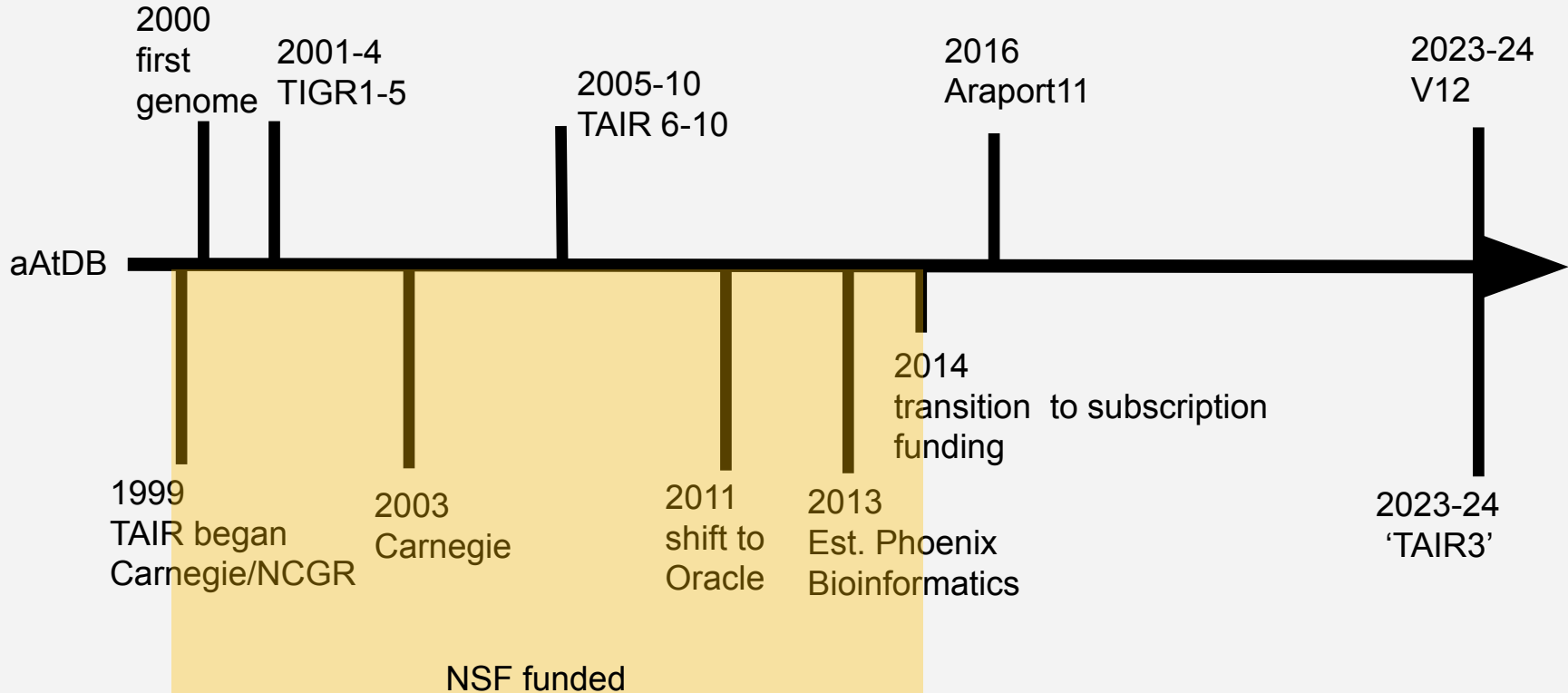


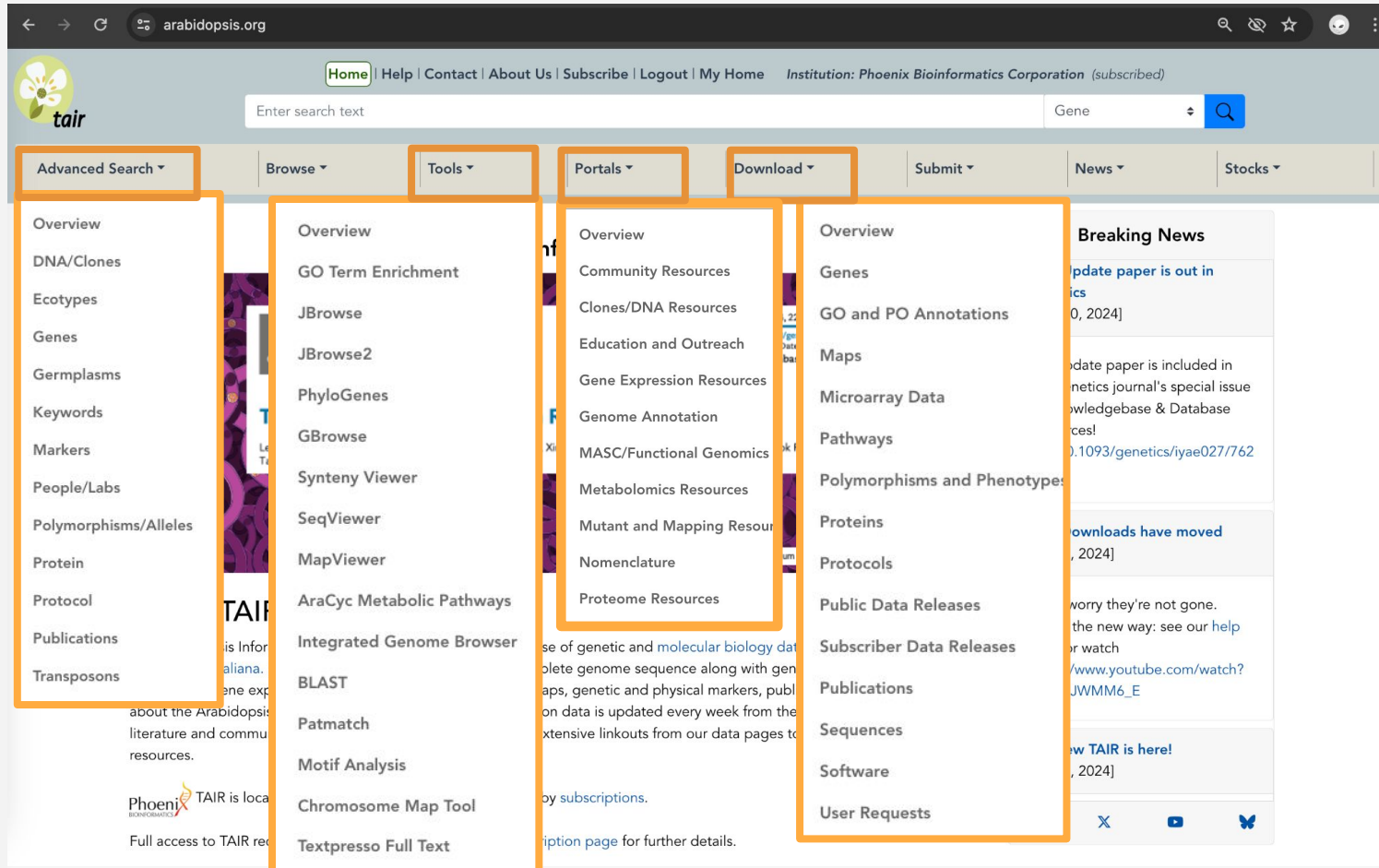
# TAIR in 2024: The Silver Anniversary Makeover Edition

GLOW UP  
  
*tair*

Leonore Reiser, Curator  
ASPB June 22, 2024  
[www.arabidopsis.org](http://www.arabidopsis.org)

# 25 years of TAIR





The screenshot shows the TAIR website homepage with several navigation menus highlighted by orange boxes. The main navigation bar includes: Home | Help | Contact | About Us | Subscribe | Logout | My Home | Institution: Phoenix Bioinformatics Corporation (subscribed). Below this is a search bar with the text "Enter search text" and a "Gene" dropdown menu.

The highlighted navigation menus are:

- Advanced Search**
  - Overview
  - DNA/Clones
  - Ecotypes
  - Genes
  - Germplasms
  - Keywords
  - Markers
  - People/Labs
  - Polymorphisms/Alleles
  - Protein
  - Protocol
  - Publications
  - Transposons
- Tools**
  - Overview
  - GO Term Enrichment
  - JBrowse
  - JBrowse2
  - PhyloGenes
  - GBrowse
  - Synteny Viewer
  - SeqViewer
  - MapViewer
  - AraCyc Metabolic Pathways
  - Integrated Genome Browser
  - BLAST
  - Patmatch
  - Motif Analysis
  - Chromosome Map Tool
  - Textpresso Full Text
- Portals**
  - Overview
  - Community Resources
  - Clones/DNA Resources
  - Education and Outreach
  - Gene Expression Resources
  - Genome Annotation
  - MASC/Functional Genomics
  - Metabolomics Resources
  - Mutant and Mapping Resources
  - Nomenclature
  - Proteome Resources
- Download**
  - Overview
  - Genes
  - GO and PO Annotations
  - Maps
  - Microarray Data
  - Pathways
  - Polymorphisms and Phenotypes
  - Proteins
  - Protocols
  - Public Data Releases
  - Subscriber Data Releases
  - Publications
  - Sequences
  - Software
  - User Requests

The right side of the page features a "Breaking News" section with several news items, including "Update paper is out in..." and "Downloads have moved...".

# Evolution of TAIR infrastructure



Datastores

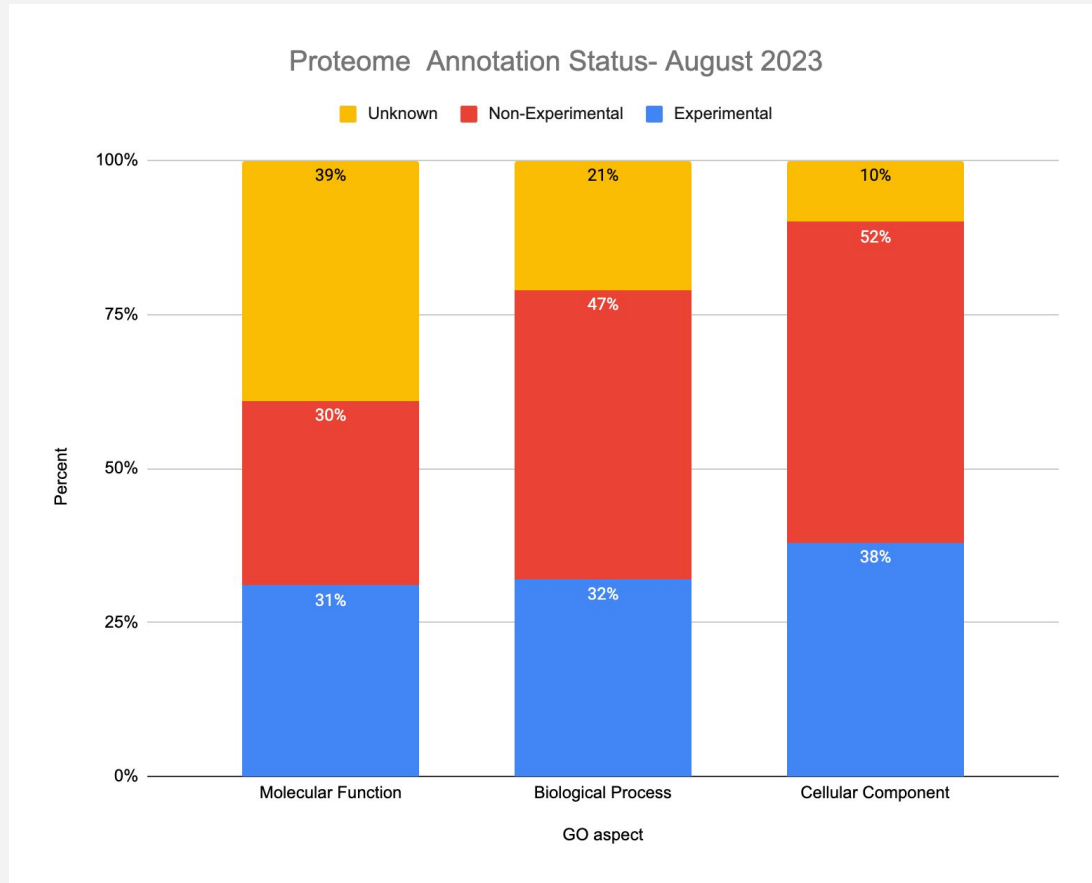


More accessible content

# Continued focus on manual curation of gene function

- Manual validation of gene –paper association
- Gene symbols
- Manually curated gene descriptions/summaries
- New and existing alleles
- Phenotypes
- Functional annotation using Gene Ontologies
- Updated weekly

# Reference Genome- GO annotation status



[DOI:10.1093/genetics/iyae027/7624473](https://doi.org/10.1093/genetics/iyae027/7624473)

# Find information for individual genes

Home | Help | Contact | About Us | Subscribe | Logout | My Home Institution: Phoenix Bioinformatics Corporation (subscribed)

tair NPH1

Advanced Search ▾ Browse ▾ Tools ▾ Portals ▾ Download ▾ Submit ▾ News ▾ Stocks ▾

Home / Search / Genes

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

Genes may be searched by name, keywords, features, and/or location. In TAIR, a Gene Model is defined as any description of a gene product from a variety of sources including computational prediction, mRNA sequencing, or genetic characterization. A locus is defined as the genomic sequence corresponding to a transcribed unit (e.g. AT2G03340) in the genome. In TAIR, many gene models can exist for a given locus, therefore a search for a gene may result in multiple hits for the same gene name.

Reset Submit Query

Search by Name or Phenotype

Search Text contains Gene Name

Include obsoleted genes

Search by Gene List

Upload or type in a list of AGI locus IDs (separated by line return)

Upload

Search by Associated Keyword

Keyword Term

Keyword Type Any GO Molecular Function

Name (AGI/Symbol)

https://www.google.com/search?q=AT3G45780&client=firefox-b-1-e&sca\_esv=10eeeb1831

AT3G45780

All Shopping Images Videos Forums More Tools

TAIR <https://www.arabidopsis.org/servlets/TairObject> na...

**Locus: AT3G45780**

Oct 4, 2023 — Blue-light photoreceptor. Contains a light activated serine-threonine kinase domain and LOV1 and LOV2 repeats. Mutants are defective in ...

People also ask

What is the structure of phot1? ▾

What is the structure and function of Phototropin? ▾

Feedback

TAIR <https://www.arabidopsis.org/servlets/TairObject> na...

**Gene: AT3G45780.2**

Nov 2, 2018 — Blue-light photoreceptor. Contains a light activated serine-threonine kinase domain and LOV1 and LOV2 repeats. Mutants are defective in ...

Advanced Search ▾
Browse ▾
Tools ▾
Portals ▾
Download ▾
Submit ▾
News ▾
Stocks ▾

Summary

Transcripts

Maps and Mapping Data

Sequences

Protein Data

Expression

Gene Ontology

Homology

Germplasm and Clones

Polymorphisms

Publications

External Links

## Locus: AT3G45780

### Publications

246 entries

View Complete List (15 of 246 displayed)

Author/Title ↕	Source ↕	Associated Loci ↕	Date ↕
<a href="#">Yoshida, T., Mergner, J., Yang, Z. et al.</a> Integrating multi-omics data reveals energy and stress signaling activated by abscisic acid in Arabidopsis	THE PLANT JOURNAL	<a href="#">AT1G01140</a> ; <a href="#">AT1G17550</a> ; <a href="#">AT1G73370</a> ; <a href="#">AT2G36460</a> ; <a href="#">AT2G38310</a> ; <a href="#">AT3G01090</a> ; <a href="#">AT3G29160</a> ; <a href="#">AT3G45780</a> ; <a href="#">AT3G47520</a> ; <a href="#">AT3G50500</a> ; <a href="#">AT3G52720</a> ; <a href="#">AT3G52930</a> ; <a href="#">AT3G59480</a> ; <a href="#">AT4G10120</a> ; <a href="#">AT4G26080</a> ; <a href="#">AT4G26530</a> ; <a href="#">AT4G38970</a> ; <a href="#">AT5G03690</a> ; <a href="#">AT5G20280</a> ; <a href="#">AT5G39440</a> ; ... more	2024
<a href="#">Tseng, T. S., Chen, C. A., Lo, M. H. et al.</a> PHOTOTROPIN1 lysine 526 functions to enhance phototropism in Arabidopsis	PLANTA	<a href="#">AT3G45780</a>	2024



Home | Help | Contact | About Us | Subscribe | Logout | My Home | Institution: Phoenix Bioinformatics Corporation (subscribed)

tair Enter search text Gene

Advanced Search \* Browse \* Tools \* Portals \* Download \* Submit \* News \* Stocks \*

Home / Search / Genes

**TAIR Gene Search** [Help]

Genes may be searched by name, keywords, features, and/or location. In TAIR, a Gene Model is defined as any description of a gene product from a variety of sources including computational prediction, mRNA sequencing, or genetic characterization. A locus is defined as the genomic sequence corresponding to a transcribed unit (e.g. AT2G03340) in the genome. In TAIR, many gene models can exist for a given locus, therefore a search for a gene may result in multiple hits for the same gene name.

Reset Submit Query

Search by Name or Phenotype

Search Text contains Gene Name

Include obsolete genes

Search by Gene List

Upload or type in a list of AGI locus IDs (separated by line return)

- AT4G30670
- AT1G30814
- AT2G31240
- AT5G44860
- AT5G11780
- AT1G19240
- AT4G25920
- AT5G55650
- AT1G23201
- AT5G15008
- AT3G42250
- AT3G21990
- AT1G16500
- AT2G35480
- AT5G45530

Search by Associated Keyword

Keyword Term

Keyword Type

- Any
- GO Molecular Function
- GO Biological Process
- GO Cellular Component

Evidence

- inferred from experiment (EXP)
- inferred from electronic annotation (IEA)
- inferred from expression pattern (IEP)
- inferred from genetic interaction (IGI)

Restrict by Features

Gene Model Type

- Any
- pre tms
- transposable element gene
- protein coding

Advanced

- has associated literature
- genetic (unsequenced) loci only

Restrict by Map Locations

Chromosome

Map Type

Range and kbp

Reset Submit Query

Name (basic) plus ID, description, phenotype (advanced)

**New Bulk Query and Download**  
(Upload list of AGI locus IDs)

Search by GO or PO annotations, gene model type

Location search

## TAIR Search Results

New Search Download All Download Checked

Get GO Annotations Get PO Annotations Get Sequences **Get Gene Descriptions** Get Locus History Get Microarray Elements

Your query for genes where taxon name is Arabidopsis thaliana,, keyword type of Any, gene model type is Any, evidence are Any, resulted in 6020 loci matches  
From the uploaded 6020 gene list, 0 were filtered out

Displaying 1 - 25 of 6020 results

Select All Clear Selected

No.	Locus	Description
1	<input checked="" type="checkbox"/> AT3G01311	Other Names: actin cross-linking protein, putative (DUF569);(source:Araport11)
2	<input checked="" type="checkbox"/> AT4G11211	Other Names: Hypothetical protein;(source:Araport11)
3	<input checked="" type="checkbox"/> AT1G30814	Other Names: hypothetical protein;(source:Araport11)
4	<input checked="" type="checkbox"/> AT4G26490	Other Names: Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family;(source:Araport11)
5	<input checked="" type="checkbox"/> AT4G30770	Other Names: Putative membrane lipoprotein;(source:Araport11)

Bulk Download Options

Make sure to select all

« < 1 2 3 4 ... > »



tair



## Gene Descriptions

Download All Download Checked

Download file

Displaying 1 - 25 of 6020 results

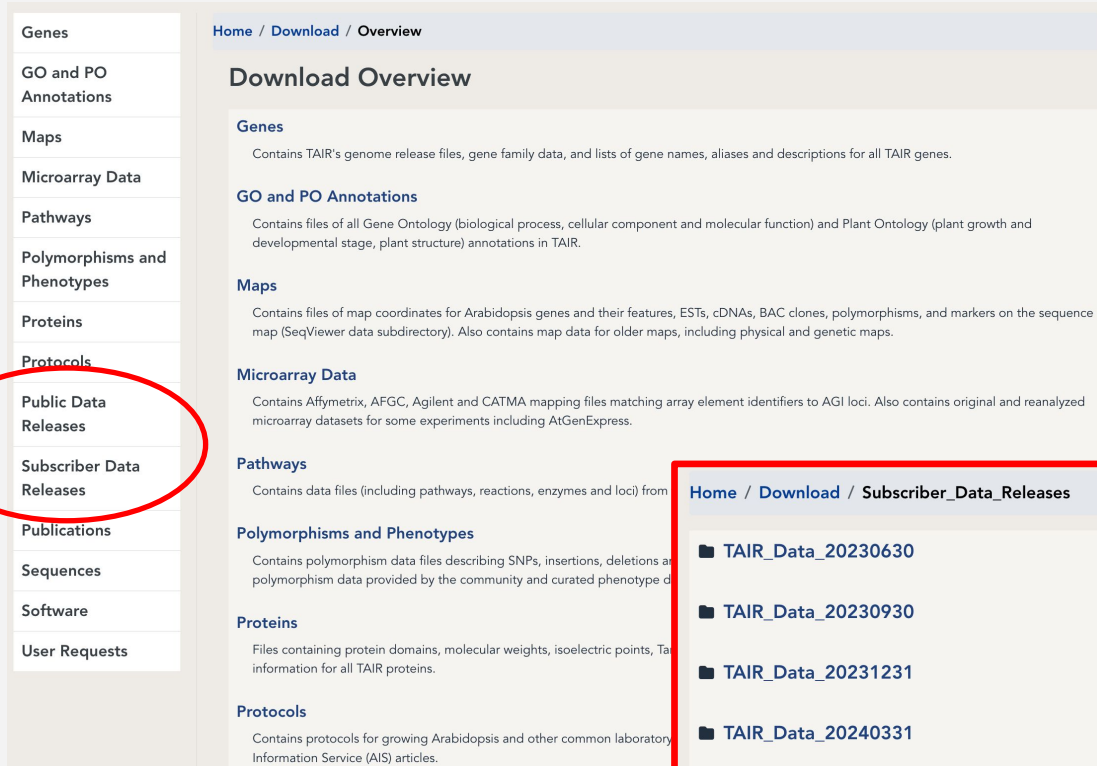
Select All Clear Selected

No.	Locus Identifier	Representative Gene Model Name	Gene Description	Gene Model Type	Primary Gene Symbol	All Gene Symbols
1	<input checked="" type="checkbox"/> AT3G01311	AT3G01311.1	actin cross-linking protein, putative (DUF569);(source:Araport11)	protein_coding		
2	<input checked="" type="checkbox"/> AT4G11211	AT4G11211.1	hypothetical protein;(source:Araport11)	protein_coding		
3	<input checked="" type="checkbox"/> AT1G30814	AT1G30814.1	hypothetical protein;(source:Araport11)	protein_coding		
4	<input checked="" type="checkbox"/> AT4G26490	AT4G26490.1	Late embryogenesis abundant (LEA) hydroxyproline-rich	protein_coding		
5	<input checked="" type="checkbox"/> AT4G30814	AT4G30814.1	hypothetical protein;(source:Araport11)	protein_coding		
6	<input checked="" type="checkbox"/> AT4G30814	AT4G30814.1	hypothetical protein;(source:Araport11)	protein_coding		
7	<input checked="" type="checkbox"/> AT5G20470	AT5G20470.1	Encodes a headless derivative of myosin XI-K, which likely arose from a partial duplication of the XI-K gene and is developmentally regulated.	protein_coding	HEADLESS DERIVATIVE OF MYOSIN XI-K (HDK)	HEADLESS DERIVATIVE OF MYOSIN XI-K (HDK)
8	<input checked="" type="checkbox"/> AT5G04000	AT5G04000.2	hypothetical protein;(source:Araport11)	protein_coding		

Make sure to select all

« < 1 2 3 4 ... > »

# Access and Reuse: Whole genome datasets



Home / Download / Overview

## Download Overview

**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**  
Contains data files (including pathways, reactions, enzymes and loci) from

**Polymorphisms and Phenotypes**  
Contains polymorphism data files describing SNPs, insertions, deletions and polymorphism data provided by the community and curated phenotype d

**Proteins**  
Files containing protein domains, molecular weights, isoelectric points, Ta

**Protocols**  
Contains protocols for growing Arabidopsis and other common laboratory

Information Service (AIS) articles.

Home / Download / Subscriber\_Data\_Releases

- TAIR\_Data\_20230630
- TAIR\_Data\_20230930
- TAIR\_Data\_20231231
- TAIR\_Data\_20240331
- .DS\_Store 6 KB Mar 9, 2024
- README.txt 10 KB Mar 9, 2024

Data and metadata

Standard, interoperable data formats if they exists

Access under CC-BY 4.0

Quarterly data releases (public releases are a year behind except GO annotations)

Archived data

# FAIRifying community data

Advanced Search ▾ | Browse ▾ | Tools ▾ | Portals ▾ | Download ▾ | Submit ▾ | News ▾ | Stocks ▾

[Home](#) / [Data Submission](#)

## Data Submission Overview

We encourage users of TAIR to share their data with the research community.

Please see below Data Types Accepted by TAIR and Other Repositories (for data types not accepted by TAIR).

**DATA TYPES ACCEPTED BY TAIR**

Please take a look at the specific guidelines below for submitting data to TAIR. You will be asked to download an Excel worksheet to fill in and submit to TAIR. If you have any questions or suggestions, please contact us at: [curator@arabidopsis.org](mailto:curator@arabidopsis.org)

**External Links - Guidelines - Submission Form**  
We provide links from TAIR detail pages to other stable and generally useful data resources.

**Gene Class Symbol Registration - Guidelines - Online Form**  
Reserve a gene symbol prior to publication to ensure uniqueness and consistency in gene nomenclature.

**Gene Function Submission Forms for Authors and Others - Online Form - Large Dataset Submission Form**  
Please submit your summary level data regarding all aspects of gene function, including gene product function, biological process, subcellular localization, expression pattern, and interactions with other gene products. See Other Repositories below for submission of quantitative experimental results.

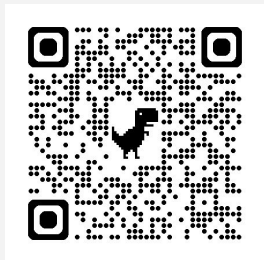
**Gene Structure Additions/Modifications - Guidelines - Submission Form**  
If you have found missing information, errors or discrepancies in the structure of an existing gene in TAIR, we would like to update our gene information to include your data.

**JBrowse tracks**  
We accept a variety of data in GFF3 format that can be displayed as JBrowse tracks. We can also provide a private space for peer review. Please contact TAIR curators about your specific data types.

**Marker and Polymorphism Data - Guidelines - Submission Form**  
Submit data on sequenced alleles and markers (including CAPS, SSLP, AFLP, RFLP and RAPD) for inclusion in TAIR.

**Phenotypes - Guidelines - Submission Form**  
Submit phenotype descriptions for plants that you have characterized, whether they are your own stocks or were obtained from a stock center like ABRC.

**Protocols - Guidelines**  
Make your protocol available through the TAIR Protocol Search.



Phoenix Bioinformatics Search

Arabidopsis Community

## Tips for making your published data about Arabidopsis genes more FAIR (Findable, Accessible, Interoperable and Reusable)

Arabidopsis Community

A. thaliana Col-0 v12 reannotation...

About Us

Data Submission

News

Portals

Resources

\*Most Wanted\* Arabidopsis Ge...

1001 EpiGenomes Browser

1001 Genomes Project

AgBioData Consortium

Arabidopsis Gene Nomenclature

Arabidopsis Next-Gen Sequenc...

Arabidopsis Single Cell Develop...

AraENCODE

AraGWAS

AraMemnon

AraPheno

Araport

Array Express

ARS-Arabidopsis RNA-seq Data...

ARVEX-Arabidopsis Root Virtua...

ATHENA (Arabidopsis THaliana ...

Atlas of Fe containing proteins

At-phasRNA

Bio-Analytic Resource for Plant ...

To ensure that your Arabidopsis gene information is accessible here are some helpful pre-publication guidelines for authors. These guidelines here are written to be specific for Arabidopsis genes, but are based on a general set of guidelines and principles (see references). This makes it easier for TAIR curators to curate your published data (which will increase its visibility) and for other researchers to find and reuse your data.

**Include AGI Locus Identifiers for genes**

To ensure that the genes described in your paper are unambiguously identified, include the systematic locus identifier for that locus. If you have identified a new gene that does not yet have an AGI locus ID, please contact TAIR curators PRIOR to publishing your gene. TAIR, and other resources, use text mining to associate publications to biological entities in databases (e.g. genes and proteins). In the absence of a unique identifier such as an AGI locus code or UniProt ID, text mining software cannot distinguish between CCRI1 and CCRI. To ensure that your published data can be curated and accurately linked to a database record, use the AGI locus ID.

**Do not reuse gene symbols**

To avoid the problem of different genes being referred to by the same symbolic name, before you publish check to see if the name is in use. Check the gene symbol registry at TAIR, and search PubMed, PubMed, Google Scholar to see if that symbolic name is in use for another Arabidopsis gene. If the name is not in use, please register the gene symbol.

**Follow gene, protein and allele nomenclature standards**

There is an established nomenclature for genes, proteins and alleles for *Arabidopsis thaliana*. For example, alleles are lowercase and distinguished by a dash and number (abc2-1, abc2-2). Again, check and make sure that the name is not already in use.

**Naming and re-naming T-DNA insertion lines**

There is a difference between a stock/germplasm and the specific T-DNA insertion that is causal for a phenotype. Therefore when referencing an allele please include the specific polymorphism and not just the name of the ABRC/NASC stock because many ABRC/NASC T-DNA stocks contain multiple insertions. Before naming your allele, check to see if it already has a name in TAIR or in the literature. At TAIR, we will update the allele name with the newly published name. If you have an allele/polymorphism/phenotype that does not already exist in TAIR, please send us the information.

**Use standard formats for reporting data (if available)**

To ensure that data is reusable and interoperable (e.g. can be integrated across platforms/tools) community based initiatives have been established to define data standards. For example the Minimal Information about any Sequence MixS standards or Minimal Information about a Plant Phenotyping Experiment (MIAPPE) standards. If you are not sure that a standard exists, check the FAIR Sharing website or contact a curator.

# Genome Browsers (phasing out old)

## JBrowse

## JBrowse2

Feature	JBrowse	JBrowse2
Views	Linear	Linear, circular and dot plot
Preloaded tracks including community data	✓	✓
Configurable tracks	✓	✓ (easier)
Load and view user tracks (e.g. GFF/BED files)	✓	✓ (inc. data hubs)
View multiple chromosomes at same time		✓
View multiple genome assemblies		✓
Sequence/regex search		✓
Synteny view		✓
View /export annotated gene sequence	✓ in Seq Lighter	✓ in Feature View

Other features and configurations possible but not yet implemented



# V12 Col-0 Reference genome and annotation updates

conf.phoenixbioinformatics.org/display/COM/A.+thaliana+Col-0+v12+reannotation+effort

Arabidopsis Community

A. thaliana Col-0 v12 reannotation effort

This site aims to collate information about the A. thaliana Col-0 v12 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](#)

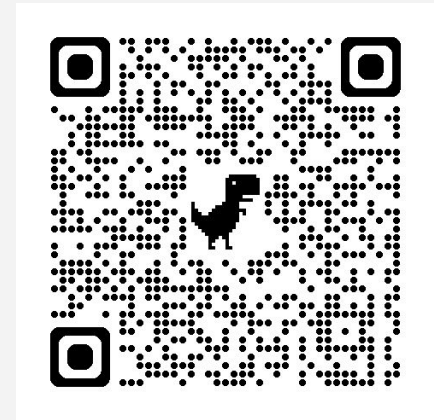
Do you want to get involved? Comments? Suggestions? Please email [curator@arabidopsis.org](mailto:curator@arabidopsis.org).

**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

Phase	Start	End	Key Events
Genome	Nov 2022	Jan 2023	S., Schneeb...
Automate	Jan 2023	Apr 2023	NCBI to re..., NCBI to run anno..., N.
Manual Review	Mar 2023	Jul 2023	TAIR to set up Apollo, TAIR to set u..., Community to conduct manual review
NCBI/Ref	Jul 2023	Oct 2023	TAIR to tes..., TAIR to ru..., TAIR to su...
Dissem	Jan 2023	Dec 2023	T., Poster and..., Write manuscript about community effort



[tinyurl.com/Athalianav12](https://tinyurl.com/Athalianav12)



- Col- CC consensus assembly submitted (virtually error free)  
-Schneeberger lab doi:10.1126/science.abi7489 (v1)
- NCBI ran eukaryotic annotation pipeline
- Grafting of chr 2 and 4 NORs from Pikard lab doi:10.1126/sciadv.adj4509 (v2)

Assembly

Automated  
Annotation

Manual Review

GenBank  
Submission

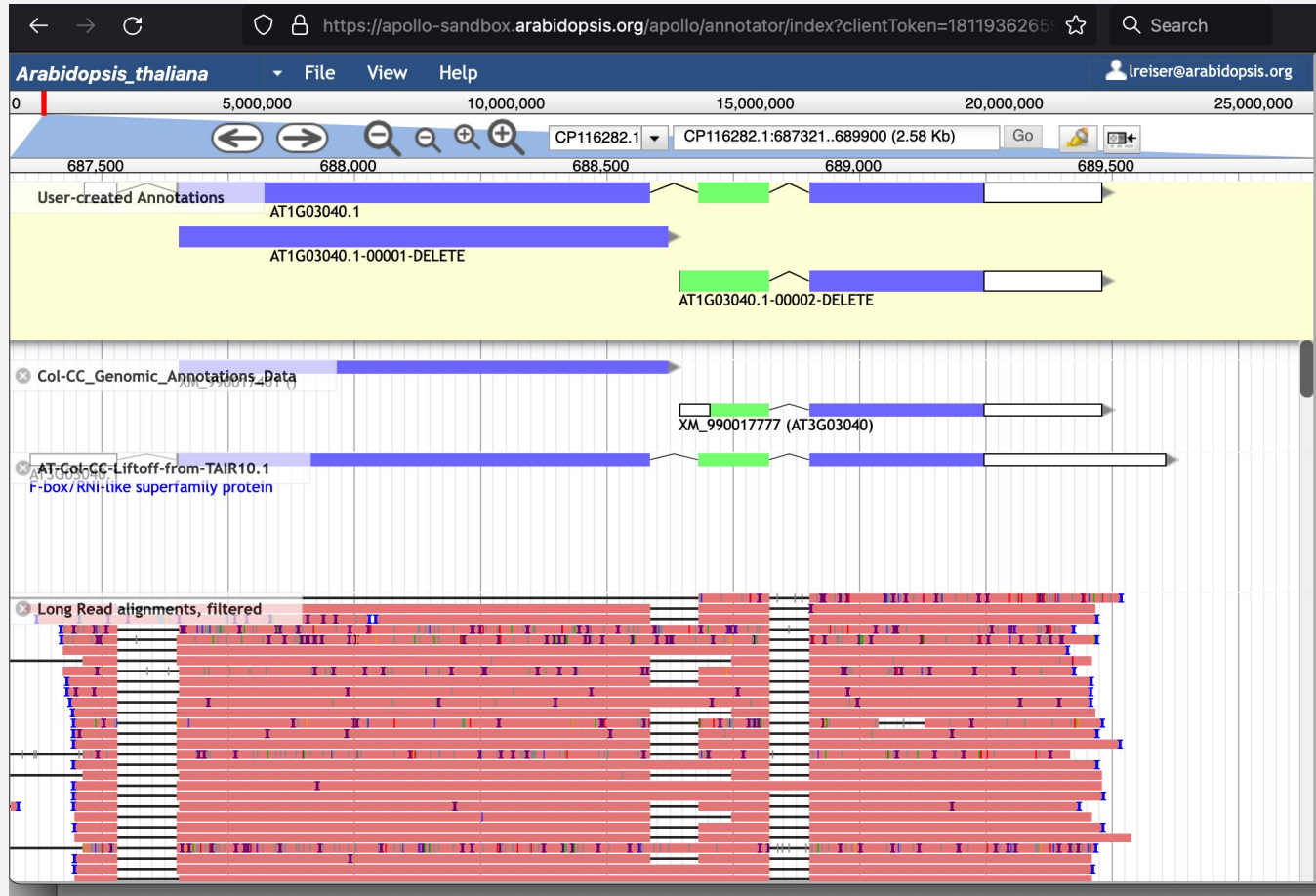
Dissemination/  
Integration

- Review of the automated annotation pipeline results
- Annotation of specific types of genome elements
  - Transposable elements
  - Tandem repeat elements
  - 5SrRNAs
  - Long non-coding RNAs
  - 45SrRNAs (NOR2 and NOR4)
  - Protein coding genes

So many amazing volunteers



# Manual Review Example



<https://apollo-sandbox.arabidopsis.org/apollo/annotator/index?clientToken=1811936265>

**Arabidopsis\_thaliana** File View Help treiser@arabidopsis.org

0 5,000,000 10,000,000 15,000,000 20,000,000 25,000,000

CP116282.1 CP116282.1:687321..689900 (2.58 Kb) Go

687,500 688,000 688,500 689,000 689,500

**User-created Annotations**

- AT1G03040.1
- AT1G03040.1-00001-DELETE
- AT1G03040.1-00002-DELETE

**Col-CC\_Genomic\_Annotations\_Data**

- XM\_990017777 (AT3G03040)

**AT-Col-CC-Liftoff-from-TAIR10.1**  
 F-box/KNI-like superfamily protein

**Long Read alignments, filtered**

# TAIR annotation tasks

- Review inserts, deletions, splits and merges*
- Secondary reviews*
- Resolve gene overlaps (lots of hyperextended UTRs)*
- Resolve internal stops, improperly assigned start or stop codons
- Comparisons/calculations
- Integrate annotations from different groups into a single GFF w other tracks
- Final QC
- Submit to GenBank

# The TAIR team (curator@arabidopsis.org)



Tanya Berardini  
TAIR Director



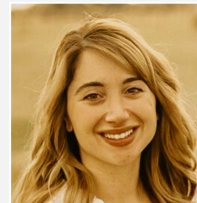
Leonore  
Reiser



Erica Bakker



Shabari  
Subramaniam



Alyssa Proia

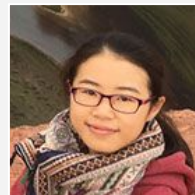
And all of our  
contributors  
and supporters  
since 1999!



Trilok Prithvi



Swapnil Sawant



Xingguo Chen

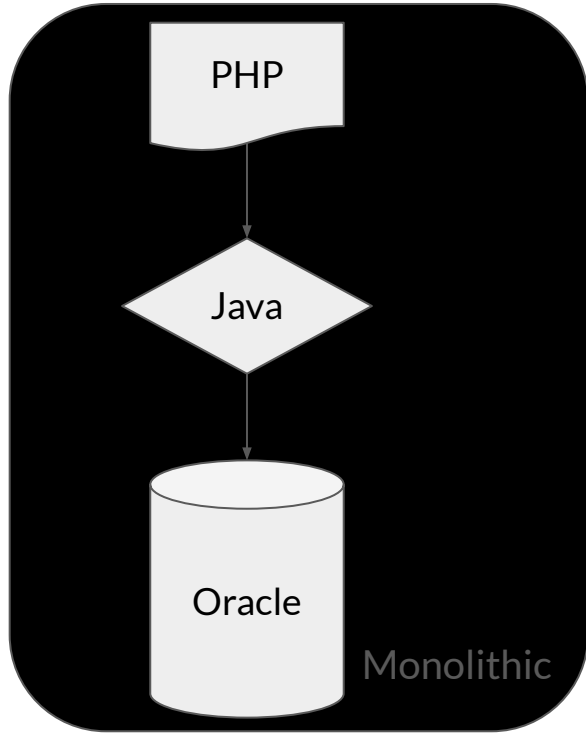


Kartik Khosa

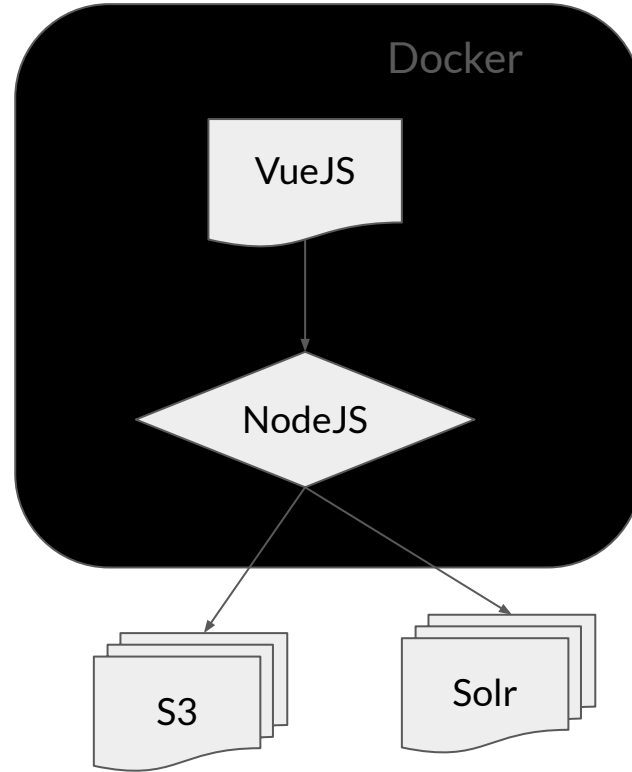


**Come see us at Booth #406**  
**Read the Genetics update paper**

# Architecture



Old Architecture



New Architecture



# Outline



- Very quick introduction to TAIR
- New and updated website and tools
- Version 12 genome annotation status updates