Data Submission

Guidelines for various datatypes that can be submitted to TAIR follow below. Please contact curator@arabidopsis.org if you have any questions or to submit your completed files for processing.

- External Links
- Gene Class Symbol Registration
- Gene Function Data
- Gene Structure Additions/Modifications
- Marker and Polymorphism Data
- Phenotypes
- Protocols

External Links

We encourage users of TAIR to submit external links associated to data objects found in TAIR. Submitted links will be added to the external link band on the object detail pages. Data objects accepted include: Loci, Clones, Genes, Polymorphisms, Genetic Markers, and Clone Ends. The locus example below would create a link on this locus detail page.

NOTE: If you want to link to ALL Arabidopsis LOCI and are using the AGI code (e.g. AT4G32520) as part of the variable for the URL, just send the External website name, base URL, and variable syntax to: curator@arabidopsis.org

If you would like to submit links for **a subset of all the data objects** found in TAIR, please submit your data in one of the following ways:

Option 1) Use a preformatted Excel spreadsheet. To submit your data, please download and complete the following Excel spreadsheet. Note that examples and instructions have been included. Before submitting your data, review your entries to ensure that the data is correct.

Note: Macros must be enabled for this form to work properly. To allow the macros in this form to run, please change your macro security level to medium (recommended) or low. From the Tools menu, choose Macro, then Security. After you change the security level to medium or low, you will have to restart Excel.

Download: external_link_data_form.xls

Option 2) Send tab-delimited data from any program. If you want to create your own file please follow the following format in submitting your data and include the column headers as the first line in your files.

Field	Description	Example
Object Name	The name of the object found in TAIR to be linked	AT4g32520 or ATDMC1.1
Object Type	Type of object. Select from: Locus, Clone, Gene, Polymorphism, GeneticMarker	Locus or GeneticMarker
External Web Site - Name	Name of the website that the Base URL refers to	View AraCyc information
External Link - Base URL	Base rule for the URL. This is usually the portion of the external link URL before the question mark (?)	http://www.arabidopsis.org: 1555/ARA/NEW-IMAGE
External Link - Variable	Variable for the URL. This is usually the portion including and following the question mark (?)	?type=REACTION-IN-PATHWAY& object=GLYOHMETRANS-RXN
Display Name	Name of the link to be displayed. For example, if the link is for an associated marker, you can provide the marker name.	formylTHF biosynthesis

Gene Class Symbol Registration

In order to register a Gene Class Symbol, **you must be registered at TAIR**.

Register New Gene Class
Symbol

These symbols typically consist of 3 or 4 letters that define either a single gene (ABC) or a gene family (ABC1, ABC2, ABC3). Although a number of classical genetic loci were assigned 2-letter symbols years ago, the continued use of 2-letter symbols to name new loci is strongly discouraged except in cases where there is a compelling reason based on the underlying science. A similar justification should be provided for the use of gene class symbols with more than 4 letters. Symbols may describe a mutant phenotype or some aspect of gene structure or function.

Please use ALL UPPERCASE letters for the Mutant Symbols (ex: EMBRYONIC LETHAL) and all lowercase letters for Gene Product Symbols (except when referring to domains or other symbols; examples: chloroplast J-like domain; TON1 recruiting motif)

ⓘ

ⓘ This registration process is designed to minimize accidental duplication in gene nomenclature by comparing the symbol being registered with both the Registered Symbol list and searching TAIR to see if the symbol is unregistered but already published. If the symbol is either registered or published the registration will not be accepted.

What Constitutes a Gene Class Symbol?

Gene class symbols have been divided into two major categories (mutant phenotype and gene product) to facilitate curation. Mutant phenotype symbols should be used when a mutant is available and the symbol describes some aspect of the phenotype. Gene product symbols should be used regardless of the availability of a mutant when the symbol describes some aspect of gene structure or function. The use of organism specific prefixes such as At or Ath is discouraged as this is redundant and leads to a lot of genes named 'Arabidopsis thaliana X'.

Examples of Mutant Phenotype Symbols:

- ABA: ABA DEFICIENT
- AGE: AUXIN-RESPONSIVE GENE EXPRESSION
- EMB: EMBRYO DEFECTIVE

Examples of Gene Product Symbols:

- ADH: alcohol dehydrogenase
- AGL: AGAMOUS-like
- COR: cold regulated

More Nomenclature documentation can be found here.

Gene Function Data

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), interacting partners, and comments (gene summaries.) 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.

The online submission form has been tested with Chrome and Firefox. Please use one of those browsers.

To use our Generic Online Annotation Tool (GOAT), you must have an ORCiD. If you have not already done so, please create one at the ORCID website. GOAT is a 'paper based' curation form, used for curating experimental gene function from a **published, peer reviewed** research article, so you will also need to have the PubMed ID or Digital Object Identifier for your paper handy. When you click on the button below, GOAT will start and prompt you for your ORCiD and password. Using GOAT, you can enter multiple loci, each with its separate set of functional annotations, and submit them all at once to the TAIR curators.

To get started click the link below and follow these steps. Or watch a You Tube tutorial first

Complete Online
Form

1. Log in with your ORCiD
2. Click on the Submission link in the upper left menubar to start a new submission or continue one in progress
3. Enter in the DOI or Pubmed ID for the paper you are curating
4. Enter in ID's for every locus you are curating from the paper. GOAT accepts UniProt IDs, AGI Locus IDs and RNA central IDs
5. Add your annotations. Choose the type of annotation from the drop down menu. You can choose to make Gene Ontology annotations (cellular component, molecular function, biological process), Plant Ontology annotations (plant developmental stages or anatomy), protein interactions or comments
6. Review your annotations before you submit
7. After you complete your submission, TAIR curators will review the information. If we have any questions, we will get in touch with you

We also provide a preformatted Excel spreadsheet. Please click on the link to download the spreadsheet. Required fields are **blue** . Fields with an asterisk (*) can have more than one entry separated by a pipe with no intervening spaces (e.g. Ecker,Joseph|Bell,Callum). If submitting a manuscript to The Plant Journal, please include this file in your TPJ submission as a supplemental file.

File Format

Field	Description	Values/Constraints	Examples
Contact Person/ Submitter *	Name of the person we can contact if we have questions about the annotation	none	Eva Huala

Locus Identifier	AGI locus id or name of genetic locus	At#g#####, see AGI coding convention if the gene is new, splitting or merging existing genes	AAAP or At1g10010
Gene Name	symbol-based name of gene, if this exists	See nomenclature guidelines	ETO1
Reference *	PubMedID or citation	PubMed ID can be obtained from TAIR Publication Search or PubMed	PubMedID:14576282; Smith (2006) Science 123:23; DOI:10.1111/j.1365-313x. 2010.04399.x; submitted to Plant Journal
Gene function, process, location or interacting partner	functional annotation that you'd like to make	For interacting partners, please use the AGI code (At1g01010.1) in addition to the gene name	ser/thr kinase, outer mitochondrial membrane, anther, gynoecium morphogenesis, At1g01010.1 (ABC1)
Method description	Method and evidence used to support the functional annotation	A list of method descriptions is given as the second sheet of the Excel workbook	enzyme assay, mutant phenotype, transcript levels, yeast two-hybrid-assay

Gene Structure Additions/Modifications

ⓘ **tl;dr** Email us. 🙂
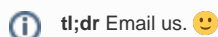
I have identified a new gene not in the current genome release. How can I add that information to TAIR?

Please contact TAIR to request a new locus identifier (i.e. AGI Locus code). Once you have the new identifier please contact TAIR to provide any information you have about the function or expression pattern of the new gene. If you are registered at TAIR, you can submit functional annotation information directly or register a new gene symbol. Registration is free.

I think the structure of a gene in the current annotation is incorrect. How can I correct it?

If you have sequence data that indicate the current annotation for a gene is incorrect, please contact TAIR to update the structure.

I think the functional annotation for a gene is incorrect. How can I correct it?

If you have data from a published article indicating that information about the expression, localization or biological role of a gene is incorrect, please contact TAIR. An example of an incorrect annotation would be if an article shows a gene to be expressed in the roots (but not shoots) and that same article is used as evidence for gene expression in shoots. If you have data that contradict other published works, we strongly encourage you to submit your own functional annotations and have them included in the gene record.

I know of published information on gene function that is missing from TAIR. How can I include it?

Please provide the missing gene function information using our online submission tool, GOAT. This form can be used by anyone to submit published data, whether you are an author of the publication or not. You will need to log in before filling out the form so that we can associate your name to the submission. If you experience difficulty or have information that will not fit within the form, please contact us.

Marker and Polymorphism Data

If you are submitting more than 150 entries, please e-mail curator@arabidopsis.org for additional instructions otherwise please submit your data by using the following preformatted spreadsheet:

Download: marker_polymorphism_data_form.xls

Instructions on how to use the preformatted spreadsheet:

If available, always indicate the digest pattern/PCR length/base pair of Col-0 as a reference (not necessary for deletions/insertions).

Examples:

Marker Name	Marker Type	Polymorphism Type	Polymorphism	Restriction Site	Ecotype
MARKER1	CAPS	digest_pattern	400;230	1	Col-0
MARKER1	CAPS	digest_pattern	630	0	Hi-0
MARKER2	SNPlex	SNP	G		Col-0
MARKER2	SNPlex	SNP	A		Hi-0

In the case where a marker or polymorphism is shared by several ecotypes, please list all ecotypes in the same row, as shown in this example:

Marker Name	Marker Type	Polymorphism Type	Polymorphism	Ecotype
MARKER3	SSLP	PCR_product_length	165	Col-0;Gr-6;Gu-0
MARKER3	SSLP	PCR_product_length	150	RLD;Wil-1;Wt-5

Fields indicated in **blue** are required. Fields marked with an asterisk (*) can contain multiple entries but these must be separated by a semicolon with no intervening spaces (e.g. Col-0;Ler-0). Note that examples have been included in the column headers; please retain headers as the first line in your files. You can find your Community ID and the Citation/Reference ID by using the links to our TAIR search pages provided in the template below.

Field	Description	Values/Constraints	Example
Contact Person Name*	Last name,First name	No space between first and last names or between semicolon	Buschmann ,Henrik
Contact Person Community ID*	unique identifier for a community member in TAIR. You must be registered to have a community ID	TAIR accession from TAIR Community Search	community: 1035
Contact Person E-mail Address	current email address		buschmann @gsf.de
Marker /Polymorphis m Name			PT1 or Salk_0228
Alternative Names*			AtPT1
Chromosome	The chromosome in which the marker/polymorphism is found	1,2,3,4,5 or unknown	5
Genetic Marker Type		CAPS, SSLP, AFLP, RFLP, RAPD, SNPlex or Taqman	CAPS
Polymorphis m Type	Dependent on Marker Type	Digest_pattern (CAPS or RFLP), PCR_product_length (SSLP, AFLP, or RAPD), substitution, insertion, deletion	Digest_patt ern
Polymorphis m Lengths *	Length of digest or PCR product	numeric, in kb or bp	0.12, 120
Polymorphis m Length Units		bp or kb	kb
Flank Sequence Type	Required for all but RFLP submissions	PCR_primer or Flank	flank
Flank Sequence 1	Required for all but RFLP submissions. The first PCR primer sequence		ATGGTGC CGTGACGT
Flank Sequence 2	Required for all but RFLP submissions. The second PCR primer sequence		AATTGGG TGTGCTAG
Special Conditions	Indicate any special conditions required for marker detection		annealing temp 62C
Restriction Enzyme Name*	Name of the restriction enzyme used to detect the polymorphism (required for CAPS, AFLP, or RFLP markers).	consult REBASE for standard enzyme name.	HindIII; EcoRV
Restriction Enzyme Number of sites*	Number of recognition sites for each restriction enzyme. The order of number of sites should match the order of restriction enzyme names in previous column.	numeric	3
Ecotype /Accession Name*	The ecotype which shows the specific polymorphism. If more than one ecotype shares a polymorphism, list all ecotypes.	Col-0 should always be shown as reference	Col-O;RLD; Ler
Accession Stock ID	For accessions that correspond to an ABRC stock, this is the stock ID number for that accession.		CS3455; CS3444
Polymorphic Sequence	Sequence of the polymorphic region for the given ecotype/accession		ATGTGGC CTCTT
Map Position1	Position of the SNP, insertion site, start position of 5' PCR primer/flanking sequence		20215474
Map Position2	End position of 3' PCR primer/flanking sequence		26877049
TAIR Version	version of the TAIR genome assembly that your experiment is based on		TAIR9
Inheritance	The mode of inheritance of the polymorphism	recessive, dominant, co-dominant, semi-dominant, unknown	dominant

Marker /Polymorphism Citation /Reference_id	For a reference that describe the marker/polymorphism, this is the unique identifier in TAIR. For articles referring to the markers that are not yet in TAIR, please contact a curator to update this information into the database when the paper is published.	TAIR accession from TAIR Publication Search	publication: 12344

Phenotypes

We accept phenotype data for all mapped and/or sequenced Arabidopsis mutants, including existing ABRC stocks, stocks from other stock centers, and lines that have not been deposited in a public repository. If you are submitting phenotype data for stocks not in ABRC we encourage you to consider also submitting the seed stock.

Please submit your phenotype data in one of the following ways:

Option 1) Use a preformatted Excel spreadsheet. To submit your data, please download and complete the following Excel spreadsheet. Examples are provided for each column.

Download: phenotype_data_form.xls

Option 2) Send tab-delimited data from any program. If you want to create your own file please follow the following format in submitting your data and send the files to: curator@arabidopsis.org. Please include the column headers as the first line in your files.

Description of column headers

Field	Description
Contact Person	Please provide name and email of the contact person/corresponding author
Reference (Pubmed ID, DOI, or other format)	PMID: 18485063 - we prefer Pubmed ID, but the following format without Pubmed ID is OK too:Plant J (2008), 55: 798
Allele Symbol / Polymorphism Name	For example: rpp3, rpp4(SALK_12345), atswc6(SAIL_1142_C03)
Accession	Example, Ws, Col-0
Locus	If this polymorphism is an allele of a locus, enter the unique locus code here. Otherwise, leave empty. Example: AT4G32520
Mutagen	EMS, X-rays, T-DNA insertion, transposon, fast neutron etc.
Mutation Site	Please describe the mutation site (3rd intron, 2nd exon, promoter, 3'UTR, 5'UTR etc). Please also give details of the mutation if known (for instance, contains a G to A subsitution at nucleotide position 123)
Genotype	Select from homozygous, heterozygous, hemizygous, and unknown.
Inheritance	Select from dominant, recessive, incompletely dominant, co-dominant, and unknown.
Allele Mode	Select from hypermorphic, hypomorphic, haplo-insufficient, antimorphic, gain-of-function, loss-of-function, and unknown.
Phenotype	Please provide a brief description of the mutant phenotype. Note: please specify if the phenotype is for double/triple mutants (for example, mop3/mop4).

Protocols

TAIR encourages the research community to share their protocols. To submit a protocol to TAIR, send the protocol in one of the following formats.

Portable Document Format (PDF)
Microsoft Word Document (.doc)
Image (e.g. for powerpoint/photoshop images)- .gif and .jpg
Any text file .txt and .rtf

Along with the above file, please send in additional data following the format. Fields indicated in **blue** are required. Fields marked with an asterisk(*) can have more than one entry but the entries must be separated by a semicolon.

To find your Community ID and the Citation/Reference ID please follow the provided links to our TAIR search pages.

Field	Description	Values /Constraints	Example
Author(s)*	full name	1000 character limit	Jerome Giraudat

Submitter's Community ID*	unique identifier for a community member in TAIR. You must be registered to have a community ID	TAIR accession from TAIR Community Search	community:4851
Title	Title of the protocol	500 character limit	Rapid RNA preparation
Description	brief text description	1000 character limit	A method for rapidly preparing RNA from leaf tissue.
Publication/ Citation/ Reference_id	For a reference that describe the protocol, this is the unique identifier in TAIR. For articles referring to the markers that are not yet in TAIR, please contact a curator to update this information into the database when the paper is published.	TAIR accession from TAIR Publication Search	publication: 501682410
Usage keywords*	One or more keywords that describe the application(s) of the method (e.g. gene expression assay, protein-protein interaction assay).	separate multiple keywords with semicolons	gene expression assay
Method keywords*	use as many keywords as necessary to indicate the methods described in the protocol (e.g. mRNA isolation, cell fractionation).	separate multiple keywords with semicolons	Northern gel blot; RNA extraction