

BAR ThaleMine

Asher Pasha¹ Nicholas Provart¹

¹Bio-Analytic Resource (BAR)
Centre for the Analysis of Genome Evolution and Function
University of Toronto
Toronto, Canada

January 30, 2020

BAR: Bio-Analytic Resource

About us:

- ▶ One of the largest Plant Biology Resources
- ▶ Paper published in 2005¹
- ▶ Based in University of Toronto, Canada
- ▶ Free to access: <http://bar.utoronto.ca>

Resources Include:

- ▶ Web applications, web services, databases
- ▶ Gene Expression data ([eFP](#), [ePlant](#))
- ▶ Protein-Protein, Protein-DNA Interactions ([AIV2](#))
- ▶ Protein structures ([ePlant](#))

¹Toufighi et al. 2005.

Araport: Arabidopsis Information Portal

A centralized resource for Arabidopsis data²

- ▶ Started in 2013 under the leadership of Prof. Chris Town (JCVI)
- ▶ JCVI, TACC, University of Cambridge
- ▶ Drupal 7 based araport.org page
- ▶ ThaleMine (InterMine instance)
- ▶ App store on araport.org
- ▶ [JBrowse](#)
- ▶ Araport11 re-annotation project
- ▶ Single Sign on to all Araport resources
- ▶ RESTful API for scientific data
- ▶ Community outreach (workshops, conferences, etc)

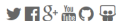
²Krishnakumar et al. 2015.

Araport: Arabidopsis Information Portal

BAR contribution to Araport

- ▶ Provided gene expression databases used by BAR eFP Arabidopsis Browser
- ▶ Attended Araport workshop in November 2014
- ▶ Interactions Viewer - the first community app published on araport.org
- ▶ Contributed several apps and web services over the next few years
- ▶ Biggest contribution: BAR ePlant app

Araport: Arabidopsis Information Portal



ARAPORT
ARABIDOPSIS INFORMATION PORTAL

[About ▾](#)[Data ▾](#)[Tools ▾](#)[Community ▾](#)[Developer ▾](#)[Help ▾](#)[Log in ▾](#)[ALL](#)[THALEMINE](#)[JBROWSE](#)

Search ThaleMine, JBrowse and other site content

☒ Search by identifiers or keywords e.g. AT4G19020 or epigenetic

New to Araport?

Araport is a one-stop-shop for *Arabidopsis thaliana* genomics. Araport offers gene and protein reports with orthology, expression, interactions and the latest annotation, plus analysis tools, community apps, and web services. Araport is 100% free and open-source. Registered members can save their analysis, publish science apps, and post announcements. [READ MORE](#)

[REGISTER](#)[LOGIN](#)

THALEMINE



Click to analyze gene lists or drill down to gene reports.

JBROWSE



Click to browse gene structures and other annotation.

APP STORE



Click to build your own analysis tools and share your data.

BAR ePlant



Visual analytic tool for exploring multiple levels of Arabidopsis data.

Araport Highlights



Official Release Jun 2016

529092123/<https://www.araport.org> | [Re-annotation](#)



Upcoming Events

ICAR, Jun 19-23

News and Job Postings

★ ABRC Funding Update

2016-

04-25

Conferences and Meetings

Frontiers and Techniques in

Plant Science

2017-

06-30

Araport: Arabidopsis Information Portal

Funding status

- ▶ Lost all funding in 2016
- ▶ Lost almost all staff
- ▶ Application for new funding failed

Future of Araport team

- ▶ JCVI, TACC, BAR, TAIR, NCGR
- ▶ First meeting at JCVI in March, 2019
- ▶ BAR got ThaleMine, TAIR got JBrowse, NCGR contributed GCV
- ▶ Everything else was deprecated
- ▶ News update: araport.org

Araport: Arabidopsis Information Portal



Welcome to Araport.org. Funding of project has been discontinued so teams from The Arabidopsis Information Resource (TAIR), the National Center for Genome Resources (NCGR), and the Bio-Analytic Resource for Plant Biology (BAR) have taken over its operation and have refreshed/expanded the functionalities that were available at Araport as in the table below. Legacy Araport data are still available at the following links, running on TACC infrastructure.

	Legacy URL	New URL
ThaleMine	Araport ThaleMine	BAR ThaleMine
JBrowse	Araport JBrowse	TAIR JBrowse
GCV		NCGR GCV



Araport ThaleMine to BAR ThaleMine

Araport ThaleMine: A few minor problems

- ▶ Not maintained since 2017
- ▶ Based on InterMine 1.8.5
- ▶ Used modified data
- ▶ Used several custom loader and displayers
- ▶ Some InterMine core and biosources changes where never made upstream
- ▶ Too difficult to upgrade to InterMine 3.1.2

BAR ThaleMine: Principles

No initial funding. So we follow these rules³

Simplicity Keep code to minimum possible lines and contribute as many changes to InterMine (upstream) as possible.

Least Surprise Test and upgrade every minor release (Principle of least astonishment)

Clarity No modification to data

Extensibility Build for the near future (Java 11, etc)

Parsimony Write loaders and displayers only when we have to.

³Raymond 2003.

BAR ThaleMine

- ▶ We started from scratch (using `make_mine` script)
- ▶ Progress based on HumanMine, FlyMine, and Araport ThaleMine
- ▶ Got most of the loaders working
- ▶ Pushed changed to upstream InterMine
- ▶ So far: ThaleMine v4.1.2-20200127
- ▶ Also on GitHub [ThaleMine](#) and [ThaleMine BioSources](#)

ThaleMine: What's Working



ThaleMine v4.1.2-20200127 Data mining on *Arabidopsis thaliana*

Home Templates Lists QueryBuilder Regions Data Sources API MyMine

Contact Us | Log in

Search: e.g. AT1G01640

GO



Search

Search ThaleMine. Enter **names, identifiers or keywords** for genes, proteins, ontology terms, authors, etc. (e.g. *FT*, *APL_ARATH*, lateral root development, Somerville).

e.g. AT3G24650, FT, APL_ARATH

SEARCH



Analyse

Enter a **list of identifiers**.

Gene

e.g. AT1G02850, AT1G05240,
AT1G05250, AT1G05260, AT1G14540,
AT1G14550, AT1G15950, AT1G24110,

[advanced](#)

ANALYSE

Welcome Back!

ThaleMine enables you to analyze *Arabidopsis thaliana* **genes, proteins, gene expression, protein-protein interactions, orthologs**, and more.

Use plain text or structured queries for interactive gene and protein reports.

TAKE A TOUR (VIA FLYMINE)

GENES AND PROTEINS

HOMOLOGY

FUNCTION

INTERACTIONS

The gene models and other genome annotations in ThaleMine are provided by a variety of sources including: NCBI, TAIR [Read more](#)

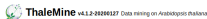
Query for genes and proteins:

- Gene ➡ Protein sequence
- Gene ➡ CDS sequence
- Gene ➡ Transposable Elements

» [More queries](#)

popular templates

ThaleMine: What's Working



Info | Help | Software |

[Home](#) [Templates](#) [Links](#) [QueryBuilder](#) [Registers](#) [Data Sources](#) [API](#) [MyMine](#)

[Contact Us](#) | [Log In](#)

Search:

Data

ThaleMine integrates data from a large number of sources into a single data warehouse. This page lists the data that are included in the current release and it is manually curated. Its contents are not indexed in our keyword search. More data sets will be added in future releases, please contact us if there are any particular data you would like to see included.

Data Category	Data	Source	PubMed
Genome	TAIR10 Genome assembly (5 chromosomes plus chloroplast and mitochondrial assemblies)	NCBI * - Release TAIR10 (2018.04.06)	Arabidopsis Genome Initiative - PubMed: 11130711 *
	Araport11 OPPID data from TAIR	TAIR * - Release Araport11 (2018.06.07)	Cheng et al., 2018 - PubMed: 27822469 *
Proteins	High quality, manually annotated, non-redundant protein sequence database.	Swiss-Prot * - Release 2019_11	UniProt Consortium - PubMed: 17142230 *
	Computationally analysed records, enriched with automatic annotation	TrEMBL * - Release 2019_11	
	Protein family and domain assignments to proteins	InterPro * - Release v76.0	Mitchell et al., 2019 - PubMed: 30386656 *
Homology	Orthologue and paralogous relationships based on the inferred speciation and gene duplication events in the phylogenetic tree.	Paranor * - Release 14.1	Mi et al. - PubMed: 22100209 *
	Phylozone Homologs generated with InParanoid	Phylozone * - realtime	Goodstein et al. - PubMed: 22110026 *
Gene Ontology	GO annotations from Gene Ontology	Gene Ontology * - Release 2020-01-01	Berardini et al., 2004 - PubMed: 15173966 * Gene Ontology Consortium - PubMed: 10802051 *
	Several electronic and manual GO annotation methods utilized by UniProt	UniProt * - Release 2019_11	UniProt Consortium - PubMed: 17142230 *
Interactions	Curated set of genetic and physical interactions for Arabidopsis thaliana	BioGRID * - Release 3.5.180	Chapman et al., 2014 - PubMed: 25426363
	Curated binary and complex protein-protein interactions for Arabidopsis thaliana	PRAC * - Downloaded 20200121	Kerren et al., 2012 - PubMed: 22121220
Expression	Electronic Fluorescent Oligonucleotide (eFP) Visualization paints gene expression information from one of the AGenExpress data sets or other compendia for a desired gene onto a diagrammatic representation of Arabidopsis thaliana plants.	BAR eFP Webserver * - realtime	Weiler et al., 2007 - PubMed: 17184562 * Brady et al., 2009 - PubMed: 19401381 *
Co-Expression	Co-regulated gene relationships deduced from microarray and RNA-seq data via ATTED-II web services	ATTED-II Co-expression * - realtime	Okuyadi et al., 2014 - PubMed: 24334350 *
Publications	Curated associations between publications and genes from UniProt	UniProt * - Release 2019_11	UniProt Consortium - PubMed: 17142230 *
	Publications from InterPro	InterPro * - Release v77.0	Mitchell et al., 2019 - PubMed: 30386656 *
GeneRIF	Publications from NCBI	NCBI * - Downloaded 20200121	Maguire et al., 2007 - PubMed: 17146475 *
	Concise phrase describing gene function and publication associated with NCBI Gene records	NCBI * - Downloaded 20200121	Maguire et al., 2007 - PubMed: 17146475 *

BAR ThaleMine: Some Future Plans and Known Issues

The follow is planned

- ▶ Arabidopsis Strains⁴
- ▶ Gene expression data (Araport11 RNA-Seq⁵)
- ▶ Salk data (TDNA-Seq)
- ▶ Gene histories from JCVI
- ▶ TAIR annotations

Known Issues

- ▶ [web properties](#) from ThaleMine web services
- ▶ BlueGenes frontend

⁴Jiao and Schneeberger 2019.

⁵Cheng et al. 2017.

Acknowledgements