

Bioinformatics

Nicholas Provar

nicholas.provar@utoronto.ca

University of Toronto



With input from MASC Bioinformatics Subcommittee members and the wider Arabidopsis community

Recently developed Open Tools and Resources for Arabidopsis Researchers

Arabidopsis Informatics

As mentioned last year, TAIR, BAR and the National Center for Genome Resources (NCGR) in New Mexico collaborated to ensure that the data and tools formerly provided by Araport remain available to the community. Araport's JBrowse instance migrated to TAIR, Thalemine was redeployed with updated data at the BAR, and a new tool – the Genome Context Viewer – for exploring micro- and macrosynteny in *Arabidopsis thaliana* ecotypes was released by the NCGR. This effort is now described in a paper published in *The Plant Cell* by Pasha *et al.* (2020).

TAIR

TAIR released a new Synteny Viewer. This tool uses the SynMap tool at CoGE (Lyons and Freeling, 2008) to provide pre-computed syntenic comparisons between *A. thaliana* and genomes from 36 other plant species. New data tracks were added to the Araport JBrowse instance that TAIR resuscitated. New/updated data include TIF-seq and plaNET-seq data from the Marquardt lab (Kindgren *et al.*, 2020; Thomas *et al.*, 2020), and Ws/Ws-2 polymorphism data from GBrowse. A new tool to enable community curation of gene function called the Generic Online Annotation Tool (GOAT, <https://goat.phoenixbioinformatics.org/>) was also introduced for collecting Gene Ontology and Plant Ontology annotations from the research community.

TAIR continued to incorporate new Arabidopsis gene function annotations made by curators and members of our community. As of March 2021, around 30% of Arabidopsis genes have at least one experimentally based annotation for GO Molecular Function, GO Biological Process, and GO Cellular Component. And that percentage or higher have GO support based on another annotation method (see TAIR report for exact numbers).

BAR

The Bio-Analytic Resource worked with Ute Roessner's group in Australia to develop an Arabidopsis Lipid Map eFP Browser for visualizing the amount of any one of hundreds of lipids across 7 different tissues (Kehelpannala *et al.*, 2021). It also worked collaboratively with Patrick Schaefer's group at the University of Warwick to develop a cell-type specific immunity view for roots in ePlant (Rich-Griffin *et al.*, 2020).

A Plant Reactome (Tello-Ruiz *et al.*, 2018) pathways viewer was added to ePlant (Waese-Perlman *et al.*, 2021). The BAR also rolled out two new tools for exploring gene regulatory networks and natural variation: AGENT (Lau *et al.*, 2021) and Variant Viewer (Cumming *et al.*, 2021), respectively. Large-scale Data Sets and Methods of Note.

The screenshot displays the 'Variant Viewer' interface. The top part is a landing page with a title 'Welcome to the A.thaliana Variant Viewer' and three numbered steps: 1. Submit your genes one by one, 2. Align your genes, and 3. Export your alignment. The bottom part shows the output page with a list of genes and their isoforms, with a detailed view of a specific variant highlighted in orange.

The Variant Viewer: an application for viewing variants from the 1001 Genomes Project.

Graphics are rendered using d3.js and plotly.js (see Experimental Procedures). Top: landing page, bottom: output page. Variants are pulled on the fly from the 1001 Genomes Project API (<http://tools.1001genomes.org/api>) for each gene entered in the search box. Selected genes are displayed as simple gene models and variants are plotted onto models to give an overview of the variants in the gene region. Hover events are triggered allowing users to inspect individual exons, UTRs, coding sequences, and variants. Users are also able to export data for their records.

As mentioned last year, Edward Marcotte's group used co-fractionation mass spectrometry to identify protein complexes in 13 plant species, including Arabidopsis. More than 3 million pairwise interactions were elucidated in this incredible study, which permits the identification of conserved and rewired protein complexes in plants (McWhite *et al.*, 2020). The Arabidopsis data consisting of 21,910 interactions have now been loaded into the BAR's Arabidopsis Interactions Viewer database (<http://bar.utoronto.ca/interactions2>) and into BioGRID.

Also in terms of proteomics, the Kuster group and colleagues released a draft of the Arabidopsis proteome (Mergner *et al.*, 2020), encompassing more than 18,000 gene products and 43,000 phosphorylation sites. The data are available in ATHENA and the ProteomicsDB.

The Marquardt Lab's plant native elongating transcripts sequencing (plaNET-seq) method showed transient genome-wide reprogramming of nascent RNAPII transcription during cold, with big implications for understanding plant stress response (Kindgren *et al.*, 2020).

Another method, TIF-seq, from the same lab (Thomas *et al.*, 2020) identified many transcripts

where promoter-proximal RNAPII stalling and premature transcriptional termination could act as checkpoint that governs plant gene expression. As mentioned, these data are available in TAIR's JBrowse instance. scRNA-Seq Search Tools. scRNA-seq data sets continued to be published.

Another tool, the Plant Single Cell RNA-Sequencing Database, for querying such data was published by Ma *et al.* (2020) at <https://www.zmbp-resources.uni-tuebingen.de/timmermans/plant-single-cell-browser/>, This is in addition to the Wang Lab's Root Cell Atlas search tool at <http://wanglab.sippe.ac.cn/rootatlas/> (Zhang *et al.*, 2019) and the BAR's eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi?dataSource=Single_Cell), which provides the ability to query scRNA-seq data from Ryu *et al.* (2019).

The Weigel Lab published a method for performing genome-wide association studies in species using incomplete genomes and a k-mer-based approach (Voichek and Weigel, 2020). They were able to recover significant, known associations for 1,582 phenotypes from 104 Arabidopsis studies but more importantly, they were able to identify new associations in cases where the reference Col-0 genome was missing a chunk of sequence as compared to other ecotypes.

Pedagogy, Policy and Outreach:

The Plant Cell Atlas project (Rhee *et al.*, 2019) continued to pick up steam with a well-attended virtual workshop last May. Keep an eye on <http://www.plantcellatlas.org/> for updates!

Nicholas Provart and colleagues published a retrospective of 20 years of Arabidopsis genomics and described the amazing advances that the publication of the Arabidopsis genome in 2000 enabled (Provart *et al.*, 2020).

Alex Mason and colleagues published a hefty book chapter "Bioinformatic Tools in Arabidopsis Research", covering 32 useful Arabidopsis tools and websites, with step-by-step instructions (Mason *et al.*, 2021). TAIR launched a webinar series using Zoom and Voov platforms for 'captive' home audiences. TAIR also continued its partnership with <https://www.micropublication.org/>, a platform for very short peer-reviewed reports, and this year curated/shepherded 10 Arabidopsis into the Micro-publication framework.

References

- Cumming, M., Esteban, E., Lau, V., Pasha, A., and Provart, N.J. (2021). A 'Variant Viewer' for investigating population-level variation in *Arabidopsis thaliana*. bioRxiv: 2021.04.21.440793.
- Kehelpannala, C., Rupasinghe, T., Pasha, A., Esteban, E., Hennessy, T., Bradley, D., Ebert, B., Provart, N.J., and Roessner, U. (2021). An Arabidopsis lipid map reveals differences between tissues and dynamic changes throughout development. *Plant J.*
- Kindgren, P., Ivanov, M., and Marquardt, S. (2020). Native elongation transcript sequencing reveals temperature dependent dynamics of nascent RNAPII transcription in Arabidopsis. *Nucleic Acids Res.* 48: 2332–2347
- Lau, V., Woo, R., Pereira, B., Pasha, A., Esteban, E., and Provart, N.J. (2021). AGENT: the Arabidopsis Gene Regulatory Network Tool for Exploring and Analyzing GRNs. bioRxiv: 2021.04.28.441830.
- Lyons, E. and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* 53: 661–673

- Ma, X., Denyer, T., and Timmermans, M.C.P. (2020). PscB: A Browser to Explore Plant Single Cell RNA-Sequencing Data Sets. *Plant Physiol.* 183: 464–467
- Mason, G.A., Cantó-Pastor, A., Brady, S.M., and Provart, N.J. (2021). Bioinformatic Tools in Arabidopsis Research. In *Arabidopsis Protocols*, J.J. Sanchez-Serrano and J. Salinas, eds, *Methods in Molecular Biology*. (Springer US: New York, NY), pp. 25–89
- McWhite, C.D. *et al.* (2020). A Pan-plant Protein Complex Map Reveals Deep Conservation and Novel Assemblies. *Cell* 181: 460-474. e14. Mergner, J. *et al.* (2020). Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature* 579: 409–414
- Pasha, A., Subramaniam, S., Cleary, A., Chen, X., Berardini, T.Z., Farmer, A., Town, C., and Provart, N.J. (2020). Araport Lives: An Updated Framework for Arabidopsis Bioinformatics. *Plant Cell*
- Provart, N.J., Brady, S.M., Parry, G., Schmitz, R.J., Queitsch, C., Bonetta, D., Waese, J., Schneeberger, K., and Loraine, A.E. (2020). Anno genominis XX: 20 years of Arabidopsis genomics. *Plant Cell*.
- Rhee, S.Y., Birnbaum, K.D., and Ehrhardt, D.W. (2019). Towards Building a Plant Cell Atlas. *Trends Plant Sci.* 24: 303–310. - Rich-Griffin, C. *et al.* (2020). Regulation of Cell Type-Specific Immunity Networks in Arabidopsis Roots. *Plant Cell* 32: 2742–2762
- Ryu, K.H., Huang, L., Kang, H.M., and Schiefelbein, J. (2019). Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiol.* 179: 1444
- Tello-Ruiz, M.K. *et al.* (2018). Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic Acids Res.* 46: D1181–D1189
- Thomas, Q.A., Ard, R., Liu, J., Li, B., Wang, J., Pelechano, V., and Marquardt, S. (2020). Transcript isoform sequencing reveals widespread promoter-proximal transcriptional termination in Arabidopsis. *Nat. Commun.* 11: 2589
- Voichek, Y. and Weigel, D. (2020). Identifying genetic variants underlying phenotypic variation in plants without complete genomes. *Nat. Genet.* 52: 534–540
- Waese-Perlman, B. *et al.* (2021). ePlant in 2021: New Species, Viewers, Data Sets, and Widgets. *bioRxiv*: 2021.04.28.441805
- Zhang, T.-Q., Xu, Z.-G., Shang, G.-D., and Wang, J.-W. (2019). A Single-Cell RNA Sequencing Profiles the Developmental Landscape of Arabidopsis Root. *Mol. Plant* 12: 648–660.