BAR ThaleMine

Asher Pasha[1] Nicholas Provart[1]

[1]Bio-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 2005[1]
- 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)

[1] Toufighi et al. 2005.

Araport: Arabidopsis Information Portal

A centralized resource for Arabidopsis data[2]

- 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)

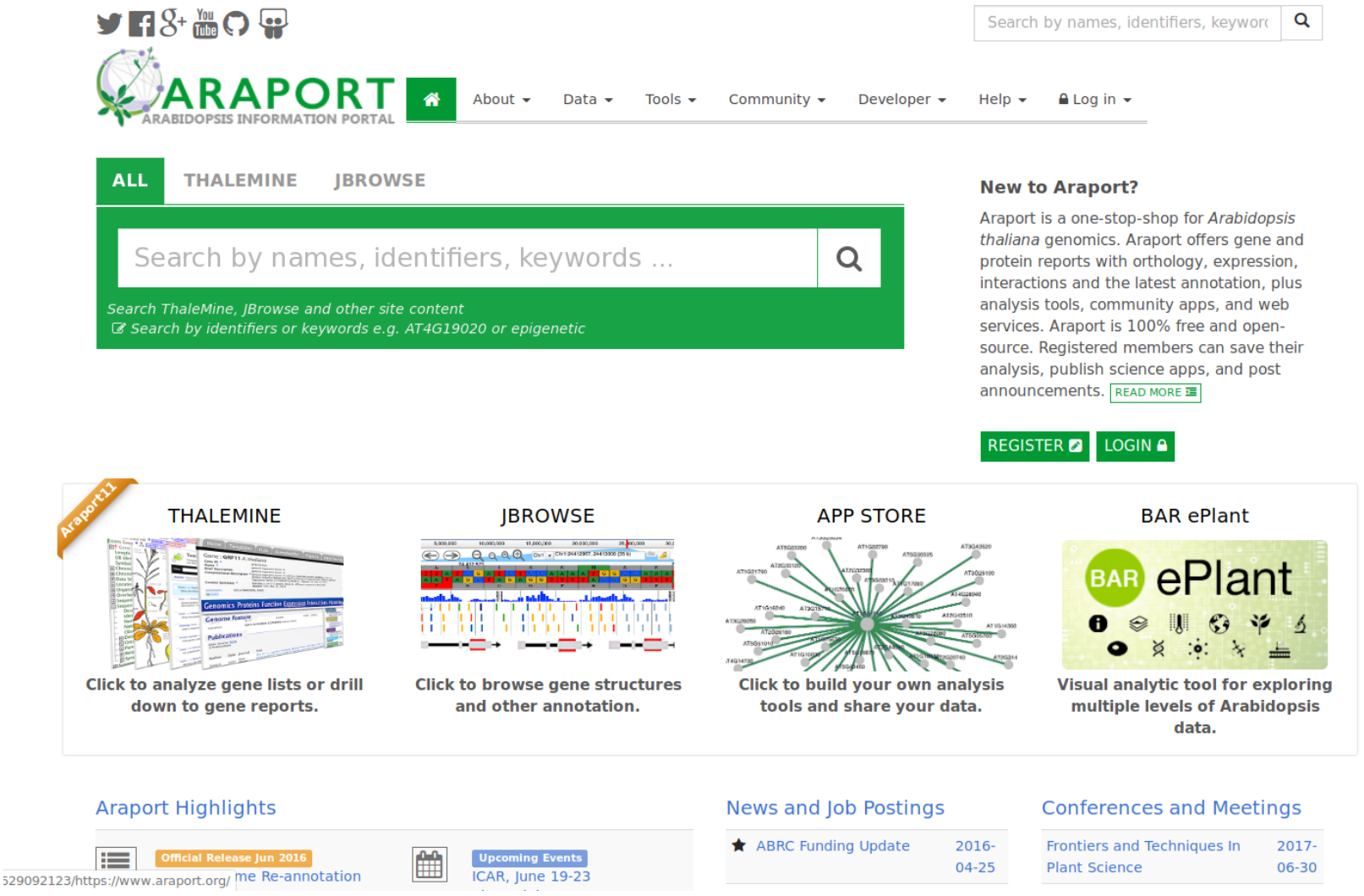
[2] Krishnakumar et al. 2015.

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 serveral apps and web services over the next few years
- Biggest contribution: BAR ePlant app

Araport: Arabidopsis Information Portal



Araport: Arabidopsis Information Portal

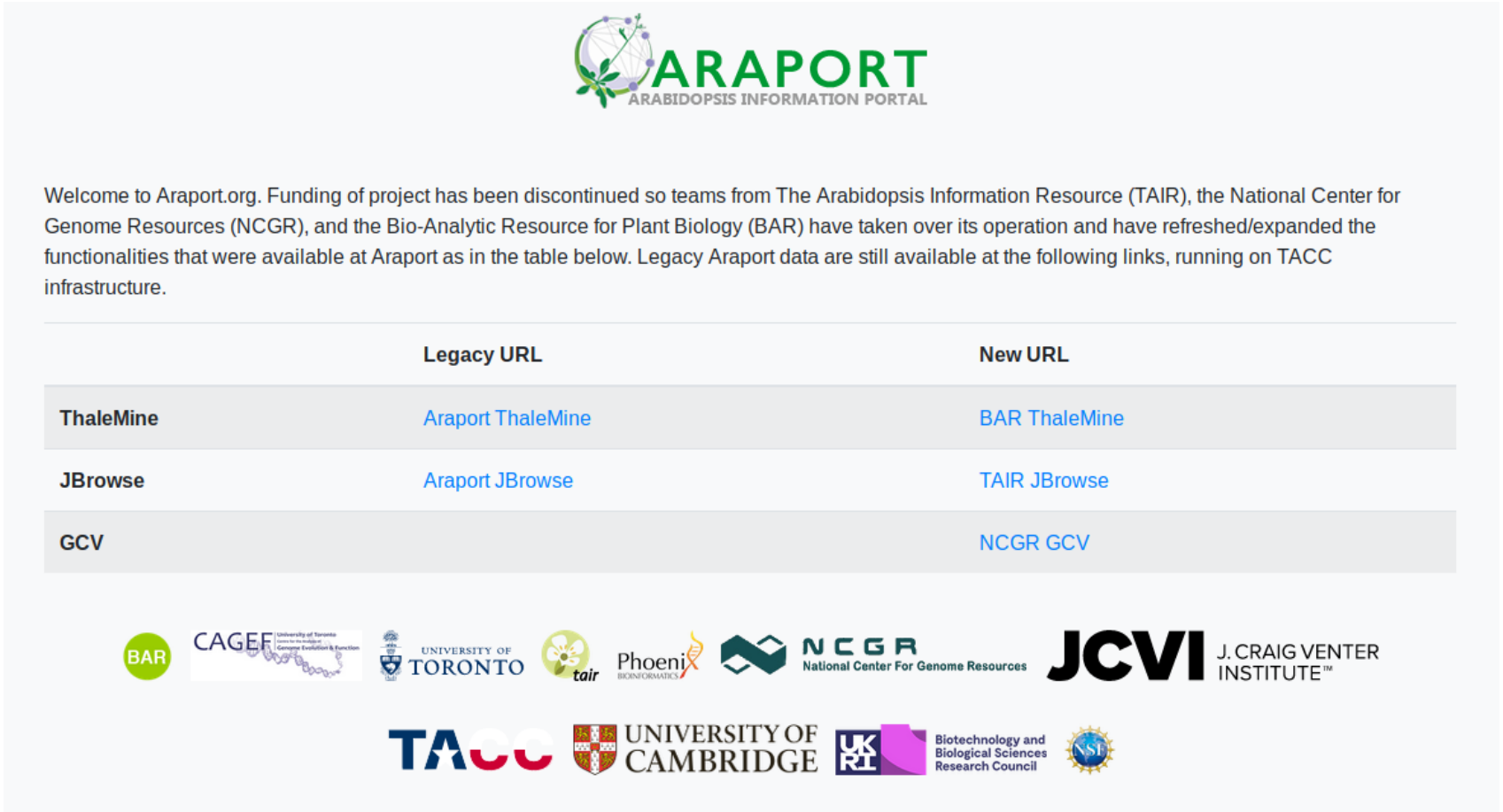
Funding status

- 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



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
- To difficult to upgrade to InterMine 3.1.2

BAR ThaleMine: Principles

No initial funding. So we follow these rules[3]

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)

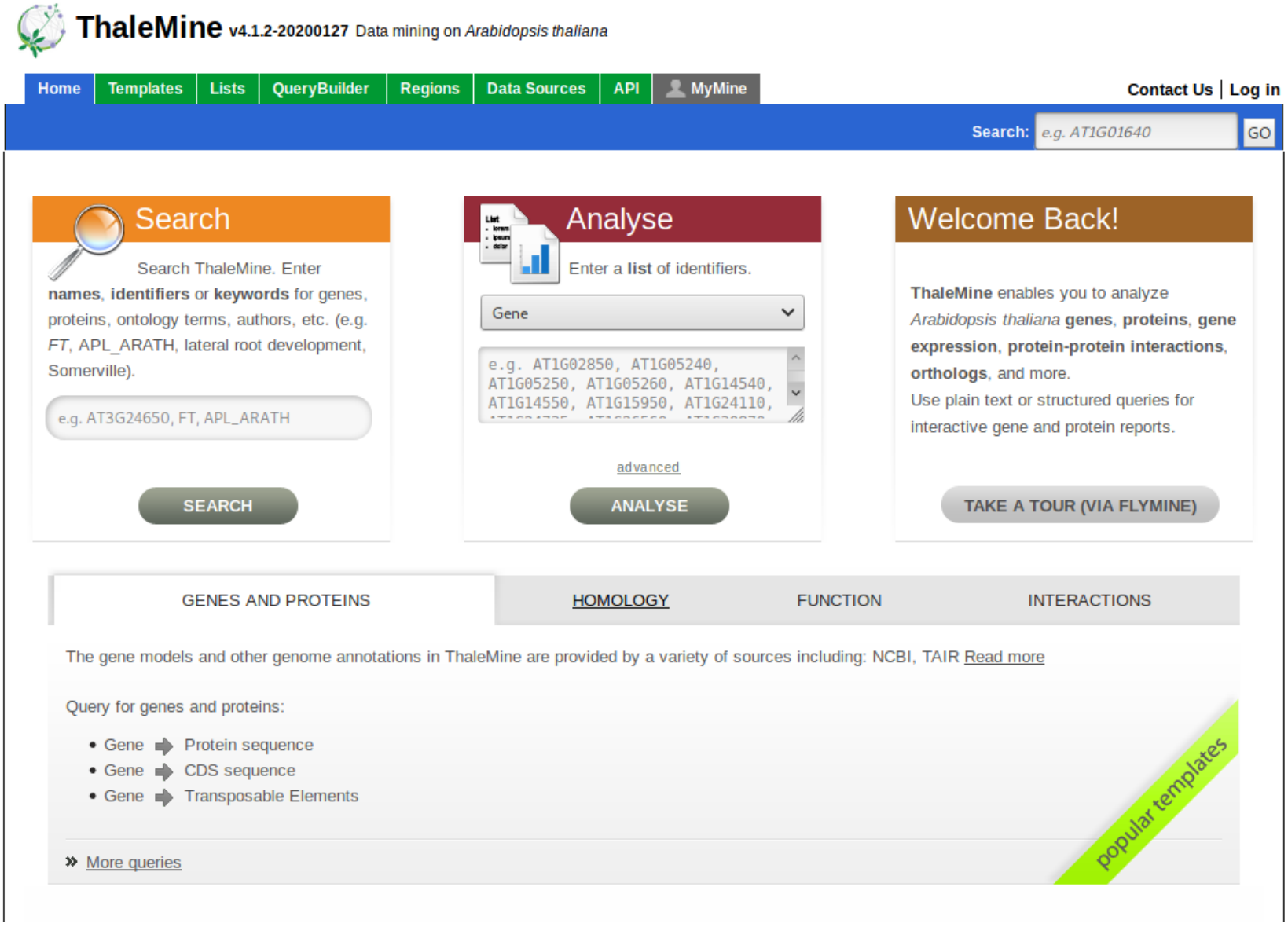
Parsimony Write loaders and displayers only when we have to.

[3] Raymond 2003.

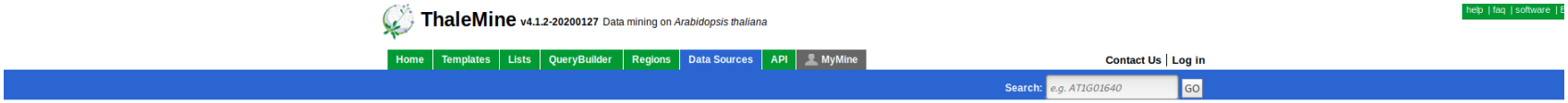
BAR ThaleMine

- 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: What's Working



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 GFF3 data from TAIR	TAIR - Release Araport11 (2016/06/17)	Cheng et al., 2016 - PubMed: 27862469
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: 30398656
Homology	Orthologue and paralogue relationships based on the inferred speciation and gene duplication events in the phylogenetic tree.	Panther - Release 14.1	Mi et al - PubMed: 23193289
	Phytozome Homologs generated with InParanoid	Phytozome - realtime	Goodstein et al - PubMed: 22110026
Gene Ontology	GO annotations from Gene Ontology	Gene Ontology - Release 2020-01-01	Berardini et al., 2004 - PubMed: 15173566 Gene Ontology Consortium - PubMed:10802651
	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	Chatr-Aryamontri et al., 2014 - PubMed: 25428363
	Curated binary and complex protein-protein interactions for Arabidopsis thaliana	IntAct - Downloaded 20200121	Kerrien et al., 2012 - PubMed: 22121220
Expression	Electronic Fluorescent Pictograph (eFP) Visualization paints gene expression information from one of the AtGenExpress data sets or other compendia for a desired gene onto a diagrammatic representation of *Arabidopsis thaliana* plants.	BAR eFP Webservice - realtime	Winter et al., 2007 - PubMed: 17684564 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	Obayashi 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: 30398656
	Publications from NCBI	NCBI - Downloaded 20200121	Maglott et al., 2007 - PubMed: 17148475
GeneRIF	Concise phrase describing gene function and publication associated with NCBI Gene records	NCBI - Downloaded 20200121	Maglott et al., 2007 - PubMed: 17148475

BAR ThaleMine: Some Future Plans and Known Issues

The follow is planned

- Arabidopsis Strains[4]
- Gene expression data (Araport11 RNA-Seq[5])
- Salk data (TDNA-Seq)
- Gene histories from JCVI
- TAIR annotations

Known Issues

- web properties from ThaleMine web services
- BlueGenes frontend

[4] Jiao and Schneeberger 2019.

[5] Cheng et al. 2017.

Acknowledgements