

Measuring Impact for TAIR

Eva Huala

Phoenix Bioinformatics



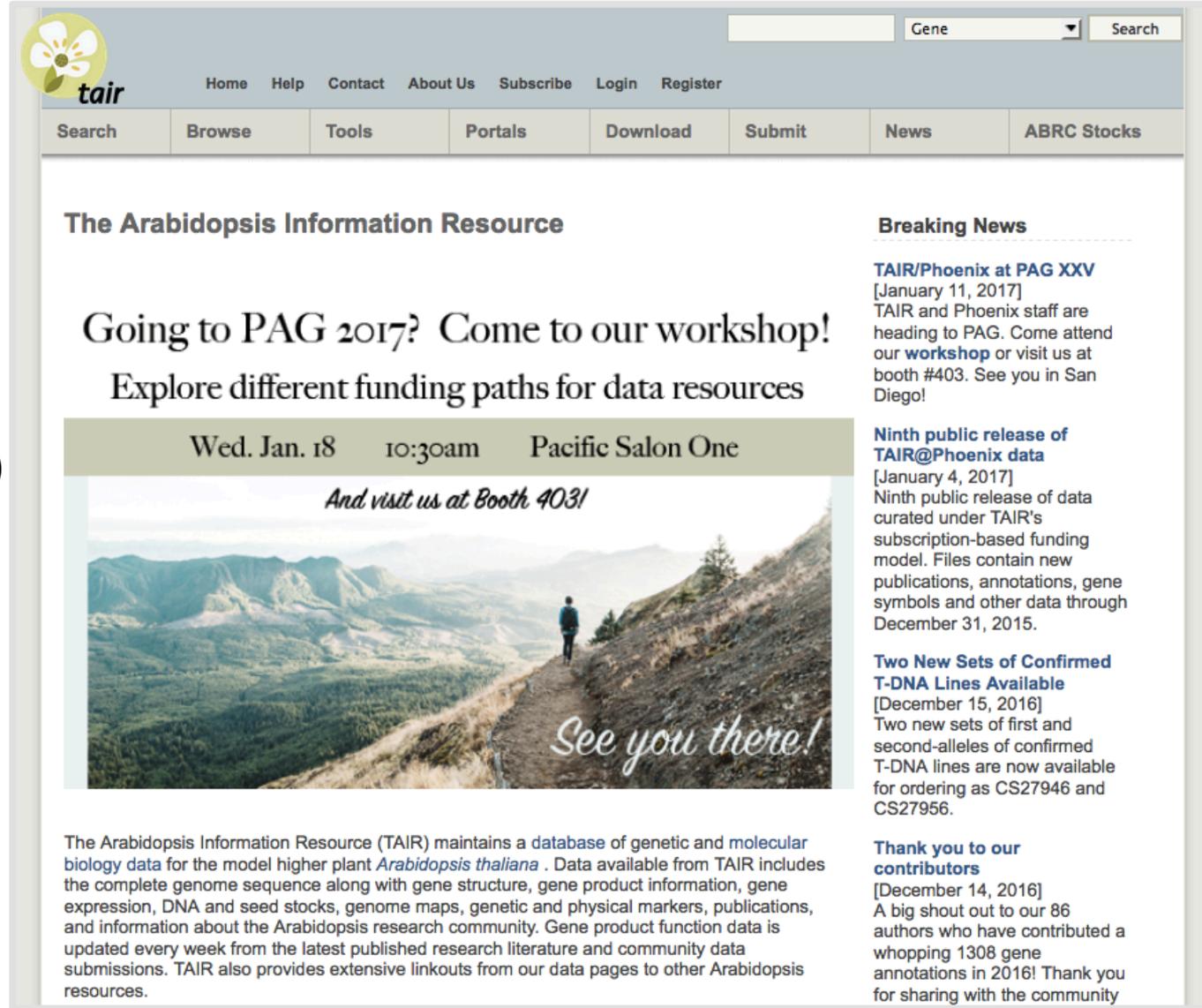
What is TAIR?

Genome database for the model plant *Arabidopsis thaliana*, established in 1999

Focus on **manual curation of gene function** (~33,000 genes)

Was NSF-funded 1999-2013 (1.1M/yr direct costs at peak funding)

Now user supported (1.1M/yr subscriptions including academic institutions and companies)



The screenshot shows the TAIR website interface. At the top left is the TAIR logo, a stylized flower. To its right are navigation links: Home, Help, Contact, About Us, Subscribe, Login, Register. Further right is a search bar with a dropdown menu set to 'Gene' and a 'Search' button. Below the navigation is a horizontal menu with tabs: Search, Browse, Tools, Portals, Download, Submit, News, and ABRC Stocks. The main content area has a heading 'The Arabidopsis Information Resource' and a large announcement: 'Going to PAG 2017? Come to our workshop! Explore different funding paths for data resources'. Below this is a date and time banner: 'Wed. Jan. 18 10:30am Pacific Salon One'. A photograph of a person on a mountain trail is shown with the text 'And visit us at Booth 403!' and 'See you there!'. To the right of the main content is a 'Breaking News' section with three items: 'TAIR/Phoenix at PAG XXV [January 11, 2017]', 'Ninth public release of TAIR@Phoenix data [January 4, 2017]', and 'Two New Sets of Confirmed T-DNA Lines Available [December 15, 2016]'. At the bottom of the main content area is a paragraph describing the TAIR database and its data sources.

The Arabidopsis Information Resource

Going to PAG 2017? Come to our workshop!
Explore different funding paths for data resources

Wed. Jan. 18 10:30am Pacific Salon One

And visit us at Booth 403!

See you there!

The Arabidopsis Information Resource (TAIR) maintains a **database** of genetic and **molecular biology** data for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every week from the latest published research literature and community data submissions. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.

Breaking News

TAIR/Phoenix at PAG XXV
[January 11, 2017]
TAIR and Phoenix staff are heading to PAG. Come attend our **workshop** or visit us at booth #403. See you in San Diego!

Ninth public release of TAIR@Phoenix data
[January 4, 2017]
Ninth public release of data curated under TAIR's subscription-based funding model. Files contain new publications, annotations, gene symbols and other data through December 31, 2015.

Two New Sets of Confirmed T-DNA Lines Available
[December 15, 2016]
Two new sets of first and second-alleles of confirmed T-DNA lines are now available for ordering as CS27946 and CS27956.

Thank you to our contributors
[December 14, 2016]
A big shout out to our 86 authors who have contributed a whopping 1308 gene annotations in 2016! Thank you for sharing with the community

Metrics must reflect purpose, priorities

Curated information source

- Selective data intake plus curation to insure quality
- Synthesized information
- Reuse over deposition
- Meeting needs of researchers

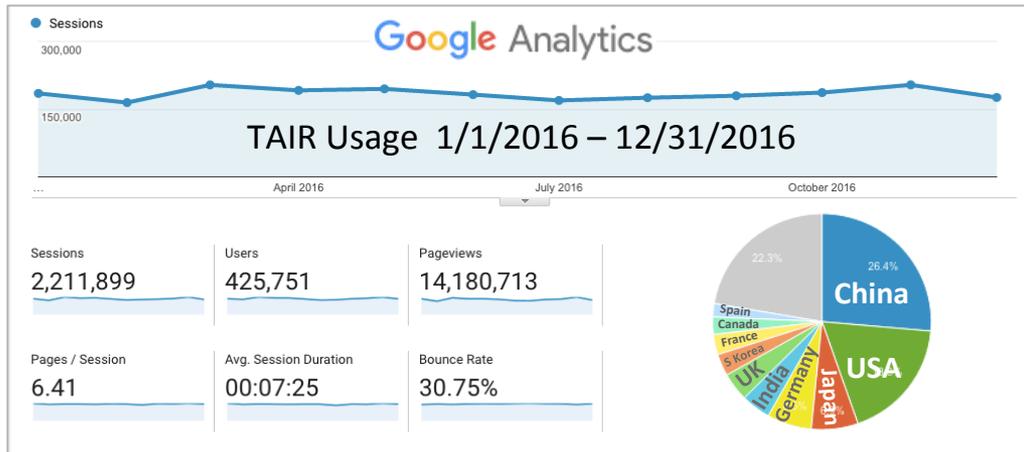
Metrics on data reuse, number of users, impact on research output, satisfaction of researchers

Data archive

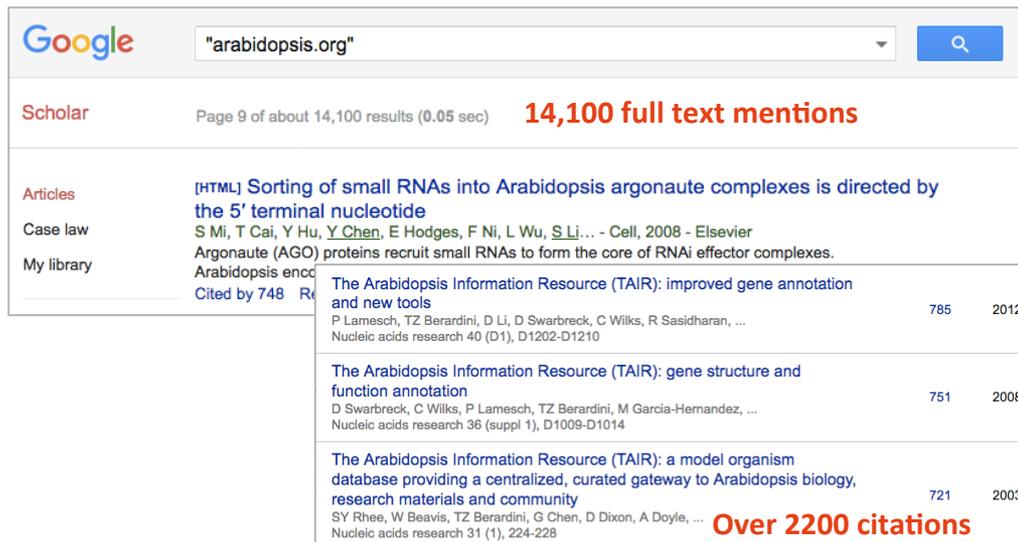
- Intake of datasets plus metadata associated to published work
- Curation of metadata
- Uses – Replication of published experiments
- Meeting needs of government, institutions and journals

Metrics on rate of data deposition, recoverability of datasets, usefulness to govt and journals

TAIR Impact Metrics – Usage and Citation

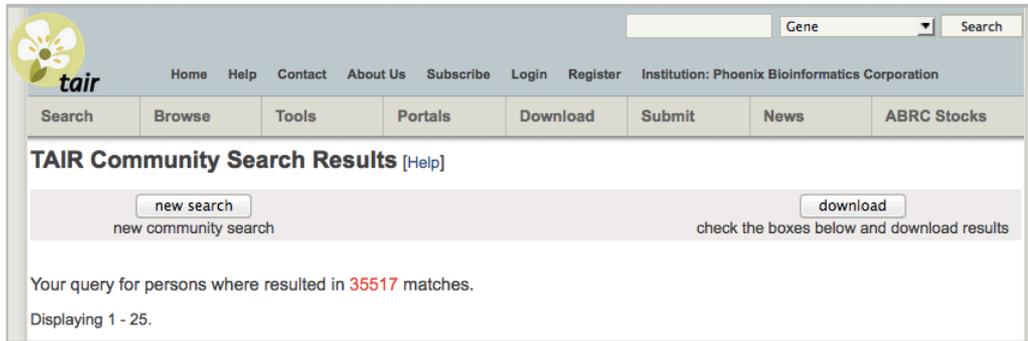


- Google Analytics usage data
 - Unique visitors, sessions, page views
 - Institutional breadth
 - Geographic breadth



- Google Scholar
 - Full text mentions (with carefully chosen search terms)
 - Citations (always an undercount, resources are not consistently cited)

Additional TAIR Impact Metrics



TAIR Community Search Results [Help]

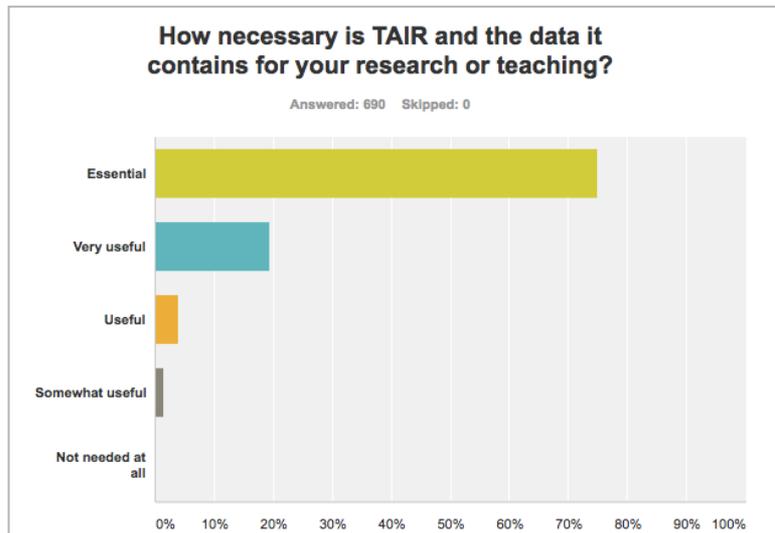
new search download

new community search check the boxes below and download results

Your query for persons where resulted in 35517 matches.

Displaying 1 - 25.

- Number of Registered Users



- Value to community and perception of quality
 - Survey assessments

Repository life cycle

- Early phase projects should show increasing impact metrics over time
- Mature projects should show high, stable or increasing impact

Social, technical, and financial issues

- Social
 - Best practices for measuring various types of impact should be developed.
 - Requirements for impact metrics should be raised up front (in calls for proposals).
 - Proposers should be asked to discuss metrics by which impact can be evaluated for future funding decisions.
 - Evaluation of metrics by grant review panel could be required as part of the renewal process.
- Technical
 - Impact metrics must be standardized for meaningful comparisons
 - They must also be easy to implement
- Financial
 - Low cost measurement and reporting options exist (e.g. zero cost for Google Analytics and Google Scholar)