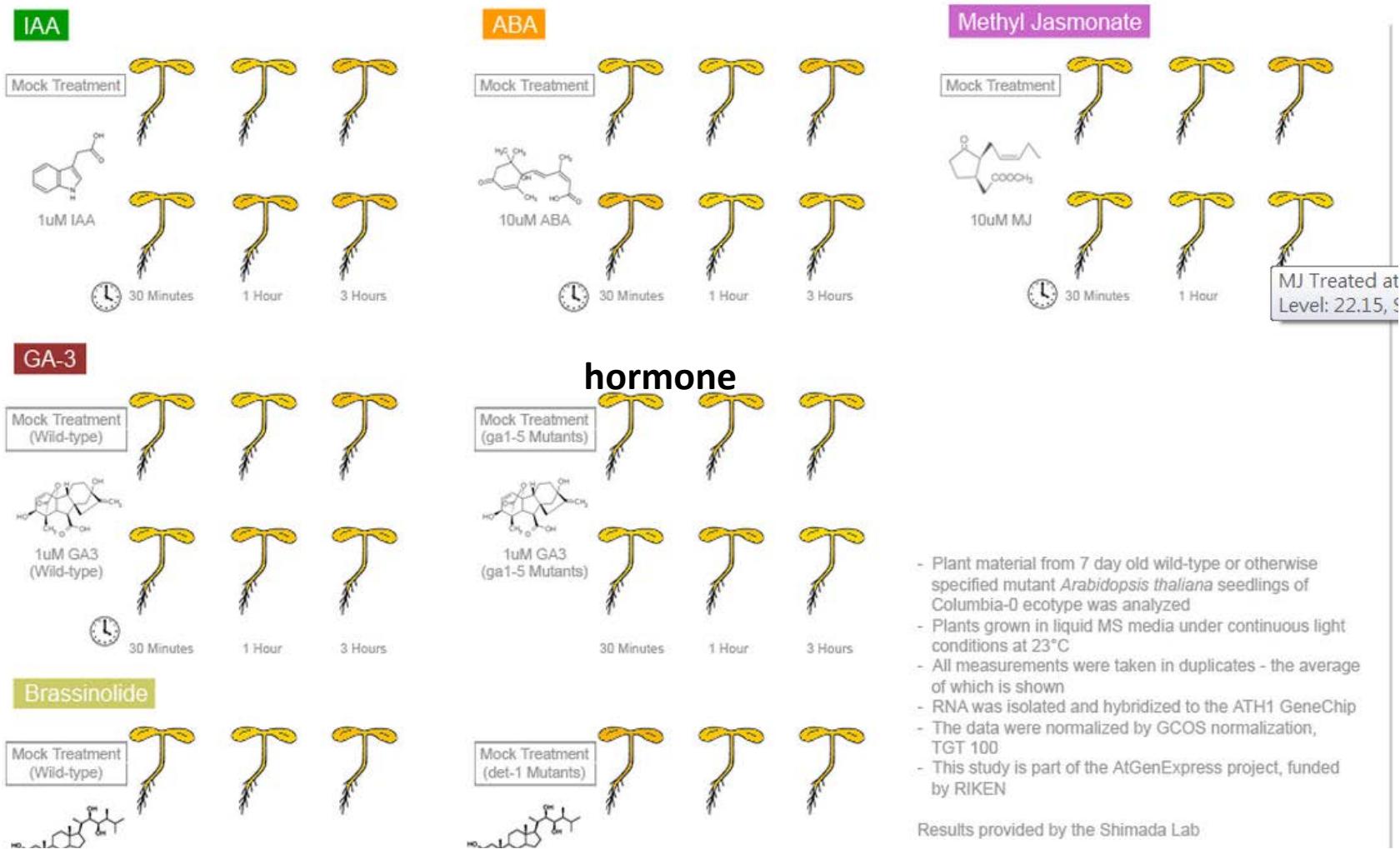
LL_LDHC

Michael et al., 2008
9d old Col-0 seedlings
agar grown, no sucrose
Constant light, 100 μ E
22°C



Gene expression :eFP Browser

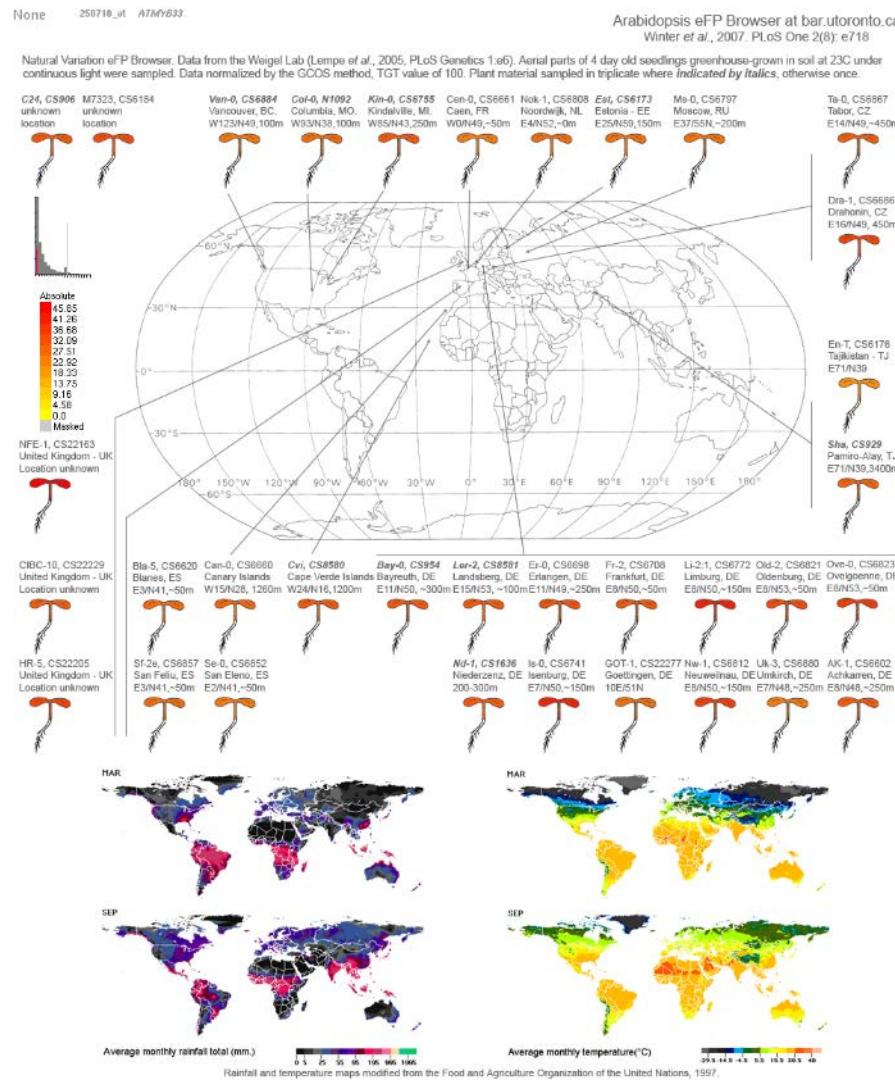
Hormone treatment



- Plant material from 7 day old wild-type or otherwise specified mutant *Arabidopsis thaliana* seedlings of Columbia-0 ecotype was analyzed
- Plants grown in liquid MS media under continuous light conditions at 23°C
- All measurements were taken in duplicates - the average of which is shown
- RNA was isolated and hybridized to the ATH1 GeneChip
- The data were normalized by GCOS normalization, TGT 100
- This study is part of the AtGenExpress project, funded by RIKEN

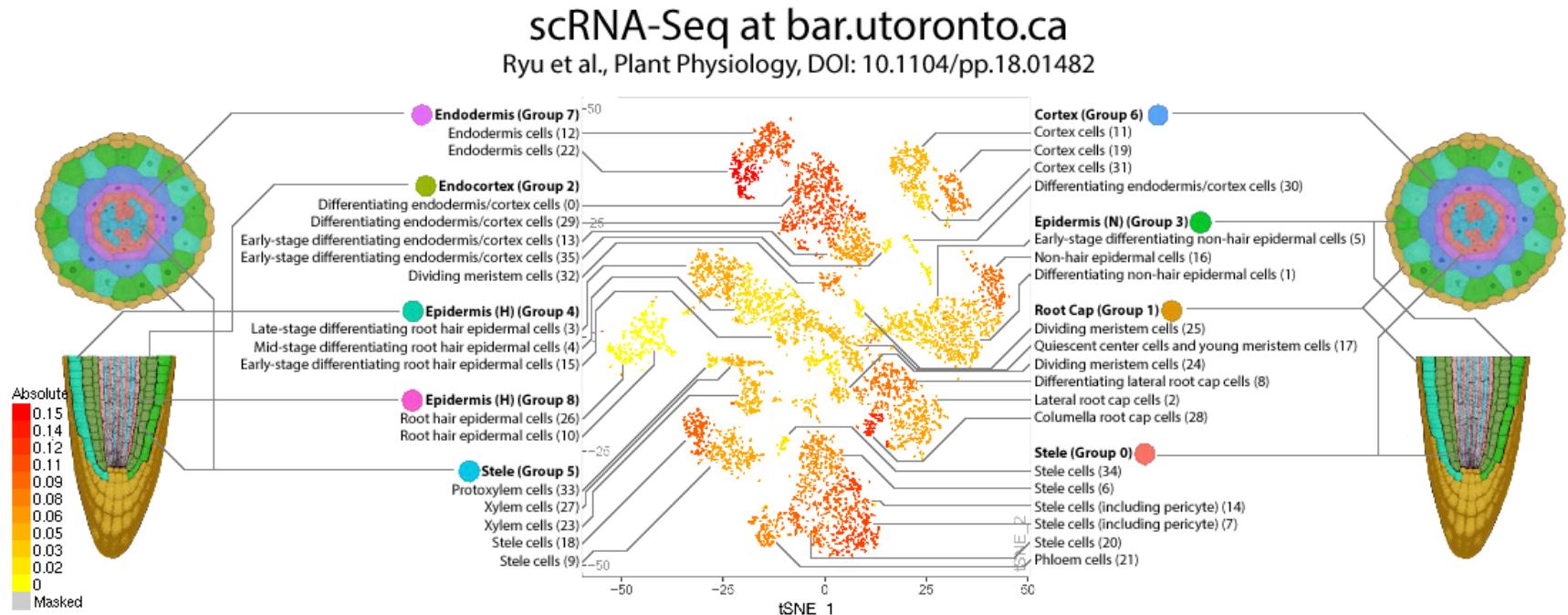
Gene expression: eFP Browser

ecotype



Gene expression: eFP Browser

Single cell





The Bio-Analytic Resource for Plant Biology

<http://bar.utoronto.ca/#GeneExpressionAndProteinTools>



Gene Expression and Protein Tools

View expression patterns as electronic fluorescent pictographs or heatmaps, explore promoters, identify protein-protein interactions and more.

Arabidopsis, maize, tomato, soybean, Medicago, Rice, Poplar, etc

Expression Angler:

Finding genes with similar expression pattern

Expression Browser:

Heatmap of gene expression



Molecular Markers and Mapping Tools

Perform Next Generation Mapping, or generate your own markers using our molecular marker tools.



Other Genomic Tools and Widgets

Remove duplicates, perform multi-dimensional Venn analyses, or generate random lists of identifiers.

DataMetaFormatter:

A clickable heatmap is generated, as well as a table of value

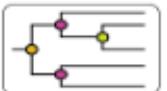
Venn Selector:

shows identifiers in common and unique to two sets of sequences.

Locus page: protein data

Protein Data ?				
name	length(aa)	molecular weight	isoelectric point	domains(# of domains)
AT5G06100.2	520	58394.7	5.6003	Homeodomain-related:IPR012287(2) Myb transcription factor:IPR015495(1) SANT domain, DNA binding:IPR001005(2) Homeodomain-like:IPR009057(2) Transcription regulator HTH, Myb-type, DNA-binding:IPR017930(2) Myb, DNA-binding:IPR014778(2) Transcription factor, GAMYB:IPR016310(1)

Gene families, phylogenetic tree, homologs

Search Gene Families	EnsemblPlants	PhyloGenes		
	InParanoid Ortholog Groups	PhylomeDB		
	PANTHER	Phytozome		
	PGDD duplications and orthologs	PLAZA		
Phylogenetic Tree	 PHYL GENES			
PANTHER 11 Plant Homologs Showing 3 of 3 entries	Organism Brachypodium distachyon (Purple false brome) Oryza sativa (Rice) Sorghum bicolor (Sorghum)	UniProt Accession I1I147 Q6ZBS8 C5YHW9	Gene BRADI3G15590 P0583B06.15 SB07G004900	Functional Information Search in QuickGO Search in QuickGO Search in QuickGO

Locus page: polymorphism and germplasm

Genetic or
epigenetic
variants

Mutants
T-DNA
EMS

Polymorphism	name	type	polymorphism site	allele type
Showing 15 of 23 entries (see all)	ATMYB33_3059	substitution	coding_region	unknown
	ATMYB33_3059	substitution	exon	unknown
	FLAG_206C04	insertion	exon	unknown
	FLAG_206G08	insertion	exon	unknown
	GK-589G04-021381	insertion	promoter	unknown
	GK-589G04-021428	insertion	promoter	unknown
	GT6511.Ds5.05.11.00.b.147	insertion	exon	unknown
	PERL0869716	substitution	promoter	unknown
	PERL0869717	substitution	intron	unknown
	PERL0869722	substitution	exon	unknown
	PERL0869723	substitution	exon	unknown
	PERL0869724	substitution	exon	unknown
	SAIL_434_C08.v1	insertion	exon	unknown
	SAIL_446_B03.v1	insertion	exon	unknown
	SALK_015891	insertion	exon	unknown
Germplasm	name	polymorphisms		
Showing 8 of 8 entries	SALK_058312	SALK_058312.35.10.x; SALKseq_058312.0; SALKseq_058312.1; SALKseq_058312.101; SALKseq_058312.2; SALKseq_058312.201; SALKseq_058312.3		
	images			
	None available			
	phenotypes			
	None available			
	SALK_065473	SALK_065473; SALKseq_065473.0; SALKseq_065473.1; SALKseq_065473.2; SALKseq_065473.3		
	images			
	None available			
	phenotypes			
	None available			

Mutants (of CIPK6)

◆ T-DNA, transposon, EMS mutants

The Arabidopsis Biological Resource Center (ABRC, not the one in Sinica)					
Germplasm: SALK_080951		Stock Information			
Stock Information	Order from ABRC	Availability	available	Donor	Joseph Ecker
Other Names	N580951	NASC stock number			
Taxon	Arabidopsis thaliana				
Pedigree	Parent Line	CS60000	Background	Ler, Ws (wild type for comparison)	
Chromosomal Constitution	Ploidy	2	Is Aneuploid	no	
Associated Polymorphisms	Name	Locus	Gene Name(s) and/or Equivalents	Genotype	Allele Mutagen
	SALKSEQ_080951.0	AT4G30960	AT4G30960.1/CBL-INTERACTING PROTEIN KINASE 6/SNF1-RELATED PROTEIN KINASE 3.14/SOS3-INTERACTING PROTEIN 3		T-DNA insertion
	SALKSEQ_080951.1				T-DNA insertion
	SALKSEQ_080951.2	AT5G62170	AT5G62170.3 ; AT5G62170.1/TON1 RECRUITING MOTIF 25		T-DNA insertion
	SALK_080951.48.55.X	AT4G30960	AT4G30960.1/CBL-INTERACTING PROTEIN KINASE 6/SNF1-RELATED PROTEIN KINASE 3.14/SOS3-INTERACTING PROTEIN 3		T-DNA insertion
Mutagen Applied to Strain	T-DNA insertion				
Associated Constructs	Clone Name	Construct Type	Vector Name	heterozygous	
	pROK2	simple_insert	pROK2		
Phenotype	no phenotype information available at this time				

Mutant order through ABRC (not our ABRC)

Need to apply an account for making order

The screenshot shows a product page for a mutant stock at the Arabidopsis Biological Resource Center (ABRC). The page includes the OSU.EDU header, the ABRC logo, and the Ohio State University navigation bar.

Name / Stock Number: SALK_080951

NASC stock number: N580951

Resource Type: seed

Availability: available

Donor:

- Joseph Ecker

Donation Date: 02/14/2003

Date Released: 02/21/2003

Description:

Sequence-indexed T-DNA insertion line, segregating for annotated insertion; kanamycin was employed for selection of T1 plants; distribution seeds are T2 or T3 generation.

Growth Requirement: none

Marker:

Background: Col (Columbia)

ABRC Comment: NOTE: kanamycin resistance gene may be silenced; PCR- or hybridization-based segregation analysis is required to confirm presence of insertion; may be segregating for phenotypes that are not linked to the insertion; may have additional insertions potentially segregating.; Please cite the Alonso et al. reference linked to this stock and acknowledge ABRC for distributing the seeds in all presentations and publications utilizing this material.

Format Shipped: 100 seeds per vial

Base / Commercial Price: \$10 / \$80

See who has ordered this mutant

A red arrow points from the text "See who has ordered this mutant" to the "View Order History" button.

Stock order histories

Friends or Competitors ?

Search Orders			
ORDER DATE	ORDER NUMBER	ORDERED BY	PI
2013-11-05	T104216	Yi-Fang Tsay	Yi-Fang Tsay
2013-09-05	T102969	Jenifer Bush	Jen Sheen
2012-09-12	T95069	Renjie Tang	Sheng Luan
2012-08-08	T94269	Yu Di-Qiu	Yu Di-Qiu
2011-09-13	T86957	Jie Shen	Wang Baichen
2011-03-22	T83298	Girdhar Pandey	Girdhar Pandey
2010-09-15	T79293	Aihua Sha	Xinan Zhou
2010-09-03	T79016	Hitoshi Onouchi	Satoshi Naito
2010-08-03	T78329	Karine David	Karine David
2010-01-07	T73798	Deng Fenglin	Xianlong Zhang
2009-12-09	T73300	Rebecca Stevenson	Jian-Kang Zhu
2009-05-11	T68713	Xue-Bao Li	Xue-Bao Li
2009-01-06	T65703	Paul Verslues	Paul Verslues

Mutants : which one to order

Mutation in coding region, 5' superior to 3'

Germplasm	name	polymorphisms	background	stock name	select
 Showing 8 of 8 entries	SALK_015891	SALK_015891		SALK_015891	<input type="checkbox"/>
	images				
	None available				
	phenotypes 				
	None available				
	WiscDsLox321D12	WiscDsLox321D12		CS851168	<input type="checkbox"/>
	images				
	None available				
	phenotypes 				
	None available				
	SALK_042186	SALK_042186.19.75.x		SALK_042186	<input type="checkbox"/>
	images				
	None available				
	phenotypes 				
	None available				
	SALK_053624	SALK_053624.195.05.x		SALK_053624	<input type="checkbox"/>
	images				
	None available				
	phenotypes 				
	None available				

T-DNA insertion sites: T-DNA Express @Salk

Locus Detail>External Link> T-DNA Express

External Link

- [Epigenomics](#)
- [Danforth Center Small RNA](#)
- [Expression/Localization](#)
- [AraQTL](#)
- [AtGenExpress Visualizations](#)
- [ATTED-II](#)
- [eFP Browser](#)
- [Eukaryotic Promoter Database](#)
- [Genevisible Expression](#)
- [NASCArrays Digital Northern](#)
- [NASCArrays Spot Hybridization](#)
- [The Subcellular Localization](#)
- [TraVA: Transcription Variation](#)
- [Gene Families](#)
- [Links](#)
- [Genomics](#)
- [AceView](#)
- [Araport](#)
- [Gramene](#)
- [MIPS](#)
- [NCBI-Entrez Gene](#)
- [Interactions](#)
- [BioGRID](#)
- [Polymorphisms/Alleles](#)
- [Salk SNP Viewer](#)
- [T-DNA Express](#)

SIGnAL
Funded by the National Science Foundation
Salk Institute Genomic Analysis Laboratory
NSF

ATSG06100.3
ATSG06100.5
ATSG06100.4

1839276

leq_066034 295757 SALKseq_056201 SALKseq_065473 GABI_39F04

WiscDsLox321D12 PAt5g06100 AY519616 U51311 CSHL_GT6

CSHL_41

T-DNA primer design: iSect Toolbox @ Salk

<http://signal.salk.edu/tdnaprimers.2.html>

1. PrimerL :

Please paste your list: like

```
Salk_000002  
SAIL_155_D07  
GABI_756F01  
FLAG_270B05  
RATM15-1976-1_G  
WiscDsLox289_292P9  
SM_3_19088
```

Output:

```
SALK_080951.48.55.x PRODUCT_SIZE 1268 PAIR_ANY_COMPL 0.00 PAIR_3'_COMPL 0.00 DIFF_TM  
0.12 LP TTCACAGTTCAACGTTGTGG Len 21 TM 59.64 GC 42.86 SELF_ANY_COMPL 0.12 3'_COMPL  
0.00 RP GTAGCGTGATTCGAGTTGG Len 21 TM 59.76 GC 47.62 SELF_ANY_COMPL 0.12 3'_COMPL  
0.00 Insertion chr4 15068952 BP+RP_PRODUCT_SIZE 605-905
```

If you find that the given primers don't work well (all seeds you tested are wild-type), you may need to re-sequence the T-DNA insertion site and design the primers by yourself.

Exercise I

TGA1 (AT5G65210)

1. number of isoforms ? (the differences)
2. the representative form in Araport11? Do you agree (check JBrwose)? If not, your answer?
3. GO biological process: number of keywords?; number of loci in each keyword?
4. In annotation, what functions are supported by experimental results?
5. Length of cDNA and CDS?
6. RNA expression: highest in which tissue, at what stage, in which ecotype? circadian gene? response to any stress or hormone?
7. Mutant: how many SALK, GABI, SAIL lines? Which mutants will you order? Why? Design primers for genotyping the mutants you select.

Locus page: External Link

External Link



Epigenomics

[Danforth Center Small RNA/PARE/Methylation](#)

Expression/Localization

[AraQTL](#)

[AtGenExpress Visualization Tool](#)

[ATTED-II](#)

[eFP Browser](#)

[Eukaryotic Promoter Database \(EPD\)](#)

[Genevisible Expression Data](#)

[The Subcellular Location of Proteins in Arabidopsis Database \(SUBA\)](#)

[TraVA:Transcription Variation Analysis](#)

Gene Families [Links moved to Gene Families band](#) [show me where]

Genomics

[AceView](#)

[Araport](#)

[Gramene](#)

[MIPS](#)

[NCBI-Entrez Gene](#)

Interactions

[BioGRID](#)

Other

[BAR ePlant](#)

Polymorphisms/Alleles/Variants

[Salk SNP Viewer](#)

[T-DNA Express](#)

Proteomics

[Functional Analysis Tools for Post-Translational Modifications\(FAT-PTM\)](#)

[Plant Proteome Database](#)

Publications

[EVEX](#)

Reagents

[Agrisera \(antibody\)](#)

[PhytoAB \(antibody\)](#)

[RIKEN BioResource Research Center \(Seed/DNA\)](#)

DNA methylation: MetC-seq @danforth

Locus page> External Link> Danforth Center Small RNA/PARE/Methylation



DNA methylation: MetC-seq @danforth

Epigenome (Methylome): Salk database

<http://signal.salk.edu/>

SIGNAL
Salk Institute Genomic Analysis Laboratory

Howard Hughes Medical Institute

About Us PBIO-E Data Release Policy cDNA Project T-DNA Express iSect Tools RiceGE Transcriptome CREB Data ATGC Data SIGnAL-2 Site Map

Plant

TTTGTGATTAGATTTGATGTAGTTTGATGTT
CCAGATCTCTGATGCAGTCCTCGATGCCCTGCTTG
CAAGACCAACATGGTCATGGCTTGGTGAGATCA
GGGACACCCAGACAAGCTCTGTGACCAGATCTCT
TTTGAATAAGATTTGATAGTAAAGTTGTTGTGA
TGATTAGATTTGATGTAGTTTGATGTTGTT
AGATCTCTGATGCAGTCCTCGATGCCCTGCTTGAA
ACCAAGACCAACATGGTCATGGCTTGGTGAGAT
AACGAGGGACACCCAGACAAGCTCTGTGACCAGAT
TGTGTTGTTGAATAAGATTTGATAGTAAAGTTG
TAAGTTTGATAGATTTGATGTAGTTTG
GCTCTGTGACCAGATCTCTGATGCAGTCCTCGATG
GTGAGACATGCACCAAGACCAACATGGTCATGGTC
CTGAGTCTGAAACGAGGGACACCCAGACAAGCTC
TAGTTTGATGTTGTTGAATAAGATTTGAT
TATTAGATAAGTTTGATAGATTTGATGT
GACAAGCTCTGTGACCAGATCTCTGATGCAGTCCT
TTGTGAGACATGCACCAAGACCAACATGGTCATGC

- *My Favourite Site*
SIGnAL T-DNA Express
- *Arabidopsis Tiling Array Transcriptome*
- *EIN3 -Ethylene Chip-seq /RNA-seq browser*
- *Arabidopsis 1,001 Genomes, Epigenomes, transcriptomes & Physical Maps*
- *Perlegen Arabidopsis Resequencing*
- **Transgenerational Inheritance of Methylation Variants**
- *Plant Cistrome /Epicistrome Atlas*
- *B-ARR-mediated cytokinin transcriptional network*
- *Abscisic Acid TF Network*
- *Phytochrome-Interacting Factors trigger chromatin dynamics in plants* **NEW**
- *Rice Functional Genomics Database*
- **Salk Insertion Sequence Database**
- *Arabidopsis Gene ORFeome Collection*
- *High Resolution Arabidopsis Exosome*
- *Salk Arabidopsis 1,001 Genomes*
- *Single Feature Polymorphism Database*
- **Single-base Resolution Epigenome Maps**
- *Arabidopsis Root Cell-type Methylomes*
- *Epigenetic Control of Ethylene Signaling*
- *Arabidopsis thaliana TF interaction network (AtTFIN-1)*
- **Soybean Epigenome Browser**
- **Salk Homozygote T-DNA Collection**
- *Arabidopsis GMUCT Uncapped & Cleaved Transcripts*
- *Arabidopsis Cyclome*
- *Complex architecture and epigenomics of plant T-DNA insertions*
- **High Resolution Arabidopsis Methylome**
- *Arabidopsis Biotic Stress Epigenome*
- *Arabidopsis Seed Methylomes*
- *Arabidopsis Interactome (AI-1)*
- *Jasmonic Acid Signaling Networks*
- *Integration of Omic Networks in Maize Atlas*

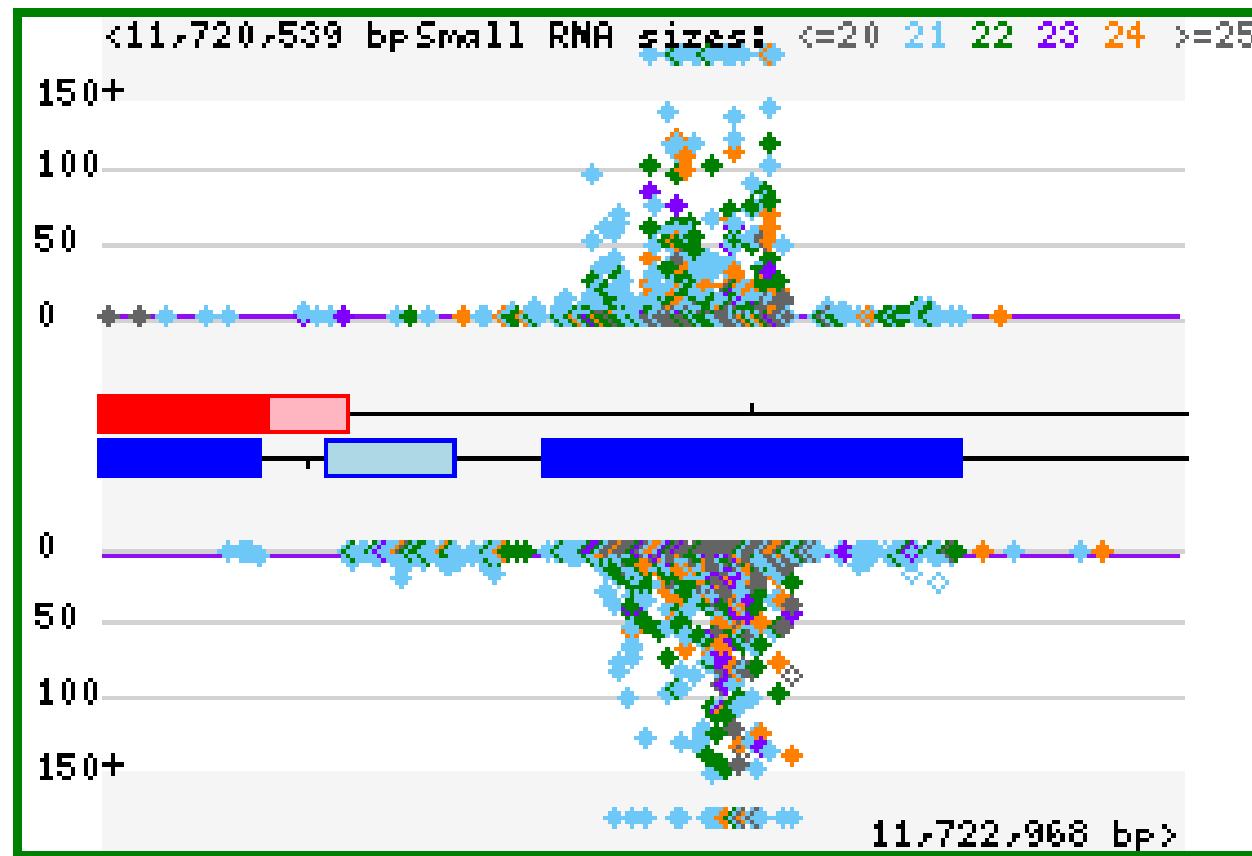
Small RNA: sRNA@Danforth

Locus page> External Link> Danforth Center Small RNA/PARE/Methylation



Small RNA: sRNA@Danforth

A tasiRNA locus

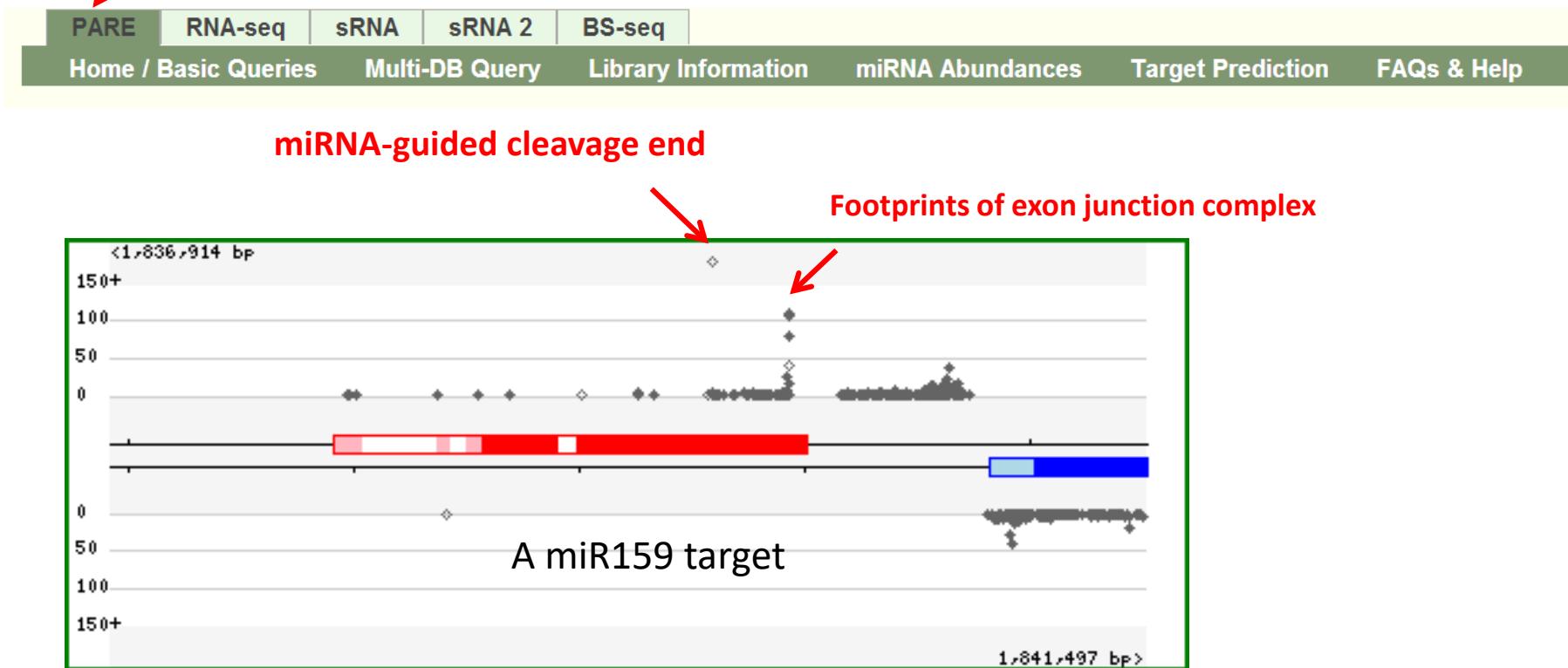


RNA degradation: PARE@Danforth

Locus page > External Link > Danforth Center Small RNA/PARE/Methylation

PARE: Parallel Analysis of RNA Ends

Uncapped RNA ends, cleavage ends of miRNA /siRNA targets, footprints of RNA binding proteins (ribosomes and exon junction complexes)



Polymorphism: Salk SNP Viewer

Locus page > External Link > Salk SNP Viewer

Not easy to have a global view

External Link 

- Epigenomics**
 - Danforth Center Smal
- Expression/Localization**
 - AraQTL
 - AtGenExpress Visualizer
 - ATTED-II
 - e-FP Browser
 - Eukaryotic Promoter Database
 - Genevisible Expression
 - NASCArrays Digital Northern
 - NASCArrays Spot Hybridization
 - The Subcellular Localization
 - TraVA:Transcription V
- Gene Families** [Links](#)
- Genomics**
 - AceView
 - Araport
 - Gramene
 - MIPS
 - NCBI-Entrez Gene
- Interactions**
 - BioGRID
- Polymorphisms/Alleles**
 - Salk SNP Viewer 
 - T-DNA Express

Arabidopsis SNP Sequence Viewer

Chromosome : chr5 Coordinates : 1838555 - 1839997

Display: Gene: AT5G06100.1 

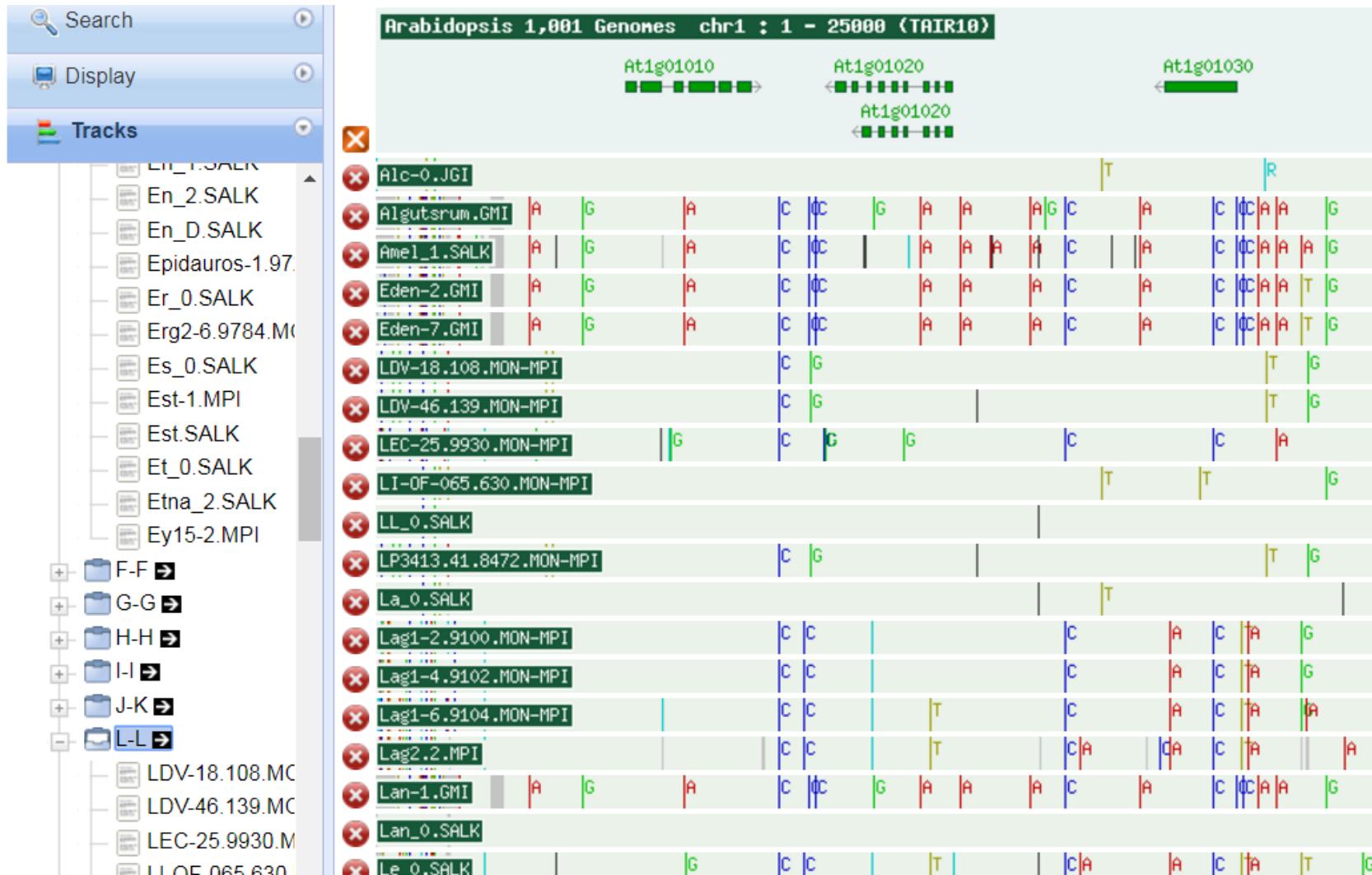
Chromosome: Coordinates: W/1838555-1838899,1838

GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT chr5.fa
GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT bay-0.fa
GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT bor-4.fa
GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT br-0.fa
GTCAACCAGCNGCTGATGATAATGGAAGTGACTGCAGNAGTAGAT bur-0.fa
GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT c24.fa
GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT col-0.fa
GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT cvi-0.fa
GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT est-1.fa
GTCAACCAGCTGCTGATGATAATGGAAGTGACTGCAGAAAGTAGAT fei-0.fa

Ecotype
DNA sequence

Polymorphism: Arabidopsis 1,001 genomes

<http://signal.salk.edu/atg1001/3.0/gebrowser.php>



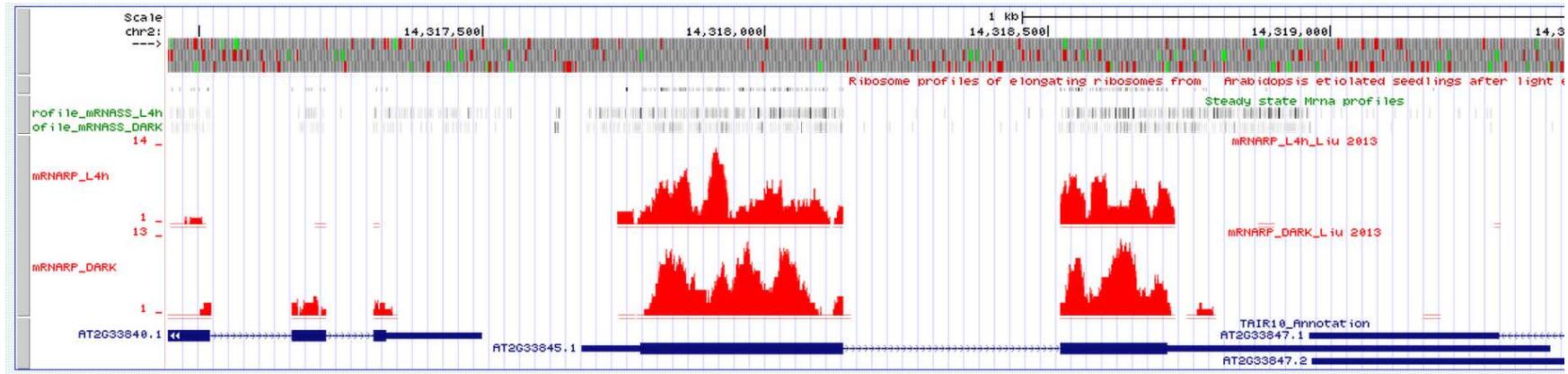
Protein Synthesis: GWIPS@UCC

**No external link in TAIR yet

Genome Wide Information on Protein Synthesis (NGS base)

<http://gwips.ucc.ie/cgi-bin/hgGateway>

- Profiling ribosome position by sequencing
- Coding or non-coding, uORF



Liu, M., Wu, S., Wu, J., Lin, W., Wu, Y., Tsai, T., Tsai, H., & Wu, S., (2013). Translation Landscape of Photomorphogenic Arabidopsis. *Plant Cell*.

Local alignment search tools @ TAIR

The screenshot shows the TAIR website interface. At the top, there is a navigation bar with links for Home, Help, Contact, About Us, Subscribe, and Log In. Below this is a secondary navigation bar with tabs for Search, Browse, Tools, Portals, and Data. The Tools tab is currently active, and a dropdown menu is displayed. The menu items listed are: GO Term Enrichment, JBrowse, GBrowse, Synteny Viewer, Seqviewer, Mapviewer, AraCyc Metabolic Pathways, Integrated Genome Browser, BLAST (with a red arrow pointing to it), Patmatch, Motif Analysis, VxInsight, Java Tree View, Bulk Data Retrieval, Chromosome Map Tool, Gene Symbol Registry, and Textpresso Full Text.

- search homologs
- map transcripts,
- examine the specificity of primers
-

BLAST program

Protein search is more sensitive than nucleotide search
Use BLASTP or BLASTX for searching homologs

TAIR BLAST 2.9.0+

This form uses NCBI BLAST 2.9.0+

Blast

BLAST™ program

Datasets:

Input:

query sequence
 locus name (At1g01030)

Upload a file

選擇檔案 未選擇任何檔案

Raw, FASTA, GCG and RSF formats accepted.

Filter query

Advanced BLAST™ Parameter Options

BLAST datasets

Home > Tools > BLAST

TAIR BLAST 2.9.0+

This form uses NCBI BLAST 2.9.0+

Blast

BLAST™ program: BLASTN: NT query, NT db

Datasets:

Input:

query sequence
 locus name (At1g01030)

Upload a file

Raw, FASTA, GCG and RSF formats accepted
 Filter query

Advanced BLAST™ Parameter Options

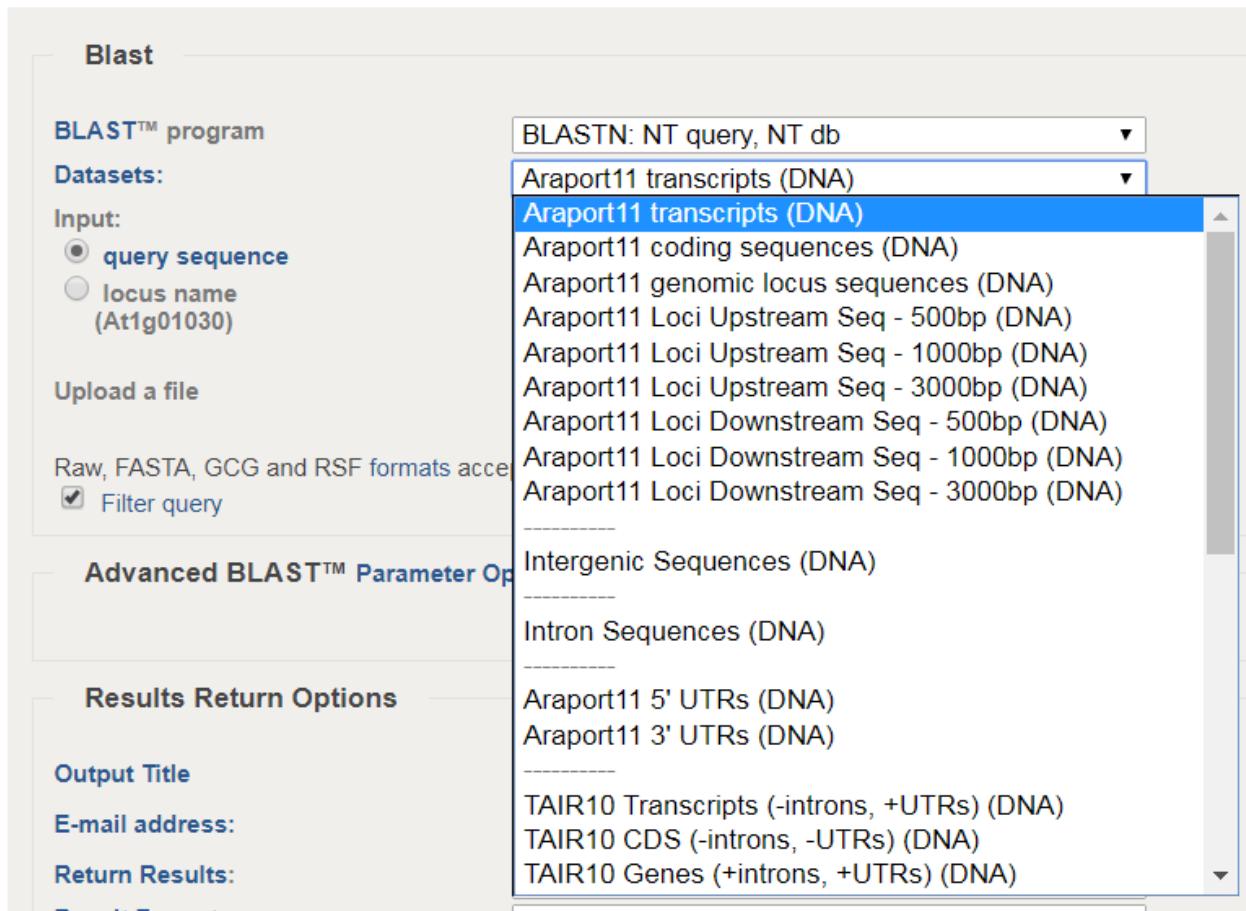
Results Return Options

Output Title

E-mail address:

Return Results:

TAIR10 Transcripts (-introns, +UTRs) (DNA)
TAIR10 CDS (-introns, -UTRs) (DNA)
TAIR10 Genes (+introns, +UTRs) (DNA)



Filter query: on or off

Blast

BLAST™ program: BLASTN: NT query, NT db

Datasets: Green plant GB experimental cDNA/EST (DNA)

Input:

query sequence
 locus name (At1g01030)

Upload a file: 選擇檔案 未選擇檔案

Raw, FASTA, GCG and RSF formats accepted.

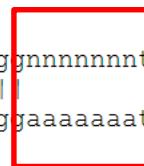
Filter query



- Filtering masks off segments of the query sequence that have low compositional complexity such as “poly A” or “CT repeat”
- Speed up the program and eliminate statistically significant but biologically uninteresting reports

Query: 61 caaaaatttacttctgaaatctctatttgaatttcgggnnnnnnnntcctaagtaaggga 120

Sbjct: 62 caaaaatttacttctgaaatctctatttgaatttcgggaaaaaaatcctaagtaaggga 121



Expectation value (E-value)

Increase E value for more hits (such as 100)

Decrease E value for highly similar hits (such as 0.0001)

Advanced BLAST™ Parameter Options

Weight Matrix: blosum62

Max Scores: 100

Max Alignments: 50

Gapped Alignments: ON

Expectation: 10

Nucleic Mismatch: -3

Nucleic Match: 2

Gap Opening Penalty: 0 (use default)

Gap Extension Penalty: 0 (use default)

Extension Threshold: 0 (use default)

Word Size: 0 (use default)

Query and Database Genetic Code: Standard or Universal (default)

A red arrow points from the text "Increase to 100 if query sequence is short (<20 nt) and expect imperfect hits" to the "Expectation" input field. Another red arrow points from the text "11->7 (for short query)" to the "Word Size" input field.

BLASTN

BLASTN: nucleotides query nucleotide database

Query= user-submitted sequence
(1662 letters)

Database: TAIR10 Transcripts (-introns, +UTRs) (DNA)
41,671 sequences; 64,867,051 total letters

Searching.....done

Sequences producing significant alignments:		Score	E
		(bits)	Value
AT3G11440.1	Symbols: ATMYB65, MYB65 myb domain protein ...	<u>3295</u>	0.0
AT5G06100.3	Symbols: MYB33, ATMYB33 myb domain protein ...	<u>359</u>	7e-98
AT5G06100.2	Symbols: MYB33, ATMYB33 myb domain protein ...	<u>359</u>	7e-98
AT5G06100.1	Symbols: MYB33, ATMYB33 myb domain protein ...	<u>359</u>	7e-98
AT4G09460.1	Symbols: AtMYB6, MYB6 myb domain protein 6 ...	<u>64</u>	6e-09
AT1G22640.1	Symbols: ATMYB3, MYB3 myb domain protein 3 ...	<u>64</u>	6e-09
AT2G26950.1	Symbols: AtMYB104, MYB104 myb domain protei...	<u>62</u>	2e-08
AT1G48000.1	Symbols: MYB112, AtMYB112 myb domain protei...	<u>60</u>	9e-08
AT4G38620.1	Symbols: ATMYB4, MYB4 myb domain protein 4 ...	<u>52</u>	2e-05
AT2G47190.1	Symbols: ATMYB2, MYB2 myb domain protein 2 ...	<u>52</u>	2e-05
AT4G21440.1	Symbols: ATMYB102, ATM4, MYB102 MYB-like 10...	<u>48</u>	4e-04
AT2G32460.2	Symbols: MYB101, ATMYB101, ATM1 myb domain ...	<u>48</u>	4e-04
AT2G32460.1	Symbols: MYB101, ATMYB101, ATM1 myb domain ...	<u>48</u>	4e-04
AT2G26960.1	Symbols: AtMYB81, MYB81 myb domain protein ...	<u>48</u>	4e-04
AT1G35515.1	Symbols: HOS10, MYB8 high response to osmot...	<u>48</u>	4e-04
AT5G59780.2	Symbols: MYB59, ATMYB59-2 myb domain protei...	<u>44</u>	0.006
AT5G59780.3	Symbols: MYB59, ATMYB59-3 myb domain protei...	<u>44</u>	0.006
AT5G62470.2	Symbols: MYB96 myb domain protein 96 chr5...	<u>44</u>	0.006
AT5G59780.1	Symbols: MYB59, ATMYB59, ATMYB59-1 myb doma...	<u>44</u>	0.006
AT5G62470.1	Symbols: MYB96, MYBCOV1, ATMYB96 myb domain...	<u>44</u>	0.006
AT3G47600.1	Symbols: MYB94, ATMYBCP70, ATMYB94 myb doma...	<u>44</u>	0.006

15 hits

← 0.001

BLASTX

BLASTX: nucleotides query amino acid database

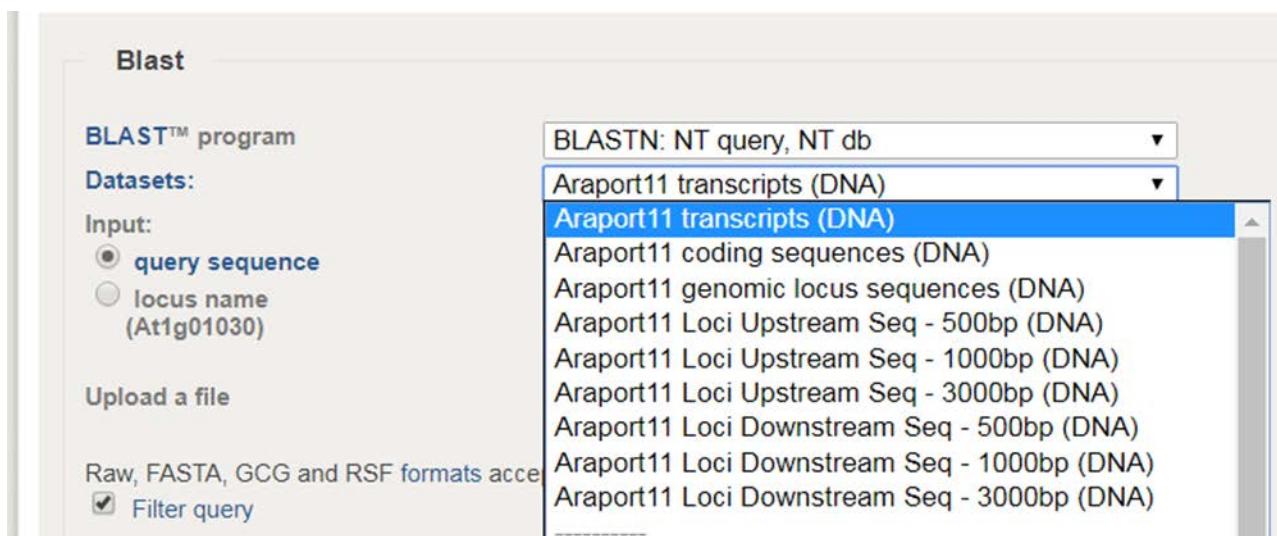
> 50 hits

← 5e-35

Drawbacks of sequence search @TAIR

For mapping sequence to the genome

- Need multiple searches if don't know where the query sequence is produced
- Difficult for the comparison of gene structure



Graph View of Alignment to Genome @ Salk



<http://signal.salk.edu/cgi-bin/atta>

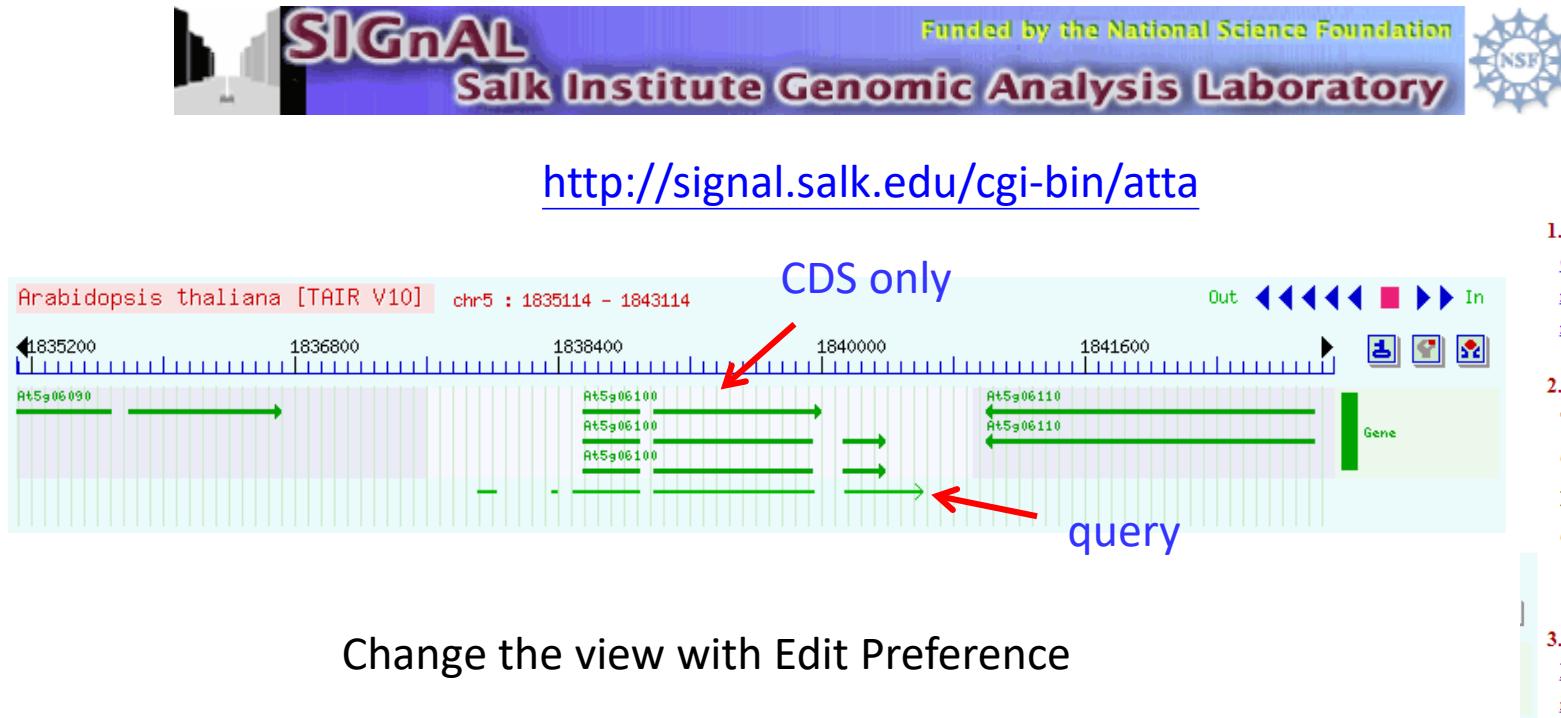
Increase E-value to 10 if query sequence is short



Summary:

Query	Chromosome	Hyperlink	Coordinates
	chr5	tdnaexpress transcriptome iPCR	W/1837914-1838033,1838371-1838396,1838498-1838900,1838987-1839954,1840152-1840621
	chr5	tdnaexpress transcriptome iPCR	W/1840152-1840727
	chr3	tdnaexpress transcriptome iPCR	W/3603179-3603429,3603550-3603638,3603703-3603765,3604122-3604276,3604711-3604757
	chr4	tdnaexpress transcriptome iPCR	W/5993623-5993669

Graph View of Alignment to Genome @ Salk



Change the view with Edit Preference

Bulk data retrieval

The screenshot shows the homepage of The Arabidopsis Information Resource (TAIR). At the top, there is a navigation bar with links for Home, Help, Contact, About Us, Subscribe, and Log In. Below this is a secondary navigation bar with tabs for Search, Browse, Tools, Portals, and Data. The Tools tab is currently active, and a dropdown menu is displayed. The menu lists various tools: GO Term Enrichment, JBrowse, GBrowse, Synteny Viewer, Seqviewer, Mapviewer, AraCyc Metabolic Pathways, Integrated Genome Browser, BLAST, Patmatch, Motif Analysis, VxInsight, Java Tree View, Bulk Data Retrieval (with a red arrow pointing to it), Chromosome Map Tool, Gene Symbol Registry, and Textpresso Full Text. The background features a photograph of an Arabidopsis plant.

- GO Term Enrichment
- JBrowse
- GBrowse
- Synteny Viewer
- Seqviewer
- Mapviewer
- AraCyc Metabolic Pathways
- Integrated Genome Browser
- BLAST
- Patmatch
- Motif Analysis
- VxInsight
- Java Tree View
- Bulk Data Retrieval** ←
- Chromosome Map Tool
- Gene Symbol Registry
- Textpresso Full Text

Bulk data retrieval

Home > Tools > Bulk Data Retrieval

A list of genes
Common properties

Multiple sequence alignment
or
cis-element analysis

Bulk Data Retrieval and Analysis

Gene Descriptions

Download gene descriptions using locus identifiers (AGI codes) or gene names.

Gene Ontology (GO) Annotations

Search and download GO annotations using locus identifiers. Functionally classify Arabidopsis genes and view the results as tables or pie charts.

Plant Ontology (PO) Annotations

Search and download PO annotations using locus identifiers.

Sequences

Download nucleotide and protein sequences using locus identifiers. Send FASTA formatted sequences to ClustalW multiple sequence alignment program at the European Bioinformatics Institute.

Proteins

Search proteins according to predicted protein properties. Now includes SCOP's structural class prediction.

Locus History

Find out how a locus has been modified during different releases of the genome by entering a locus name.

Microarray Elements

Search array elements on the Affymetrix 8K, 25K, and all the AFGC arrays.

Bulk sequence retrieval for sequences

Home > Tools > Bulk Data Retrieval > **Sequences**

Sequence Bulk Download and Analysis

This tool can be used to download a variety of sequences from the Arabidopsis Genome Initiative (AGI) in FASTA or tab-delimited formats. Individual or sets of AGI locus identifiers (e.g. At1g01030) can be typed into the textbox below or uploaded from your desktop. FASTA formatted sequences can also be sent to the multiple sequence alignment program Clustal W. More information on this tool can be found by following the link to the Help document.

Please use the ftp server if you want to download entire datasets. Click [HERE](#) to obtain details about the sequence datasets.

For Intron Sequences ONLY: A special format is required for obtaining intron sequences using the Intron Sequences data type. The identifier must include the locus identifier followed by the gene model suffix. To specify which intron, the model id.suffix is followed by the number of the intron you wish to retrieve. For example, to obtain the sequence of the first intron on the AGAMOU1 locus would be AT4G18960.1-1

For Upstream/Downstream/Intergenic sequences: For these data types we ONLY provide the sequence relative to the locus (based on the maximum extent of all gene models for that locus) regardless of whether a locus name or a specific gene model is used for the search.

Sequences	Araport11 transcripts
Locus/Gene Model Identifiers or Sequences:	Araport11 coding sequences Araport11 genomic locus sequences Araport11 protein sequences Araport11 Loci Upstream Seq - 500bp Araport11 Loci Upstream Seq - 1000bp Araport11 Loci Upstream Seq - 3000bp Araport11 Loci Downstream Seq - 500bp Araport11 Loci Downstream Seq - 1000bp Araport11 Loci Downstream Seq - 3000bp Intergenic Sequences Intron Sequences Araport11 3' UTRs Araport11 5' UTRs
Upload file:	
Dataset:	Araport11 transcripts ▾

Promoter analysis

Motif Analysis

in Promoter or Upstream Gene Sequences

Home > Tools > Motif Analysis

Statistical Motif Analysis in Promoter or Upstream Gene Sequences

The program compares the frequencies of 6-mer "words" in your query set of sequences (on both strands) with the frequencies of the words in the current genomic sequence set of 33518 sequences, each containing 500 (or 1000) bp upstream of the start codon of each gene. You can type in sets of AGI locus identifiers (e.g. At1g01030) or sets of fasta sequences. Make sure each fasta header is formatted as such, fasta symbol (>), immediately followed by a unique ID, a space, then all other descriptions (e.g. >ABCD1.1 my gene). Ensure that there are no sequences appearing more than once in your query set.

Upload file: 未選擇任何檔案

Dataset:

- 500 bp upstream 1000 bp upstream 3000 bp upstream

Output type:

- HTML Text

Input: AGI or fasta sequence

Output: 6-mer motif

ACCGGT	192	1918	90/1235	916/34187	4.49e-	AT1G59860 AT1G66090 AT3G02840 AT1G56060 AT5G64870
17						AT4G24570 AT4G25380 AT3G55980 AT1G51920 AT2G24600
						AT4G11480 AT4G38420 AT5G28630 AT2G22470 AT1G72520
						AT4G30430 AT2G32210 AT5G41740 AT5G18470
						AT5G06320 AT4G27657 AT4G34440 AT4G01950
						AT1G76600 AT5G25450 AT5G02220 AT4G34150
						AT3G62720 AT1G28600 AT5G10380 AT5G64900
						AT3G15370 AT3G46600 AT5G41100 AT3G05200
						AT1G24150 AT5G64310 AT1G17380 AT1G19770
						AT1G51620 AT3G25900 AT1G62440 AT1G08860
						AT2G24130 AT5G37770 AT1G70130 AT4G30210
						AT5G13200 AT5G59490 AT1G66400 AT5G22680
						AT4G26690 AT1G76080 AT1G23710 AT1G67060
						AT2G20960 AT1G02660 AT3G06500 AT5G41550
						AT5G58120 AT1G43000 AT2G24860 AT1G51660
						AT5G06860 AT1G59910 AT3G05320 AT2G36770
						AT1G19180 AT4G34210 AT1G75860 AT4G30090
						AT5G42050 AT1G75390 AT1G30040 AT3G12740
						AT4G25390 AT3G48090 AT3G57280 AT5G53990
						AT4G08500 AT5G04720 AT3G51920 AT5G36260
						AT3G22370 AT3G61640 AT4G37640 AT1G33560
						AT3G55270 AT5G01950 AT1G03740
AACCGG	148	1787	123/1235	1568/34187	1.41e-	AT1G59860 AT2G28210 AT1G66090 AT1G56060 AT5G46295
15						AT5G64870 AT4G24570 AT3G17690 AT4G25380 AT1G51920
						AT1G35140 AT3G28210 AT2G24600 AT1G05575 AT2G46400
						AT1G66160 AT4G38420 AT2G29500 AT5G28630
						AT3G08970 AT2G22470 AT1G72520 AT1G66500
						AT4G30280 AT4G30430 AT2G32210 AT5G41740
						AT5G18470 AT5G54720 AT1G14040 AT4G27657
						AT4G34440 AT5G25450 AT3G12580 AT1G53980

Motif Analysis

in Promoter or Upstream Gene Sequences

Sequence Bulk Download and Analysis

This tool can be used to download a variety of sequences from the Arabidopsis Genome Initiative (AGI) in FASTA or tab-delimited formats. Individual or sets of AGI locus identifiers (e.g. At1g01030) can be typed into the textbox below or uploaded from your local FASTA-formatted sequences can also be sent to the multiple sequence alignment program Clustal W. More information or tool can be found by following the link to the Help document.

Please use the ftp server if you want to download entire datasets. Click [HERE](#) to obtain details about the sequence datasets.

For Intron Sequences ONLY: A special format is required for obtaining intron sequences using the Intron Sequences dataset. It must include the locus identifier followed by the gene model suffix. To specify which intron, the model id.suffix is followed by the number of the intron you wish to retrieve. For example, to obtain the sequence of the first intron on the AGAMOU would be AT1G18960.1-1.

For Upstream/Downstream/Intergenic sequences: For these data types we ONLY provide the sequence relative to the locus (based on the maximum extent of all gene models for that locus) regardless of whether a locus name or a specific gene name is used for the search.

Sequences Locus/Gene Model Identifiers or Sequences: Upload file: Dataset:	Araport11 transcripts Araport11 coding sequences Araport11 genomic locus sequences Araport11 protein sequences Araport11 Loci Upstream Seq - 500bp Araport11 Loci Upstream Seq - 1000bp Araport11 Loci Upstream Seq - 3000bp Araport11 Loci Downstream Seq - 500bp Araport11 Loci Downstream Seq - 1000bp Araport11 Loci Downstream Seq - 3000bp Intergenic Sequences Intron Sequences Araport11 3' UTRs Araport11 5' UTRs
--	---



For further information on how to interpret these results please access <https://meme-suite.org/meme/doc/meme.html>.
To get a copy of the MEME software please access <https://meme-suite.org>.

If you use MEME in your research, please cite the following paper:
Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", *Proceedings of the National Academy of Sciences of the United States of America*, 92(13), 6295-6300, 1995.

[DISCOVERED MOTIFS](#) | [MOTIF LOCATIONS](#) | [INPUTS & SETTINGS](#) | [PROGRAM INFORMATION](#) | [RESULTS IN TEXT FORM](#)

DISCOVERED MOTIFS

	Logo	E-value	Sites	Width	More	Submit/Download
1.		2.5e-019	437	10	<input type="button" value="I"/>	<input type="button" value="..."/>
2.		1.6e-002	239	10	<input type="button" value="I"/>	<input type="button" value="..."/>
3.		6.0e+002	24	10	<input type="button" value="I"/>	<input type="button" value="..."/>

Stopped because requested number of motifs (3) found.

MEME Suite 5.5.1

- Motif Discovery
- Motif Enrichment
- Motif Scanning
- Motif Comparison
- Gene Regulation
- Utilities
- Manual
- Guides & Tutorials
- Sample Outputs
- File Format Reference
- Databases
- Download & Install
- Help
- Alternate Servers
- Authors & Citing
- Recent Jobs

MEME
Multiple Em for Motif Elicitation

Version 5.5.1

Data Submission Form

Perform motif discovery on DNA, RNA, protein or custom alphabet datasets.

Select the motif discovery mode

Classic mode Discriminative mode Differential Enrichment mode

Select the sequence alphabet

Use sequences with a standard alphabet or specify a custom alphabet.

DNA, RNA or Protein Custom 選擇檔案 未選擇任何檔案

Input the primary sequences

Enter sequences in which you want to find motifs.

Select the site distribution

How do you expect motif sites to be distributed in sequences?

Zero or One Occurrence Per Sequence (zoops)

Select the number of motifs

How many motifs should MEME find?

Advanced options

What should be used as the background model?

How wide can motifs be?

How many sites must each motif have? Minimum sites: Maximum sites:

Can motif sites be on both strands? (DNA/RNA only) search given strand only

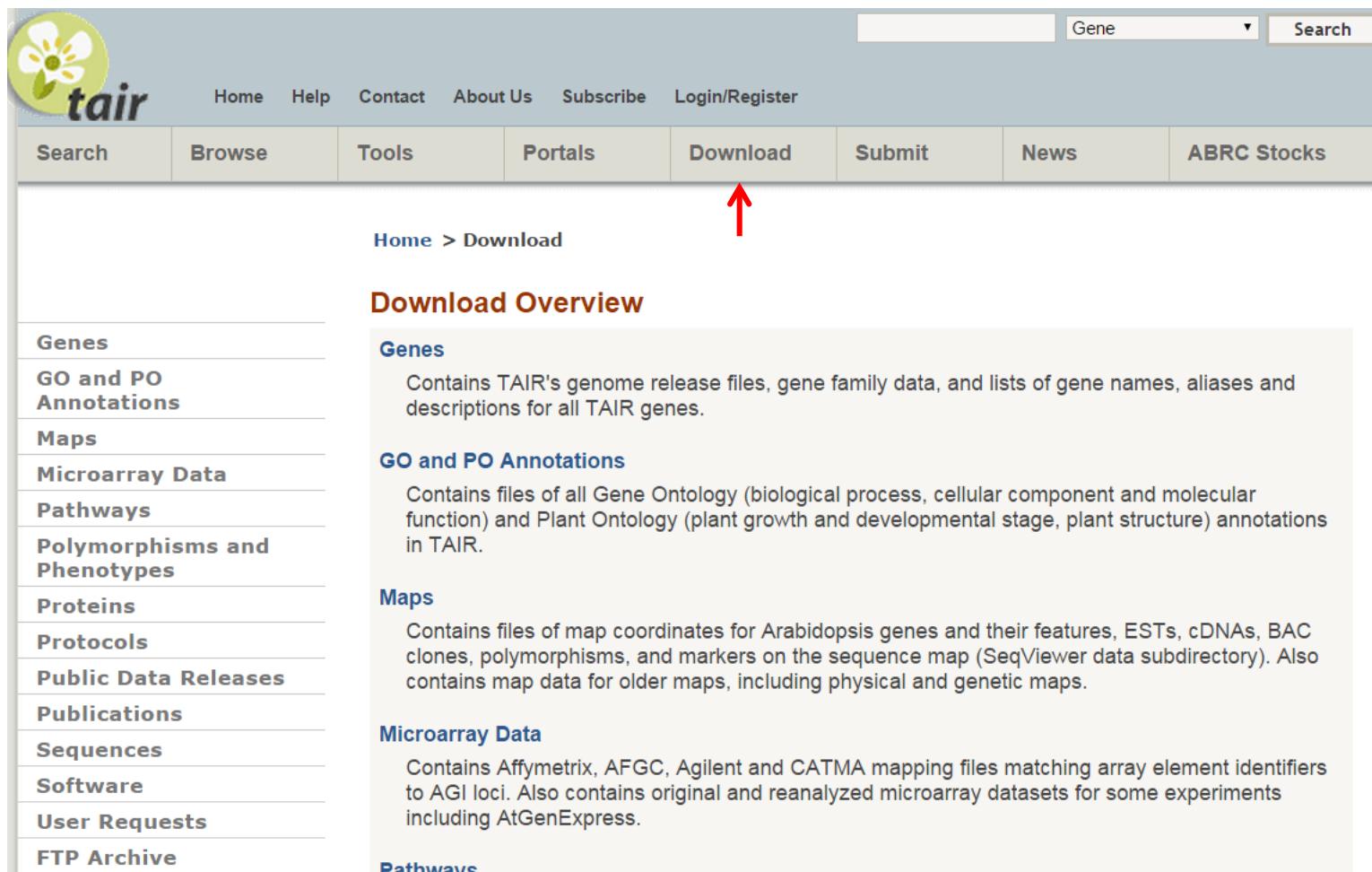
Should MEME restrict the search to palindromes? (DNA only) look for palindromes only

Should MEME shuffle the sequences? Shuffle the sequences

Note: if the combined form inputs exceed 80MB the job will be rejected.

TAIR FTP download

Home > Download > Download Overview



The screenshot shows the TAIR website interface. At the top, there is a search bar with a dropdown menu set to 'Gene' and a 'Search' button. Below the search bar is a navigation menu with links: Home, Help, Contact, About Us, Subscribe, Login/Register, Search, Browse, Tools, Portals, Download (which has a red arrow pointing to it), Submit, News, and ABRC Stocks.

Below the navigation menu, the URL 'Home > Download' is displayed. The main content area is titled 'Download Overview'. It contains several sections with descriptions:

- Genes**: Contains TAIR's genome release files, gene family data, and lists of gene names, aliases and descriptions for all TAIR genes.
- GO and PO Annotations**: Contains files of all Gene Ontology (biological process, cellular component and molecular function) and Plant Ontology (plant growth and developmental stage, plant structure) annotations in TAIR.
- Maps**: Contains files of map coordinates for Arabidopsis genes and their features, ESTs, cDNAs, BAC clones, polymorphisms, and markers on the sequence map (SeqViewer data subdirectory). Also contains map data for older maps, including physical and genetic maps.
- Microarray Data**: Contains Affymetrix, AFGC, Agilent and CATMA mapping files matching array element identifiers to AGI loci. Also contains original and reanalyzed microarray datasets for some experiments including AtGenExpress.
- Pathways**

TAIR FTP download: genes

Download - Genes

- [Araport11 genome release](#)
- [gene_aliases_20130831.txt 876](#)
- [Gene families](#)
- [**TAIR10 genome release**](#) 
- [TAIR6 genome release](#)
- [TAIR7 genome release](#)
- [TAIR8 genome release](#)
- [TAIR9 genome release](#)
- [TAIR gene confidence ranking](#)
- [TIGR release versions](#)

Download - TAIR10 genome release

- [gene_description_20131231.txt.gz 4,474 KB 21](#)
- [README_TAIR10.txt 5 KB 2012-08-22](#)
- [TAIR10-Subcellular_Predictions.xlsx 3,964 KB](#)
- [**TAIR10 blastsets** !\[\]\(37f8ca622a0422c2a4bf3cdf1d9784da_img.jpg\)](#)
- [**TAIR10 chromosome files** !\[\]\(b8e3fa35029d134b67387e0955b3e3a0_img.jpg\)](#)
- [TAIR10_domain_architectures.tab.t10 2,608 KB](#)
- [TAIR10_functional_descriptions 25,396 KB 201](#)
- [TAIR10_functional_descriptions.bk 25,396 KB](#)
- [TAIR10_functional_descriptions_20130831.txt 25](#)
- [TAIR10 gene confidence ranking](#)
- [TAIR10 gene lists](#)
- [TAIR10 gene transcript associations](#)
- [TAIR10 gff3](#)
- [TAIR10_locushistory.txt 2,053 KB 2010-11-23](#)
- [TAIR10 NCBI mapping files](#)
- [TAIR10_sequence_edits.txt 0 KB 2010-12-07](#)
- [TAIR10_TAIRAccessionID_AGI_mapping.txt 792](#)
- [TAIR10_TAIRlocusaccessionID_AGI_mapping.txt](#)
- [TAIR10 transposable elements](#)
- [Tair10 XML](#)

TAIR FTP download: sequences

Download - Sequences

-  Araport11 blastsets
-  TAIR10 blastsets

> Sequences

Download - TAIR10 blastsets

-  downstream sequences
 -  Readme_blastdatasets_TAIR10.txt 4 KB 2012-04-16
 -  TAIR10_3_utr_20101028 8,718 KB 2010-11-10
 -  TAIR10_5_utr_20101028 6,027 KB 2010-11-10
 -  TAIR10_bac_con_20101028 137,399 KB 2010-11-10
 -  TAIR10_cdna_20101214_updated 71,717 KB 2012-04-16
 -  TAIR10_cdna_20110103_representative_gene_model_updated 56,587 KB 2012-04-16
 -  TAIR10_cds_20101214_updated 49,453 KB 2012-04-16
 -  TAIR10_cds_20110103_representative_gene_model_updated 38,019 KB 2012-04-16
 -  TAIR10_exon_20101028 81,803 KB 2010-11-10
 -  TAIR10_intergenic_20101028 51,663 KB 2010-11-10
 -  TAIR10_intron_20101028 41,688 KB 2010-11-10
 -  TAIR10_pep_20101214_updated 20,006 KB 2012-04-16
 -  TAIR10_pep_20110103_representative_gene_model_updated 15,437 KB 2012-04-16
 -  TAIR10_seq_20101214_updated 101,193 KB 2012-05-07
 -  TAIR10_seq_20110103_representative_gene_model_updated 76,879 KB 2012-04-16
-  upstream sequences

Database/Tool Help

The screenshot shows the TAIR website's "Help" page. At the top, there is a navigation bar with links for Home, Help (which has a red arrow pointing to it), Contact, About Us, Subscribe, and Login/Register. Below the navigation bar is a horizontal menu with links for Search, Browse, Tools, Portals, Download, Submit, News, and ABRC Stocks. The main content area displays the "Help" page, which includes a "Home > Help" breadcrumb, a "Tutorials" section with links to AraCyc Tutorial, TAIR Gene Ontology Tools, and TAIR Microarray Tools, and a "Getting Started Guide". A sidebar on the left lists links for Database/Tool Help Pages, FAQ, TAIR Glossary, and TAIR User Guide. The main content area also contains instructions about using the Definitions help feature and a search box for help. A note at the bottom states that tutorials work best with specific browsers.

This section of TAIR's website contains links to documents to help you maximally utilize TAIRs database and tools.

Use the Definitions help. Wherever you see a symbol next to a word , click on the symbol to see a definition for the word or phrase.

If you have a specific topic you need help with, you can search ONLY within the help section of TAIR using the search box below.

Tutorials

Note: These tutorials work best with the following browsers: Firefox, Internet Explorer 6.X(Win),

The screenshot shows the TAIR Bulk Data Retrieval GO Annotations page. The URL in the address bar is "Home > Tools > Bulk Data Retrieval > GO Annotations". The main content area features a heading "GO annotation search, functional categorization and download Gene Ontology at TAIR" with a red arrow pointing to the "Help" link. Below the heading, there is a paragraph of text explaining how to use the tool, mentioning the submission of locus identifiers and the use of functional categorization.

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

[Help]

Exercise II

TGA1 (AT5G65210)

1. Does TGA1 produce a lot of small RNAs? Any specific size of small RNA?
2. Any major RNA degradation products (PARE) in TGA1 ? location?
3. Any SNP in the CDS of TGA1 in C24 ecotype?
4. Based on the Ribo-seq data in GWIPS@UCC, does TGA1 have a translated ORF in 5'UTR?
5. Besides itself, which gene is the best hit of TGA1 cDNA? How many hits with E-value <e-20?
6. Besides itself, which gene is the best hit of TGA1 protein? How many hits with E-value <e-20?
7. For the sample list in GO term enrichment, which enriched term of molecular function and cellular component has the best FDR? their fold enrichment?
8. How big is the file size of Araport11_cds.20220914.gz (download/sequences/)?



Thank you !

For questions, you can contact

Ho-Ming Chen 陳荷明

Office: Agricultural Technology Building A638

Email: homing@gate.sinica.edu.tw

Phone: 2787-2073