# Raw Materials for Literature Curation

Where do the publications come from?

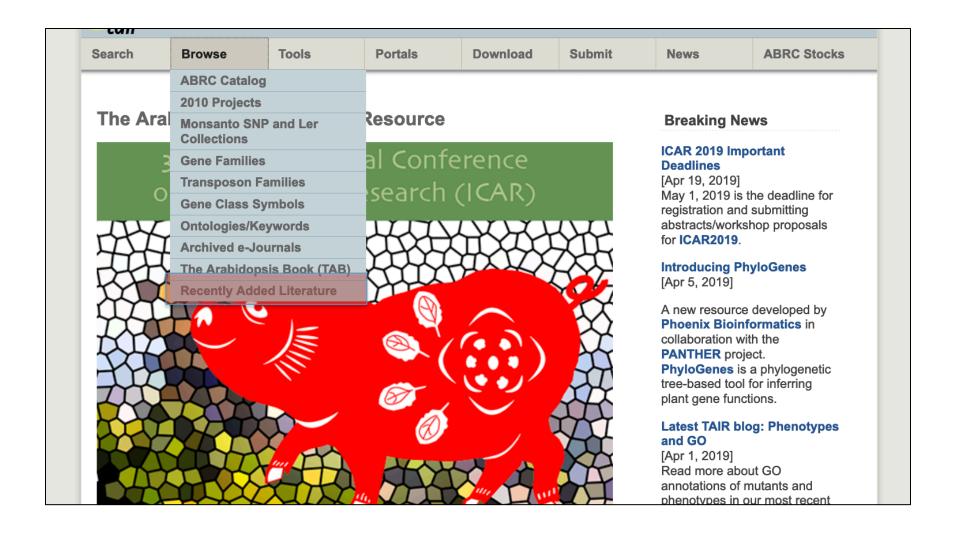
# Locus Page

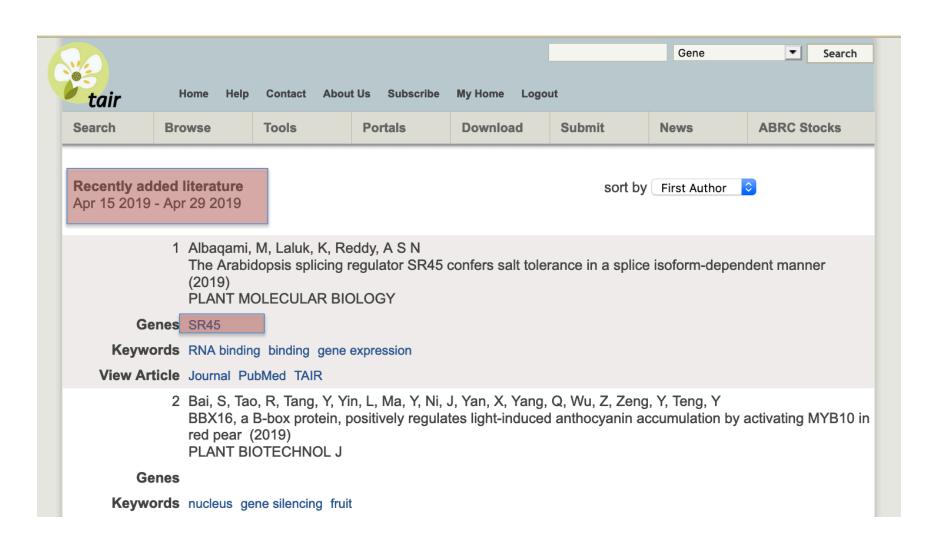
Locus: AT	IG16610	What's new on this page	A	dd a Comment	
Representative Gene Model @	AT1G16610.3				
Gene Model Type	protein_coding				
Other names:	ARGININE/SERINE-RICH 45, ATSI	R45, RNPS1, SR45			
Description 0	Encodes SR45, a member of the highly conserved family of serine/arginine-rich (SR) proteins, which play key roles in pre-mRNA splicing and other aspects of RNA metabolism. SR45 is a spliceosome protein, interacts with SR33 and the U1-70K protein of the U1 snRNP. Also involved in plant sugar response. sr45-1 mutation confers hypersensitivity to glucose during early seedling growth.				

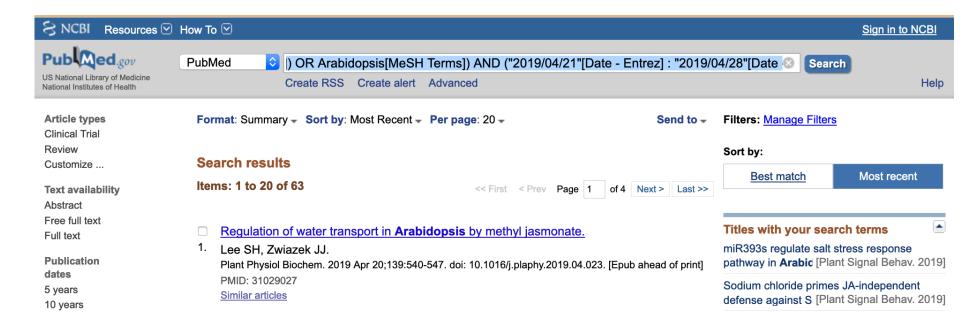
Publication 0	author/title	source	associated loci	date
	Albaqami, M, Laluk, K, Reddy, A The Arabidopsis splicing regulator SR45 confers salt tolerance in a splice isoform-dependent manner	PLANT MOLECULAR BIOLOGY	AT1G16610	2019
	Reddy, A CORRECTION: Transcriptome- Wide Identification of RNA Targets of Arabidopsis SERINE/ARGININE- RICH45 Uncovers the Unexpected Roles of This RNA Binding Protein in RNA Processing	THE PLANT CELL	AT1G16610	2017
	Zhang, X N, Shi, Y, Powers, J J, Transcriptome analyses reveal SR45 to be a neutral splicing regulator and a suppressor of innate immunity in Arabidopsis thaliana	BMC GENOMICS	AT1G16610	2017

# Polymorphism Page

	hism: dcl4	1-4							
Name 0	dcl4-4								
Date last modified @	2005-12-08								
Tair Accession	Polymorphism:	1009287001							
Type 🛭	substitution								
Chromosome	5								
Associated Genes	Gene Model	Locus 0		Polymorphism site	Associa	ation Type 🛭			
	AT5G20320.1	AT5G20320		exon	is an all	ele of			
	Description   ②	in a distinct small siRNAs from both	RNA biogenes inverted-repea	me that catalyzes procession pathway. The protein is at constructs and endogered in long-range cell-to-cell	also involved in the pro nous sequences, as well	duction of 21-nt p as the RDR6-de	orimary pendent		
Associated Loci	AT5G20320								
Mutagen 0	ethylmethane s	sulfonate							
Allele Type @	loss-of-function	1							
Description	The substitution in exon 4 generates a premature stop codon at amino acid 186 (Q186STOP).								
	lymorphisms 🛭	,							
Substitution	Species Varia	int (attribution) 🕡	Length Pol	lymorphic Sequence 0	Polymorphism Verifie	d			
Germplasm	Name/Image	Polym	orphisms	Background	Stock Na	me 🛭 S	elect		
Showing 1 of	DCL4-4	dcl4-4			not an AB	RC stock			
1 entries	Phenotype @	1							
<ul> <li>Plants have altered silencing movement with respect to the parent line. Whereas the parent line has low levels of SUL mRNA and displays non-cell autonomous silencing, the mutants lost the silencing movement phenotype and accumulated high levels of SUL mRNA. 21-nt SUL siRNAs are undetectable in these plants, wherease the 24-nt siRNAs accumulated to similar levels as in the SUC-SUL parental line. Dunoyer, et al. (2005)</li> </ul>									
Community Co	omments 🛭 (sh	ows only the most	recent comn	nents by default)					
		Add My Comn	nent Hi	de Comments Show	v All Comments				
Publication	title					source	date		
0		is required for RN/ A component of the		and produces the 21-nucle cell silencing signal.	eotide small	NATURE GENETICS	2005		







### Semi-automated Approach

- NCBI's E-utilities: ESummary and EFetch
  - Send keywords and time restrictions
    - ((Arabidopsis[Title/Abstract]) OR Arabidopsis[MeSH Terms]) AND ("2019/04/21"[Date - Entrez] : "2019/04/28"[Date - Entrez])
  - Return XML records of publications
- Custom scripts
  - Parse XML and load into TAIR curation DB

### **Pubsearch**

[Add Article / Articles in Bulk / Gene / Locus / Term / Communication / Allele / Symbol / Community Reference / Germplasm / User / Community Submission Other Annotations Bulk Annotation Upload ]
[Browse Gene Ontology / Tair Ontology / Annotation Tasks ]
[Submit SourceForge Bug Report ]

#### LoadArticles - Fetching Literature Data from Pubmed

Add Journals: <

Do Fuzzy Search: <a>V</a>

Search Type: Relative Range

Range Search: 2019/04/21 - 2019/04/28

yyyy/mm/dd yyyy/mm/dd

Reset Submit

# Triage

my papers grab	new papers	my sc	anned pape	ers in	nvalid pap	ers	temp p	apers				
	<<	1	2	3 4	5	6	7	8	9	10	11	>>
Papers Total: 54	[93400]	THE PLAN	T JOURNAL	L (2019)	]	HIGH PRIC	ORITY ]	[H	as Hit]	[NO PDF]	[2019-04-29	) ]
Order By ? priority  ▼ GO				lgi network t ative tetherir			acuolar ti	raffickin	g and T	GN		
	DisplayAbstr	act			l	JpdateArtic	cle					
Search For ?												
GO	[93377]	THE PLAN	T JOURNAL	L (2019)	l	HIGH PRI	ORITY ]	[H	las Hit]	[NO PDF]	[2019-04-29	]
				ctor MYB6 p ondary cell v								
	DisplayAbstr	act				UpdateArtic	cle					
Reload Papers	[93385]	PLoS bio	logy (20)	19)		[HIGH PRI	CORITY ]	[	las Hit]	[NO PDF]	[2019-04-29	)]
Import Papers		ear phenoi lopsis thali		tion uncover:	s the eme	rgence of	heterosis	in				
	DisplayAbstr	act				UpdateArti	icle					

# Link

lcus AT1G17980 ( Article [ 93404 ] (2019) [research_article] The poly(A) polymerase PAPS1 interacts 12 [title] Yes				
to de (2000)	Term Name	Article Title and Abstract	Hit details	Hit Valid?
Symbols: PAPS1 opoly(A) polymerase 1> TAIR:locus:2030943 (Accession) Searched on 2019-04-30 Maybe Université Italicus: Control of the public development TAIR: No Searched on 2019-04-30 Maybe Université Université	pub:147032) Symbols: PAPS1 <poly(a) 1="" polymerase=""> TAIR:locus:2030943</poly(a)>	with the RNA-directed DNA methylation pathway in sporophyte and pollen development THE PLANT JOURNAL Zhang, Y., Ramming, A., Heinke, L., Altschmied, L., Slotkin, R. K., Becker, J. D., Kappel, C., Lenhard, M. (TAIR Reference:null)  RNA-based processes play key roles in the regulation of eukaryotic gene expression. This includes both the processing of pre-mRNAs into mature mRNAs ready for translation and RNA-based silencing processes, such as RNA-directed DNA methylation (RdDM). Polyadenylation of pre-mRNAs is one important step in their processing and is carried out by three functionally specialized canonical nuclear poly(A) polymerases in	[abstract] Searched on	No Maybe

# Ready for rest of curation

- Gene summary
- Symbols
- Alleles
- Phenotypes
- Germplasms
- GO and PO annotations



### MaizeGDB Curation of Data in References



### MaizeGDB Curation of Data in References

Editorial Board: 5 community members recommend papers monthly.

Paper with large

Paper on few genes
--> genes, alleles,
phenotype Data
determined

alleles,
e Data
ed
ed
explicit asks, or curator approval.

Curation tools used (2003), Curator only

Custom spreadsheet, Data entered directly, Curator and Database Manager

Seed Stock curator goes through Pubmed daily, for Papers on genes, alleles, phenotypes



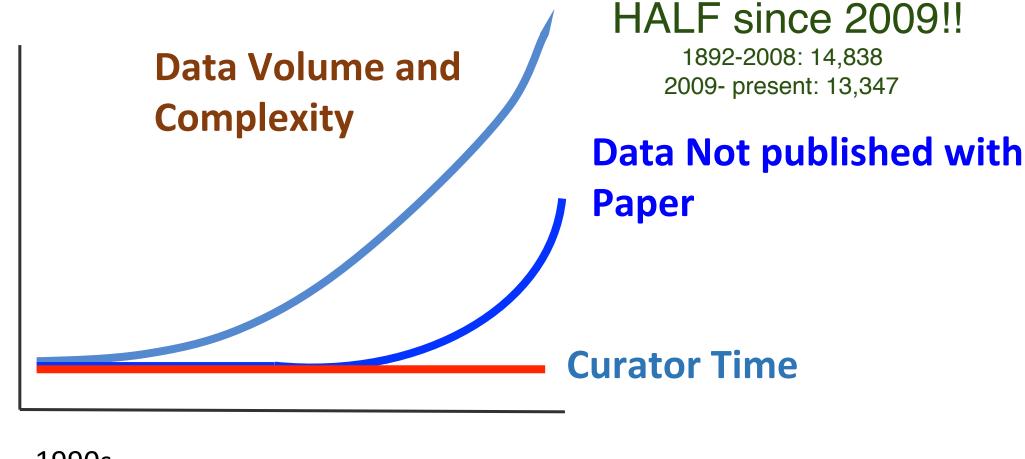
genes, alleles, phenotype Data determined



Curation tools used (2003), Curator only

## Data vs Curation: Maize Publications

- ----- >28,000 Articles-Pubmed "maize" title/abstract



1990s

3