MASC Subcommittee Reports

Bioinformatics

Nicholas Provar
nicholas.provart@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 Freel- ing, 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.
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).


References


