

TAIR Workshop

Part III: Community annotation

Coupling data submission with publication -
TAIR's innovative journal collaboration effort

- Why it matters
- How it works

Challenge for scientific database

TAIR:

How to keep up with the pace of data increase with limited resources?



Arabidopsis literature

- 200+ new articles each month
- 26,000+ full text articles
- 2 FTE gene function curators with many other tasks besides literature curation

Challenge for research community

Community:

Data accessibility?

New tools to analyze/maximize the value of data?

Research community and database working together

TAIR:

How to keep up with the pace of data increase with limited resources?



**increased demand for
data management,
analysis and access**

Community:

Data accessibility?

New tools to analyze/maximize the value of data?

Research community and database working together

TAIR:

How to keep up with the pace of data increase with limited resources?

**enhanced productivity
new methods for discovery**



**increased demand for
data management,
analysis and access**

Community:

Data accessibility?

New tools to analyze/maximize the value of data?

'First of its kind' journal database collaboration to promote community curation

Plant Physiology
Physiology



with the
Illumina Genome Analyzer.

HOME | HELP | FEEDBACK | SUBSCRIPTIONS | ARCHIVE | SEARCH | TABLE OF CONTENTS

Institution: **Stanford University** [Sign In as Member](#)

Perform your original search, **plant physiology tair**, in *Plant Physiol.* [Search](#)

Plant Physiology 146:1022-1023 (2008)
© 2008 [American Society of Plant Biologists](#)

EDITORIAL

***Plant Physiology* and TAIR Partnership**

We are pleased to announce a unique partnership between *Plant Physiology* and The Arabidopsis Information Resource (**TAIR**) that we anticipate will increase the annotation of *Plant Physiology* in the **TAIR** database to nearly 100%. This partnership is the first of its kind, and, if successful, we hope this model can be extended to other journals, organisms, and databases.

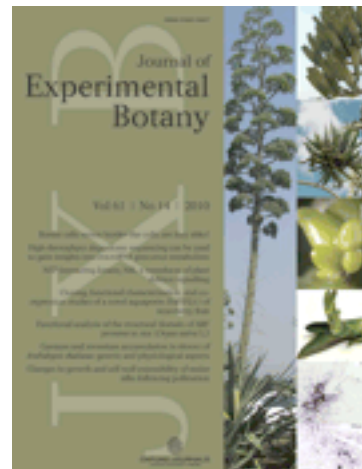
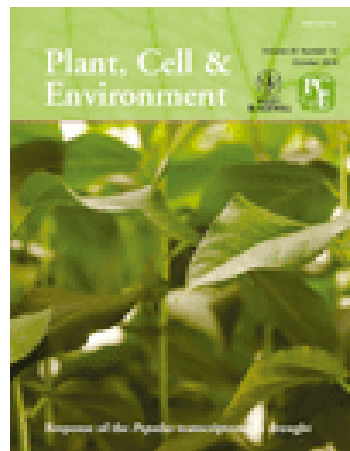
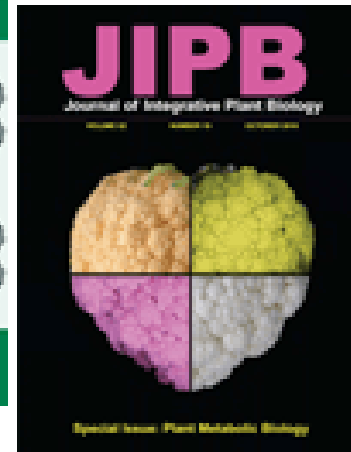
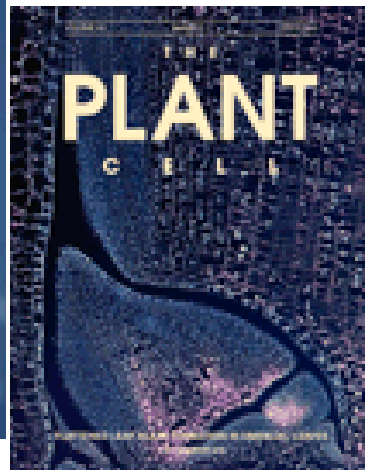
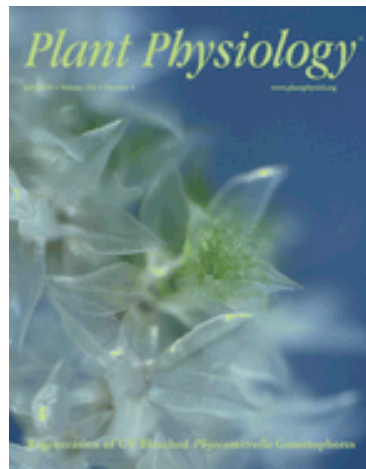
This Article

- ▶ [Full Text \(PDF\)](#)
- ▶ [Alert me when this article is cited](#)
- ▶ [Alert me if a correction is posted](#)

Services

- ▶ [Email this article to a friend](#)
- ▶ [Similar articles in this journal](#)
- ▶ [Similar articles in Web of Science](#)
- ▶ [Alert me to new issues of the journal](#)
- ▶ [Download to citation manager](#)
- ▶ [© Get Permissions](#)

TAIR journal partners



TAIR journal partners



ELSEVIER



 WILEY-BLACKWELL

The Wiley-Blackwell logo features a circular symbol on the left, containing a stylized 'W' and 'J'. To the right of the symbol, the words "WILEY-BLACKWELL" are written in a bold, sans-serif font.

How to submit your data

Tools	Portals	Download	Submit	News	ABRC Stocks
			Submit Overview		
			ABRC Stock Donation		
			External Links		
			Gene Class Symbol Registration		
			Gene Family Data		
			Online Submission for Authors and Others		
			Gene Structure Additions/Modifications		
			Locus Identifier Request		
			Marker and Polymorphism Data		
			Metabolic Pathway Data		
			Phenotypes		
			Protocols		
			2010 Functional Genomics Gene List		

[Home](#) > [Data Submission](#) > [Literature-Based](#)

Online Gene Function Submission

Authors are encouraged to submit their gene function data. We also welcome submission of data from older articles by the original author on the article. Gene function data accepted includes gene expression (for example, Northern blot), protein kinase, localization (cellular, sub-cellular), development, and interacting partners. If you have data, please choose the appropriate form from our [Submit Overview](#) page.

All submissions will be reviewed by a curator and will not be released until the relevant publication is published.

Have just a few genes?

Login or register to fill out our online submission form. Login page, and you will be forwarded to the online submission page.

TAIR curators will review your submission and contact you via the e-mail from your user profile.

[Fill in Online Form](#)

Have a large dataset?

Last modified on May 6, 2010.

What do we ask from users?

Have just a few genes?

Fill out our online submission form. * means a field is **required**.

TAIR curators will review your submission and will get in touch with any issues or questions using the e-mail from your user profile.

1

*Article Id:

Must enter article id

☒ Digital Object Identifier (DOI)

☐ Pubmed ID

Gene Search

2

*Locus Name

AT2G23380

Symbol

CLF

Symbol Full Name

CURLY LEAF

*Add Information

Click links below to add...

3

Wooden Leg

Molecular Function

Biological Process

Subcellular Localization

Expression

Interacting Partner/s

Other

What do we ask from users?

3

Molecular Function Annotations

Annotating locus from article

Molecular Function
Examples:
ion transmembrane transporter activity
protein kinase activity
phytoene synthase activity
Start typing, then choose from list or add a new term,
then click outside the field to accept the new term.

Method
Example:
Enzyme assays
Choose a method or enter a new one.

histidine

Choose method or enter a new one... Delete

Term	Id	Synonym
CAU codon-amino acid adaptor activity	GO:0033425	histidine tRNA
CAC codon-amino acid adaptor activity	GO:0033426	histidine tRNA
protein histidine kinase binding	GO:0043424	histidine kinase binding
protein histidine kinase activity	GO:0004673	histidine kinase activity
polar-amino acid-transporting ATPase activity	GO:0015426	histidine permease activity
histidine-tRNA ligase activity	GO:0004821	histidine translase activity
L-histidine:2-oxoglutarate aminotransferase activity	GO:0008110	histidine transaminase activity
L-histidine transmembrane transporter activity	GO:0005290	L-histidine transporter activity
histidine decarboxylase activity	GO:0004398	
histidine ammonia-lyase activity	GO:0004397	
L-phenylalanine:pyruvate aminotransferase activity	GO:0047312	histidine aminotransferase activity
histidine N-acetyltransferase activity	GO:0047981	histidine acetyltransferase activity

What do we ask from users?

'Article Id: ☐ Digital Object Identifier (DOI)
☒ Pubmed ID

Must enter article id

Gene Search

'Locus Name <i>AT2G23380</i>	Symbol <i>CLF</i>	Symbol Full Name <i>CURLY LEAF</i>	'Add Information <i>Click links below to add...</i>	
<input type="text" value="AT2G01830"/>	<input type="text" value="WOL"/>	<input type="text" value="Wooden Leg"/>	Molecular Function (1) Biological Process (3) Subcellular Localization Expression Interacting Partner/s (1) Other (1)	<input type="button" value="Delete"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	Molecular Function Biological Process Subcellular Localization Expression Interacting Partner/s Other	<input type="button" value="Delete"/>

Multiple genes?

[Home](#) > [Data Submission](#) > [Literature-Based Gene Function Data](#)

Online Gene Function Submission Form for Authors and Others

Authors are encouraged to submit their gene function data to TAIR at the time of publication. We also welcome submission of data from older articles by any community member whether or not you are an author on the article. Gene function data accepted by this form include molecular function (for example, protein kinase), localization (cellular, sub-cellular or gross anatomy), biological role (for example, seed development), and interacting partners. If you have other types of data to submit please choose the appropriate form from our [Submit Overview](#) page.

All submissions will be reviewed by a curator before making the data public and will not be released until the relevant publication is published.

[List of submitters](#)

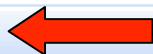
Direct Submission to TAIR

Login or register to fill out our online submission form. Please click on the button below to go to the Login page, and you will be forwarded to the online submission form once you've logged in.

TAIR curators will review your submission and will get in touch with any issues or questions using the e-mail from your user profile.

[Fill in Online Form](#)

Downloadable Spreadsheet



Community annotation on TAIR

TAIR Annotation Search [\[Help\]](#)

Your query for annotations based on the publication **An allelic mutant series of ATM3 reveals its key role in the biogenesis of cytosolic iron-sulfur proteins in Arabidopsis** resulted in 5 records.

Displaying 1 - 5 of 5 records on page 1 of 1 pages.

Locus/ Gene Model	Gene Symbol/Full Name	Relationship Type	Keyword	Keyword Category	Evidence Code [?] : Evidence Description [?] : Evidence With: Reference [?]	Annotated By/ Date Last Modified
AT5G58270	ABCB25/ ATP-BINDING CASSETTE B25	involved in	root development	biological process	<i>inferred from mutant phenotype:</i> analysis of visible trait: none: Bernard, et al. (2009)	Janneke Balk/ 2009-09-29
AT5G58270	ABCB25/ ATP-BINDING CASSETTE B25	involved in	regulation of chlorophyll biosynthetic process	biological process	<i>inferred from mutant phenotype:</i> biochemical/chemical analysis: none: Bernard, et al. (2009)	Janneke Balk/ 2009-09-29
AT5G58270	ABCB25/ ATP-BINDING CASSETTE B25	involved in	chloroplast organization	biological process	<i>inferred from mutant phenotype:</i> analysis of visible trait: none: Bernard, et al. (2009)	Janneke Balk/ 2009-09-29



Progress so far

Community annotation at TAIR

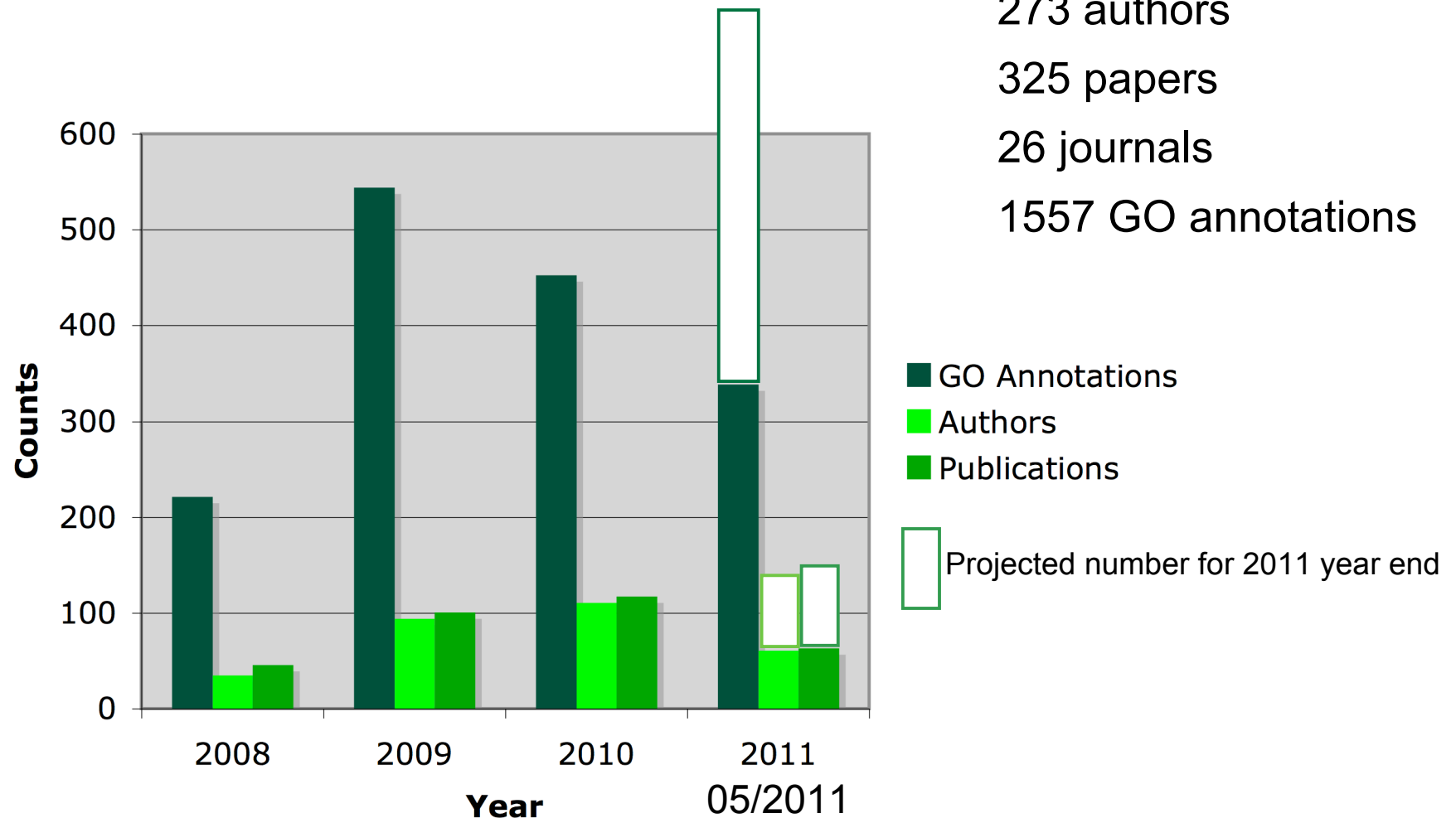
02/2008 - 05/2011

273 authors

325 papers

26 journals

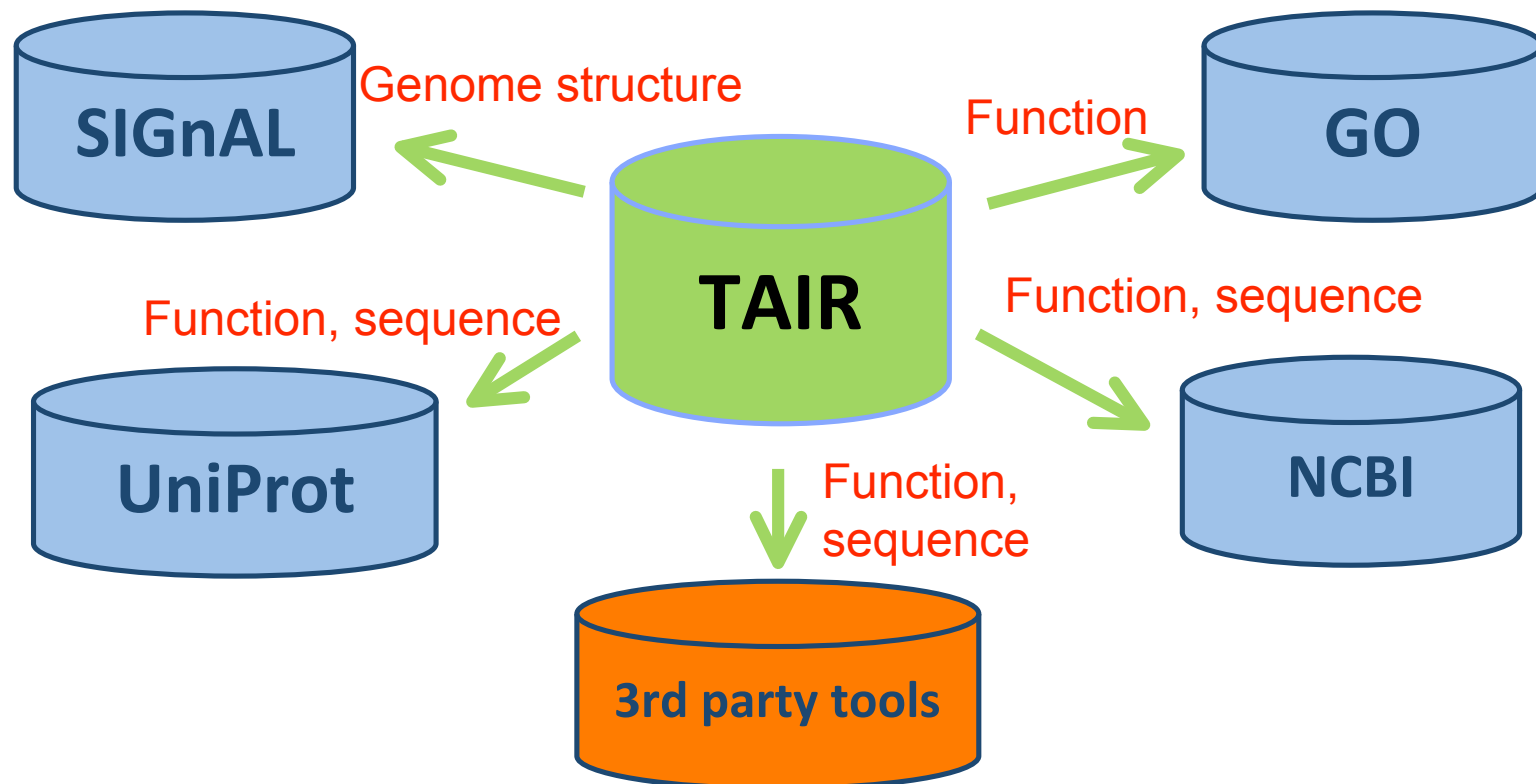
1557 GO annotations



Community feedback

- “I think the new system is a nice idea and may facilitate new submissions.”
- “I do profit a lot from the data on TAIR, thus this submission is a small contribution to extend the data present on TAIR.”
- “I gratefully did it [data submission] because I already benefit from similar information for other genes.”

Why should everyone participate - TAIR is the primary source for Arabidopsis function/structure data



★ Reviewed, UniProtKB/Swiss-Prot **Q9SP32** (DICER_ARATH)

Last modified October 13, 2009. Version 70. [History...](#)

Clusters with 100%, 90%, 50% identity | Documents (2) | Third-party data | Customize

Gene Ontology (GO)

Biological process

RNA interference, production of lsiRNA

Inferred from mutant phenotype. Source: TAIR

RNA interference, production of ta-siRNAs

Inferred from mutant phenotype. Source: TAIR

cytokinesis [Ref.1](#)

Inferred from mutant phenotype. Source: TAIR

embryonic pattern specification

Inferred from mutant phenotype. Source: TAIR

flower development

Traceable author statement. Source: TAIR

gene silencing by miRNA, mRNA cleavage

Inferred from mutant phenotype. Source: TAIR

Data from TAIR

NCBI Entrez Gene

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals

Search Gene for

Limits Preview/Index History Clipboard Details

Display Full Report 20

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 0

☐ 1: DCL1 DCL1 (DICER-LIKE 1); ATP-dependent helicase/ double-stranded RNA binding / protein binding / ribonuclease III [*Arabidopsis thaliana*]

GeneID: 839574 updated 04-Oct-2009

Summary

Gene name	DCL1
Primary source	TAIR:AT1G01040
Locus tag	AT1G01040
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Arabidopsis thaliana (ecotype: Columbia)
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids 11; brassicales; brassicaceae; Arabidopsis
Also known as	ABNORMAL SUSPENSOR 1; ASU1; ATDCL1; CAF; CARPEL FACTORY; DCL1; DICER-LIKE 1; EMB60; EMB76; EMBRYO DEFECTIVE 60; EMBRYO DEFECTIVE 76; SHORT INTEGUMENTS 1; SIN1; SUS1; SUSPENSOR 1; T25K16.4; T25K16_4

Why should everyone participate - increased exposure of your work

p	Keyword	Keyword Category	Evidence Code ②:	Annotated By/ Date Last Modified
			Evidence Description ②: Evidence With: Reference ②	
	root development	biological process	<i>inferred from mutant phenotype:</i> analysis of visible trait: none: Bernard, et al. (2009)	Janneke Balk/ 2009-09-29
	regulation of chlorophyll biosynthetic process	biological process	<i>inferred from mutant phenotype:</i> biochemical/chemical analysis: none: Bernard, et al. (2009)	Janneke Balk/ 2009-09-29
	chloroplast organization	biological process	<i>inferred from mutant phenotype:</i> analysis of visible trait: none: Bernard. et al. (2009)	Janneke Balk/ 2009-09-29

Why should everyone participate - increased exposure of your work

OPEN  ACCESS Freely available online



Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

Background. Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available. **Principal Findings.** We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. **Significance.** This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

A few things to keep in mind

1. Pre-publication

Register your gene symbol to minimize accidental duplications in gene nomenclature


AT1G56650 [PAP1](#)
PRODUCTION OF ANTHOCYANIN PIGMENT 1

AT2G01180 [PAP1](#)
PHOSPHATIDIC ACID PHOSPHATASE 1

AT2G27190 [PAP1](#)
PURPLE ACID PHOSPHATASE 1

AT3G16500 [PAP1](#)
PHYTOCHROME-ASSOCIATED PROTEIN 1

Submit	News
Submit Overview	
ABRC Stock Donation	
External Links	
Gene Class Symbol Registration	



A few things to keep in mind

2. Preparing your manuscript

Include AGI locus identifiers !!! (*main text preferred*)

One name, different genes

AT1G56650 PAP1 PRODUCTION OF ANTHOCYANIN PIGMENT 1

AT2G01180 PAP1 PHOSPHATIDIC ACID PHOSPHATASE 1

AT2G27190 PAP1 PURPLE ACID PHOSPHATASE 1

AT3G16500 PAP1 PHYTOCHROME-ASSOCIATED PROTEIN 1

One gene, many names

AT1G01040

ASU1: ABNORMAL SUSPENSOR 1

DCL1: DICER-LIKE 1 / ATDCL1:DICER-LIKE 1

CAF: CARPEL FACTORY

SIN1: SHORT INTEGUMENTS 1

SUS1: SUSPENSOR 1

EMB60: EMBRYO DEFECTIVE 60

EMB76: EMBRYO DEFECTIVE 76

conflicting nomenclature / error not uncommon

conflicting nomenclature / error in publication not uncommon

Plant Cell Physiol. 2010 Jun;51(6):866-76

PSII. Recent bioinformatic and proteomic studies suggested that the two PQL proteins, PQL1 (At1g14150) and PQL2 (At3g01440), might function in the chloroplast NAD(P)H

Plant Cell Physiol. Jun;51(6):877-83

et al. 2007). The Arabidopsis luminal proteome has also revealed two proteins, **AT3G01440** and **AT1G14150**, with similarity to PsbQ (Friso et al. 2004). Following the nomenclature of Ishihara and co-workers, we named these proteins PsbQ-like 1 (**PQL1**) and **PQL2**, respectively. Recently, co-migration of the

A few things to keep in mind

THE **PLANT** CELL

[HOME](#) | [ABOUT](#) | [SUBMIT](#) | [SUBSCRIPTIONS](#) | [ADVERTISE](#) | [ARCHIVE](#) | [CONTACT](#)

Institution: Stanford University Sign In as Member / Individual

© 2011 American Society of Plant Biologists



Accession Numbers

PMID:21447788

Sequence data from this article can be found in the Arabidopsis Genome Initiative database under the following accession numbers: *pPLAI*, At1g61850; *pPLAI α* , At2g26560; *pPLAI β* , At4g37050; *pPLAI γ* , At4g37070; *pPLAI δ* , At4g37060; *pPLAI ϵ* , At5g43590; *pPLAII α* , At2g39220; *pPLAII β* , At3g54950; *pPLAII γ* , At4g29800; *pPLAII δ* , At3g63200; *UBQ10*, At4g05320; and β -tubulin, At5g23860.

A few things to keep in mind

3. Post-publication

Submit your annotation to us

1) Any journal - not limited to a journal that has an existing collaboration with TAIR

2) Not limited to papers you authored

Developing new methods for function annotation

1. Semi-automated GO annotation in collaboration with WormBase (Caltech)
2. Functional annotation across genomes using PAIN (Protein Annotation Inferencing Tool) (in collaboration with GO Consortium)

Acknowledgements

Many thanks:

- o Collaborating journals, publishers
- o Authors who have contributed data to TAIR
- o Collaborating databases
- o The community!

CARNEGIE
INSTITUTION FOR
SCIENCE



TAIR Sponsors



Dow AgroSciences

syngenta®



GRASSROOTS
BIOTECHNOLOGY

download this talk from <http://min.us/u/Tair>

Come see our posters!

#360

An Equation For A Vibrant Database: Curators + Journals + Community = Success

#371

Arabidopsis genome annotation and large-scale analysis tools at TAIR

Stop by TAIR booth!

Booth 12

Making your community database better by providing your feedback:
curator@arabidopsis.org

Quality of community annotation

We analyzed submissions from 50 papers, used online tool
Reviewed author submissions, looked for errors/missing data

Findings:

1. Few errors: community-submitted annotations can almost always be integrated into the TAIR database (98%)
2. Specificity: Curator and author annotations sometimes differ in the specificity of the terms used. [protein binding \(author\)](#) vs [protein self association \(curator\)](#)
3. Completeness: 25/50 did not miss any annotation; 10/50 missed 1 annotation; 15/50 missed 2 (or more) annotations

Bottom line:

annotations are almost always correct, just different levels of specificity and focus