



LARGE-SCALE BIOLOGY ARTICLE

ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology^{OPEN}

Jamie Waese,^a Jim Fan,^b Asher Pasha,^a Hans Yu,^a Geoffrey Fucile,^c Ruian Shi,^a Matthew Cumming,^a Lawrence A. Kelley,^d Michael J. Sternberg,^d Vivek Krishnakumar,^e Erik Ferlanti,^e Jason Miller,^e Chris Town,^e Wolfgang Stuerzlinger,^f and Nicholas J. Provart^{a,1}

^a Department of Cell and Systems Biology/Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario M5S 3B2, Canada

^b Department of Computer Science, University of Waterloo, Ontario N2L 3G1, Canada

^c SIB Swiss Institute of Bioinformatics, sciCORE Computing Center, University of Basel, CH-4056 Basel, Switzerland

^d Imperial College London, London SW7 2AZ, United Kingdom

^e Araport.org/J. Craig Venter Institute, Rockville, Maryland 20850

^f School of Interactive Arts and Technology, Simon Fraser University, British Columbia V3T 0A3, Canada

ORCID IDs: 0000-0002-9315-0520 (A.P.); 0000-0002-3210-2656 (L.A.K.); 0000-0002-1884-5445 (M.J.S.); 0000-0002-5227-0200 (V.K.); 0000-0001-5128-1584 (E.F.); 0000-0002-6912-2925 (J.M.); 0000-0001-5551-7232 (N.J.P.)

A big challenge in current systems biology research arises when different types of data must be accessed from separate sources and visualized using separate tools. The high cognitive load required to navigate such a workflow is detrimental to hypothesis generation. Accordingly, there is a need for a robust research platform that incorporates all data and provides integrated search, analysis, and visualization features through a single portal. Here, we present ePlant (<http://bar.utoronto.ca/eplant>), a visual analytic tool for exploring multiple levels of *Arabidopsis thaliana* data through a zoomable user interface. ePlant connects to several publicly available web services to download genome, proteome, interactome, transcriptome, and 3D molecular structure data for one or more genes or gene products of interest. Data are displayed with a set of visualization tools that are presented using a conceptual hierarchy from big to small, and many of the tools combine information from more than one data type. We describe the development of ePlant in this article and present several examples illustrating its integrative features for hypothesis generation. We also describe the process of deploying ePlant as an “app” on Araport. Building on readily available web services, the code for ePlant is freely available for any other biological species research.

INTRODUCTION

Many data visualization tools have been created to provide visual depictions of information that is normally not easily visible. Yet, to explore complex phenomena at multiple levels of analysis using existing tools, researchers must visit multiple websites, each with their own user interface, data input requirements, methods for organizing and categorizing information, and design language for visualizing the particular layer of information that they were created to display.

There are many connections between biological entities at different levels of analysis. Figure 1 illustrates the depth and complexity of the relationships between these levels. An investigation of seed development in *Arabidopsis thaliana* may focus on one particular transcription factor gene, *ABI3*, but at which level of analysis? Environmental factors can lead to DNA

polymorphisms being retained, leading to natural variation in different populations at the level of protein sequences and structures. Subtle variations at these levels can then affect signaling and signal transduction cascades, protein interaction networks, metabolism, subcellular localization, spatio-temporal distribution, and ultimately phenotypic response across various ecotypes.

A thorough investigation of *ABI3* would require visiting Araport (Krishnakumar et al., 2015) or TAIR (Berardini et al., 2015) for annotation and sequence information, the BAR Arabidopsis Interactions Viewer (Geisler-Lee et al., 2007) or BioGRID (Chatr-Aryamontri et al., 2015) for protein-protein interactions, SUBA3 (Tanz et al., 2013) for subcellular localization information, Phyre2 (Kelley et al., 2015) for 3D molecular models, Reactome (Joshi-Tope et al., 2005) for signal, metabolic, and gene regulation pathway charts, and dozens of other sites, each dedicated to their own particular layer of analysis.

It is difficult to develop hypotheses about complex processes when the information is hard to assemble and laborious to interpret. When researchers must devote a portion of their cognitive load to a computer interface instead of the subjects they are exploring, their train of thought will be disrupted and their overall productivity will decrease (Ware, 2012). The effects of

¹ Address correspondence to nicholas.provart@utoronto.ca.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Nicholas J. Provart (nicholas.provart@utoronto.ca).

^{OPEN}Articles can be viewed without a subscription.

www.plantcell.org/cgi/doi/10.1105/tpc.17.00073

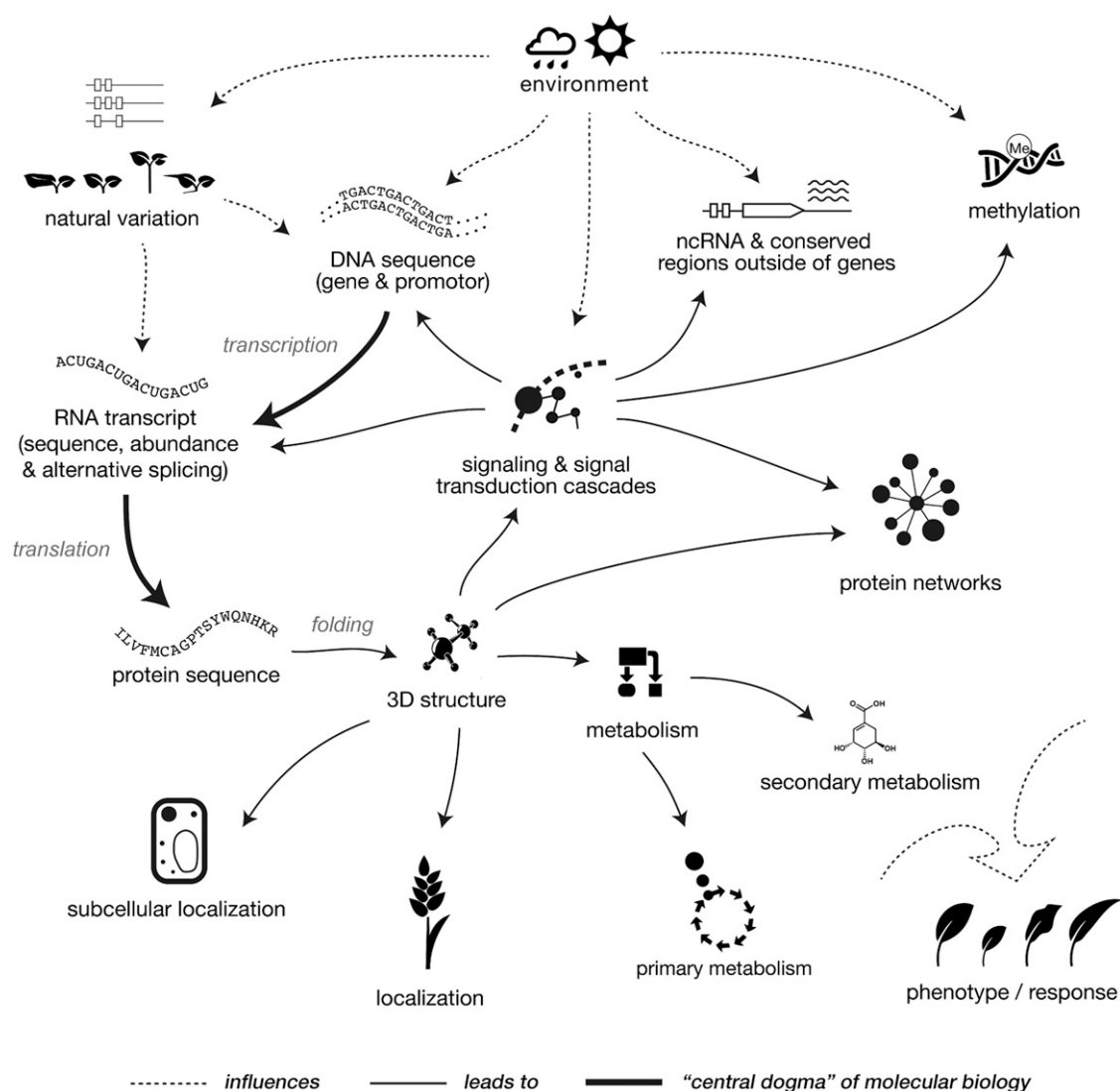


Figure 1. Connections between Biological Entities at Different Levels of Analysis.

interruptions and distractions on cognitive productivity are well described: information overload, increased stress, decreased decision-making accuracy, and the narrowing of attention resulting in the ability to process fewer information cues (Speier et al., 2003). The systems biology research workflow could be improved with the availability of an integrated software platform, with what is known in the information visualization community as a "transparent" user interface, i.e., an interface that "is so easy to use that it all but disappears from consciousness" (Ware, 2012).

This project addresses these challenges by combining several data visualization tools into the same interface, ordering them into a hierarchy of scale and providing zoom transitions and integrative connections between the layers so users can explore multiple levels of biological data in new ways (Figure 2). We postulate that applying the principles of user-centered design to build an integrated visual analytic tool for exploring multiple levels of plant data should improve the ability of researchers to extract

information from their data, identify connections between layers, facilitate hypothesis generation, and, ultimately, promote a deeper understanding of biological processes and functioning.

RESULTS

We have developed a data integration software tool, ePlant, that not only applies tailored visualizations to more than 10 data types but also integrates data across at least 10 orders of magnitude, from the kilometer scale (natural variation data) to the nanometer scale (protein structure and sequence data), into one easy-to-use interactive framework. We have developed ePlant based on Arabidopsis data, and in this case it taps into >35 million gene expression measurements, experimentally documented subcellular localizations for 10,910 Arabidopsis proteins (with predictions for most of the proteome), ~100,000 protein-protein and ~2.7 million protein-DNA interactions, Phyre2-predicted structures covering

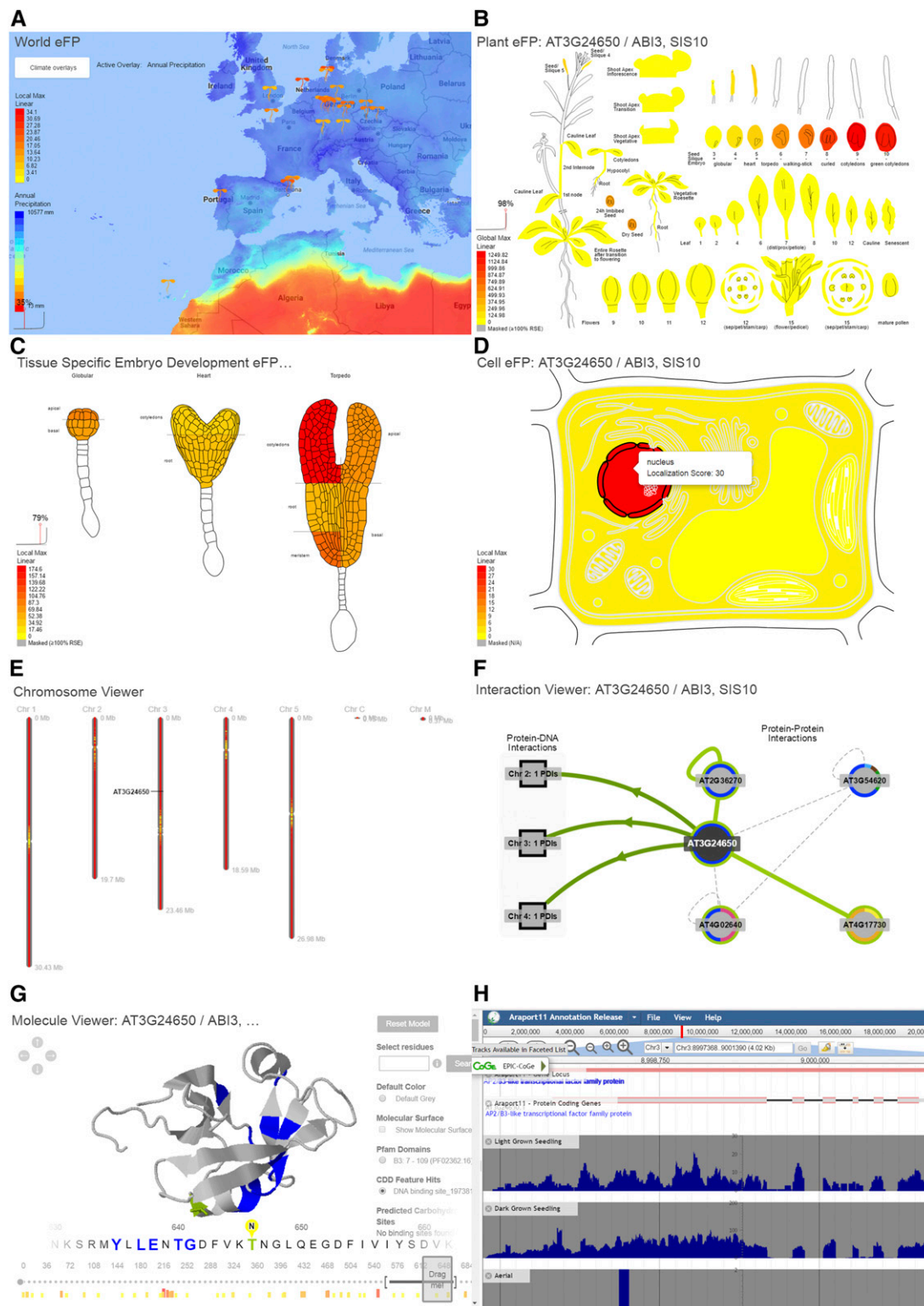


Figure 2. ePlant's Module Viewers (Each Displaying Data for *ABI3*), Illustrating ePlant's Hierarchy of Scale. To see detail, view the online version. World eFP Viewer (A), Plant eFP Viewer (B), Tissue and Experiment eFP Viewer (C), Subcellular eFP Viewer (D), Chromosome Viewer (E), Interaction Viewer (F), Molecule Viewer (G), and Sequence Viewer (H). High-resolution vector graphics are available for most views (see example in Supplemental File 1). ePlant outputs are freely usable via an "open" license.

23,091 gene products, and 6.19 million nonsynonymous single nucleotide polymorphisms (SNPs) from the 1001 Proteomes website (Supplemental Table 1). In addition, more than a dozen nucleotide-resolution data types (including 100 gigabases of RNA-seq data used to reannotate the Arabidopsis genome in the Araport 11 release) are also available via Araport's JBrowse instance that has been incorporated into ePlant. We have also created linkages across data scales such that it is possible to ask questions such as, "Is there a polymorphism that causes a nonsynonymous amino acid change close to the DNA binding site of my favorite transcription factor?"

System Architecture and User Interface

ePlant is a collection of programs written with HTML, CSS, JavaScript, and jQuery, bundled together within a custom zoomable user interface (ZUI) framework (Figures 2 and 3A). It is HTML5 compliant and runs within a web browser on most laptops, desktops, and some tablets.

ePlant was designed to support data fed dynamically from web services. Upon entering a gene name, alias, or AGI ID in the gene selection box in the upper left corner, a data loading management script sends queries to multiple web services (Supplemental Table 1) to retrieve data for each of the ePlant modules. Data are returned asynchronously via AJAX so the program does not freeze while waiting for data to download. Once everything that has been requested has been returned, the data are passed to a function that initializes each module's viewer for each loaded gene.

The ePlant user interface has two main elements (Figure 3B): the gene panel and navigation icons on the left and the module viewer panel on the right. For users who do not know which gene (or genes) they want to look at, the "Expression Angler" button opens a tool that helps identify genes based on a user-defined expression pattern (Austin et al., 2016), and the "Mutant Phenotype Selector" button opens a tool that helps identify genes based on Lloyd and Meinke's mutant phenotype classification system (Lloyd and Meinke, 2012). Both of these features are discussed later in this article.

Downloaded genes appear as rectangular bars in the gene panel. The currently selected gene is colored green. A vertical stack of icons for selecting the ePlant module to be viewed separates the gene panel from the module viewer panel. The viewers currently include *Gene Information Viewer*, *Publication Viewer*, *Heat Map Viewer*, *World eFP Viewer*, *Plant eFP Viewer*, *Tissue and Experiment eFP Viewer*, *Cell eFP Viewer*, *Chromosome Viewer*, *Protein Interaction Viewer*, *Molecule Viewer*, *Sequence Viewer*, and *Links to External Tools*. Icons appear gray when a module is unavailable, turn black once the data have loaded, and are highlighted green when the module is active (i.e., has been selected for viewing by the user).

The module viewer panel shows the content of whichever ePlant module is currently selected. A tab selector at the top of the screen enables users to create multiple views. Beneath the tab selector, a toolbar contains icons for controlling various features, such as *Session history*, *Screen grab*, *Zoom in/out*, *Absolute/relative display*, *Compare genes*, *Filter data*, *Custom color palette*, *Global/local/custom color gradient*, *Get citation and experiment information*, and *Download raw data for the currently selected view*. A global options menu in the top right corner of the page allows users to toggle *Zoom transitions*, *Tooltips*, and *New user*

information pop-ups on and off. There is also an option to *Create a custom URL* that automatically restores the current session upon sharing it with a colleague.

ZUIs take advantage of human spatial perception and memory to display more information than could otherwise fit on a desktop computer display. ZUIs have been shown to significantly improve users' ability to find information across multiple layers of data (Helt et al., 2009; Bederson, 2010), and animated navigation transitions have been shown to increase user task performance, even taking into account the time of the transitions themselves (Klein and Bederson, 2005). A custom ZUI framework was created to handle zoom transitions between data visualization modules in ePlant. These transitions do not attempt to map spatial and size relationships between the layers. Rather, they produce a "2.5D effect" to indicate a conceptual relationship between the layers.

The following ePlant module viewers are organized following a hierarchy of scale from "big" to "small."

Gene Information and Publications Viewer

The Gene Information Viewer (Figure 4A) provides top level access to aliases, full name, description, computational description, curator summary, location, and a visual representation of the gene model structure with intron/exon information about the currently selected gene. It also provides DNA and protein sequences. Immediately beneath it, the Publications Viewer (Figure 4B) provides a list of publications and gene "reference into function" records (GeneRIFs) about the currently selected gene, along with links to the actual papers on PubMed. Both are powered by web services from Araport (Krishnakumar et al., 2015). These modules are atypical for ePlant because they are primarily text based. They were added after user testing feedback.

Heat Map Viewer

The Heat Map Viewer displays all the expression levels for all the samples of all the genes that are loaded, along with the corresponding subcellular localizations of their gene products in one neatly formatted table. This view is near the top of the conceptual hierarchy because it provides a Gestalt sense of the similarities or differences of all the genes/gene products that are loaded. Table cells that are colored red denote high expression levels (or high levels of confidence in a protein's subcellular localization), and table cells that are colored yellow denote low levels. The Heat Map Viewer provides an overview, but determining what sample a red cell corresponds to requires extra mouse-over steps.

Figure 5 shows a heat map of 25 genes that were identified with the Expression Angler tool for having similar expression patterns as At3g24650 (*ABI3*). This can be quickly confirmed by scanning the heat map to see if the red cells (which represent samples with high expression) are mostly aligned to the same columns or if there are any obvious outliers. In this case, there are not.

When mapping expression levels across several genes, it is important to clarify whether the color gradient should be determined locally or globally. ePlant can map expression levels locally, which means the color gradient for each view is determined by the minimum and maximum expression levels of that view. This can be useful to help discern a gene's expression pattern even if its

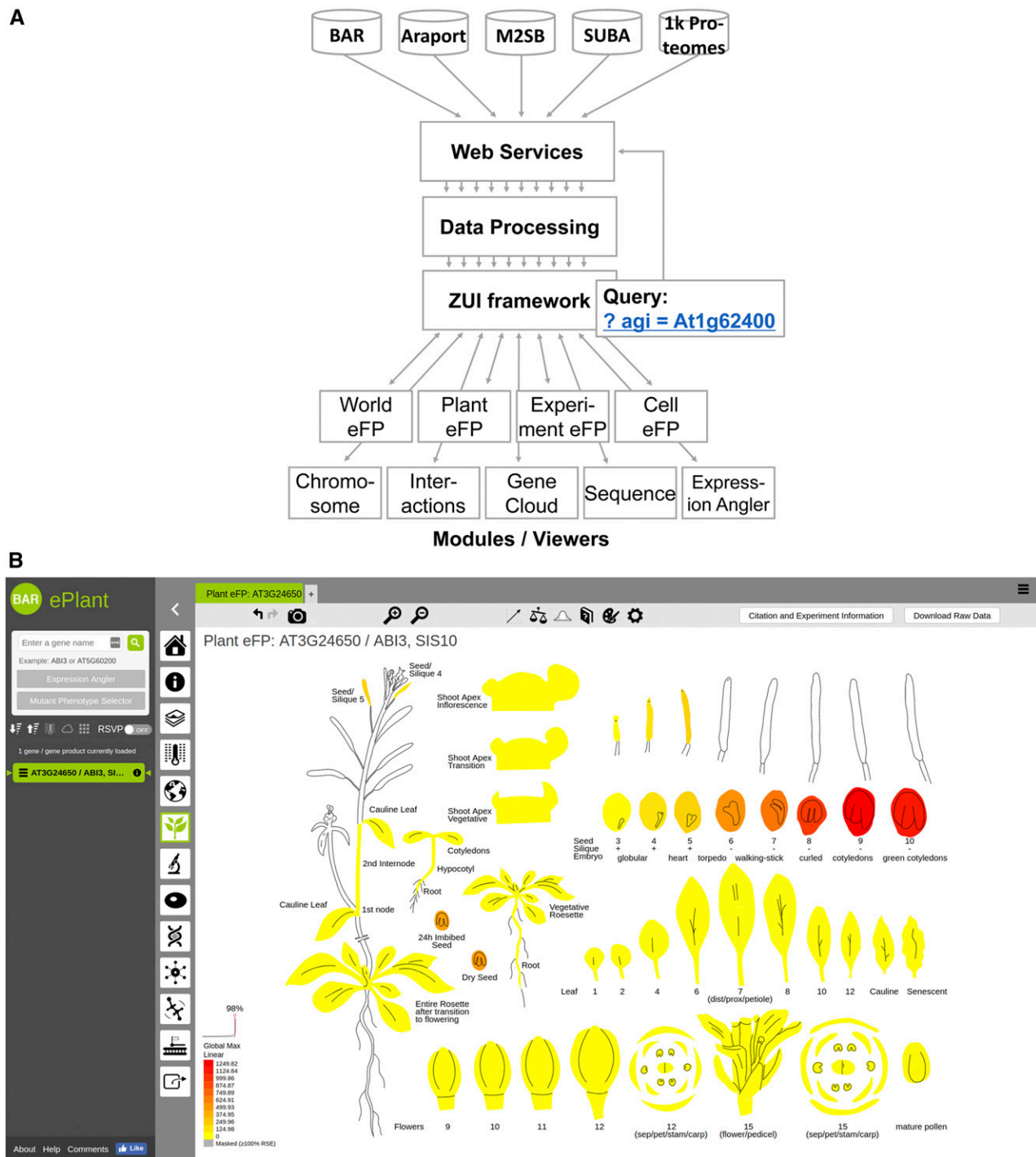


Figure 3. ePlant System Design and User Interface.

(A) ePlant system architecture.

(B) ePlant user interface showing the expression pattern of *ABI3* with the Plant eFP Viewer. See detail in online version of figure.

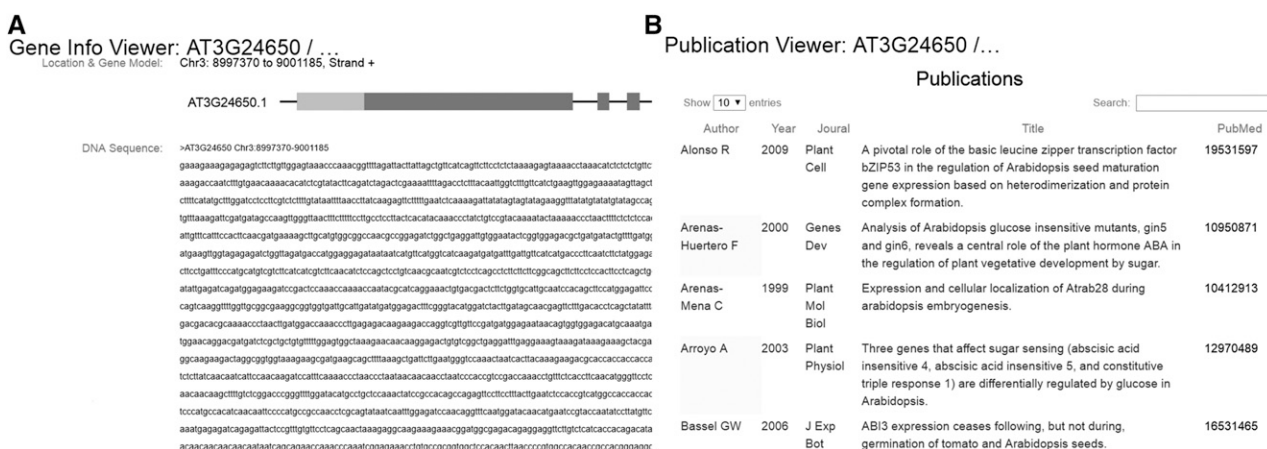


Figure 4. ePlant Views Providing Information from Araport.

(A) ePlant Gene Information Viewer.

(B) ePlant Publications Viewer.

maximum expression level is significantly lower than other genes being displayed. The “global” color gradient (default) is useful for comparing gene expression levels, with the minimum and maximum values determined by the lowest and highest expression

level of all the genes that have been downloaded. In Figure 5, which uses the “global” color gradient setting, the maximum expression level for *ABI3* is 1249 in the stage 9 seeds (as noted in the mouse-over tooltip). Its most similar coexpressed gene is *At2g27380*

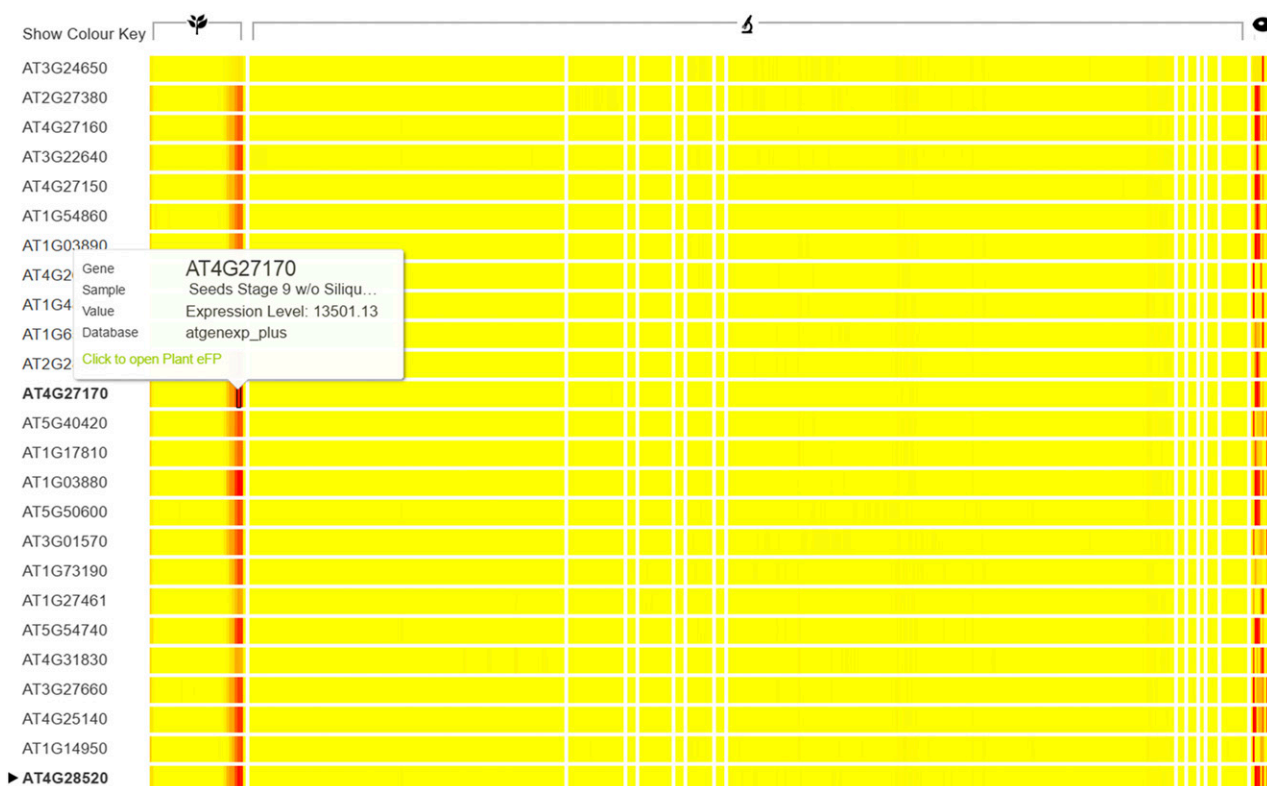


Figure 5. The Heat Map Viewer Showing 350+ Expression Level Samples for 25 Genes Identified with the Expression Angler for Having Similar Expression Patterns to *ABI3* (At3g24650).

The “global” color gradient is selected, making it easy to see the variability in the expression levels of the various genes.

(*EPR1*) with a coexpression coefficient r value of 0.979 and a maximum expression level of 8722 (also in the stage 9 seeds). Since *EPR1*'s maximum expression level is nearly 7 times higher, *ABI3*'s highest levels are lower on the color gradient; thus, the heat map cells for the stage 9 seeds appear almost yellow. ePlant also provides a "custom" color gradient setting so users can map colors to a user-defined threshold.

World eFP Viewer

The World eFP Viewer (Figure 2A) displays natural variation of gene expression levels from 34 ecotypes collected from different parts of the world but grown in a common chamber (Lempe et al., 2005). It draws pin markers (designed to look like a seedling because the data were collected from seedlings) that are colored according to the expression level for the selected gene for that given ecotype and placed according to the geographic coordinates of their source on a Google Maps image of the world. Climate data (annual precipitation, maximum temperature, and minimum temperature) is also projected onto the map using the Google Map API raster layer function. Combining ecotype expression data from the Weigel Lab (Lempe et al., 2005) with climate data from the World Bank Climate Portal (Harris et al., 2014) enables researchers to quickly see how ecotypes might differ in response to their environments.

This is an update of a similar tool originally included with the Arabidopsis eFP Browser (Winter et al., 2007). The original version used server-side image processing to draw a noninteractive chart, with little concern for data visualization best practices such as "details-on-demand" (Schneiderman, 1996) and "data/ink ratio" (Tufte and Graves-Morris, 1983). A simple task such as determining whether *ABI3* is up- or downregulated in arid climate regions took considerable effort as the answer required processing information from all over the screen. In the new version, this task can be answered with a single glance.

Plant eFP Viewer

The Plant eFP viewer (Figures 2B and 3B) displays the selected gene's expression pattern by dynamically coloring the tissues of a pictographic representation of a plant according to gene expression levels from multiple experiments. This visualization method is known as an electronic fluorescent pictograph (eFP), and it is a reimplement of the developmental map view of Arabidopsis eFP Browser by Winter et al. (2007).

Several new features have been added. For one, the chart has been redrawn as an SVG image (a vector image instead of a bitmap) and redundant black outlines have been omitted to improve the data/ink ratio (Tufte and Graves-Morris, 1983). Also, since SVG shapes can be filled with any color programmatically, it is now possible to adjust color gradients on the fly without having to redownload the image. This makes it possible to toggle between absolute and relative views with almost no latency between screen updates. Finally, expression patterns for several dozen genes occupy much less bandwidth than separate image files for each gene. This makes it possible to switch between eFP images for multiple genes at intervals of 150 ms or faster, a technique known as rapid serial visual presentation (RSVP), discussed later in this article.

Figures 2B and 3B show the spatio-temporal distribution of *ABI3* gene expression levels across the various developmental stages of Arabidopsis based on data from Schmid et al. (2005) and Nakabayashi et al. (2005). With a single glance, it is possible to see that *ABI3* has a narrow expression pattern that is limited to the maturing seeds.

Tissue and Experiment eFP Viewer

The Tissue and Experiment eFP Viewer (Figure 6) provides detail level information about gene expression in individual tissues and the results of perturbation response experiments. The 22 views in this ePlant module display information for 640 separate tissues based on 1385 samples. Multiplying that by the 22,814 genes on the ATH1 array (RNA-seq data are available in the Sequence Viewer module) produces a data set of 31,597,390 records into which each query taps. This is an example of how big data can be explored with a simple graphical interface.

Many of the views in this module are based on supplemental views that have been added since the original publication of the Arabidopsis eFP Browser (Winter et al., 2007). They have been updated in keeping with current data visualization best practices. Three new features have been added. First, a vertical stack of thumbnail images along the left side of the window provides a visual method for selecting the active view. Second, the thumbnail images can be sorted either alphabetically or by the maximum expression level for each of the views, making it easy to identify which tissues or experimental conditions are associated with high expression levels for the selected gene. Also, simply by glancing at the proportion of red or yellow in the stack of thumbnail images (a visualization technique sometimes referred to as "small multiples"; Tufte, 1990), it is possible to get a Gestalt sense of the expression pattern for the selected gene in various contexts without opening a single view (e.g., "Does my gene of interest have a narrow or wide expression pattern?"). Finally, the global color gradient option (discussed in the Heat Map Viewer section above) is especially useful here because it enables viewers to compare spatio-temporal, tissue-specific, and perturbation response expression levels all on the same scale (as in Figure 6). The views represent separate experiments, but they are easily comparable because the results are mapped to a common scale. This makes it possible to quickly answer the question, "In which tissue, and under what circumstance, does my gene of interest have the highest expression?"

As with all the tools in ePlant, the raw data used to generate the charts are available for download as a text file by clicking the "Download Raw Data" button on the toolbar. Table 1 provides a list of all the views available in the Experiment and Tissue eFP Viewer along with their data sources.

Subcellular Localization eFP Viewer

The Subcellular Localization eFP Viewer displays the documented and predicted localization of a gene product within the cell, with a color gradient representing a confidence score that the selected gene's protein product is found in a given compartment. The data for this module/view originate from SUBA3 database (Tanz et al., 2013) via a web service hosted by the BAR.

Like the other eFP viewers in ePlant, the Subcellular eFP Viewer is a reimplement of an earlier tool by Winter et al.

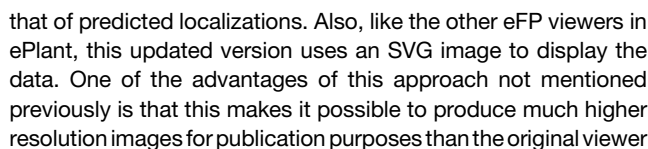


Table 1. Experimental and Tissue eFP Views and Data Sources

View	Data Source
Abiotic stress	Kilian et al. (2007)
Biotic stress: <i>Botrytis cinerea</i>	AtGenExpress initiative
Biotic stress: Elicitors	AtGenExpress initiative
Biotic stress: <i>Erysiphe orontii</i>	AtGenExpress initiative
Biotic stress: <i>Hyaloperonospora arabidopsidis</i>	Wang et al. (2011)
Biotic stress: <i>Myzus persicae</i>	Couldridge et al. (2007)
Biotic stress: <i>Phytophthora infestans</i>	AtGenExpress initiative
Biotic stress: <i>Pseudomonas syringae</i>	AtGenExpress initiative
Chemical	Goda et al. (2008)
Guard cell: meristemoids	Pillitteri et al. (2011)
Guard cell: mutant and wild-type guard cell ABA response	Pandey et al. (2010)
Guard cell: suspension cell ABA response with ROS scavenger	Böhmer and Schroeder (2011)
Tissue specific: embryo development	Casson et al. (2005)
Tissue specific: guard and mesophyll cells	Yang et al. (2008)
Tissue specific: microgametogenesis	Honys and Twell (2004)
Tissue specific: pollen germination	Qin et al. (2009)
Tissue specific: root	Birnbaum et al. (2003); Naway et al. (2005)
Tissue specific: shoot apical meristem	Yadav et al. (2009)
Tissue specific: stem epidermis	Suh et al. (2005)
Tissue specific: stigma and ovaries	Swanson et al. (2005)
Tissue specific: trichomes	Gilding and Marks (2010); Marks et al. (2009)
Tissue specific: xylem and cork	NASCArrays experiment #92

could generate. Figure 7A shows a screen grab of the Subcellular eFP view for *ABI3*, while Figure 7B demonstrates how the content can be scaled without image degradation. An example SVG file downloaded from the Plant eFP Viewer for *ABI3* has been included as Supplemental File 1. Such files may be easily used in any vector graphics program to generate high-quality figures. ePlant outputs are freely available under an “open” Creative Commons Attribution license (CC-BY version 4.0).

Chromosome Viewer

The Chromosome Viewer (Figure 2E) provides a pictographic overview of the plant’s chromosomes as a series of vertical bars with markers indicating the positions of all the genes that have been downloaded. Spatial relationships within the genome can sometimes indicate functional relationships (Chae et al., 2014; Wisecaver et al., 2017). This viewer can be used to quickly determine the physical location of coexpressed genes, for instance.

Clicking on the chromosomes opens a menu listing all the genes at the location that was selected. Since each chromosome contains several thousand genes but the display panel height is typically <700 pixels, each pixel represents the location of several genes. This limits the practicality of using this feature as a gene selection method. However, several users reported during user testing that they appreciated how it conveys the sheer number of genes in the genome. Clicking the “thermometer” icon in the toolbar generates a heat map indicating

the density of genes within the chromosome. This makes it easy to see if the selected gene is located in a gene-rich region. An annotation tool (accessed by clicking the “pencil” icon) allows users to adjust label colors and sizes in order to make custom charts.

Protein and DNA Interaction Viewer

The Protein and DNA Interaction viewer displays documented and predicted protein-protein interactions (PPIs) and protein-DNA interactions (PDIs) for the selected gene. It uses a node-link charting method, in which the nodes represent proteins or DNA sequences and the links represent interactions between these elements (Figure 8A). This module is a reimplement of the Arabidopsis Interactions Viewer at the BAR (Geisler-Lee et al., 2007). Several features have been added and the interface has been modified to improve usability.

The design of the chart takes advantage of preattentive visual processing (Healey and Enns, 2012) to help users explore multiple levels of data in the same window. DNA elements are drawn as squares and have curved lines to indicate interactions with other proteins. Protein elements are drawn as circles and have straight lines to indicate interactions with other proteins. Interaction line thickness is determined by the interaction confidence value. Line colors are determined by the coexpression coefficient with a yellow-to-red scale. Interactions that have been experimentally determined are drawn with green lines, and clicking them opens a window with the associated paper on PubMed. The borders of protein nodes are colored according to where each protein is localized within a cell. For instance, a blue border indicates a protein that is mostly found in the nucleus, and an orange border indicates a protein that it is mostly found in the plasma membrane. This makes it possible to quickly answer the question, “Does my gene of interest mostly interact with proteins in the same cell compartment, or across several compartments?” DNA nodes have black borders because subcellular localization data do not apply to them.

The center color of each node changes from light gray to dark gray when the data for that gene have been downloaded. This makes it easy to see if a set of downloaded genes are also interaction partners. Previously, answering this question would require a multistep list collation process. It can now be done with a glance, simply looking for multiple dark gray node centers.

Hovering over the various nodes will open a popup box with annotation information for the gene and a “Get Data” button that downloads all the data for that gene into ePlant. This makes it possible to “surf” from one gene to another and explore ideas on a whim. Researchers may not initially know which genes to load, but loading one gene could take them on a journey that links to a whole set of genes.

In Arabidopsis, interacting proteins have an average of 11 interacting partners (Geisler-Lee et al., 2007) but some genes, such as At4g26840, have as many as 172. Drawing that many nodes and links would produce a “hairball” that cannot be easily deciphered. To accommodate these cases, we added a data filtering function that permits users to hide interactions with confidence or correlation values below a customizable threshold.

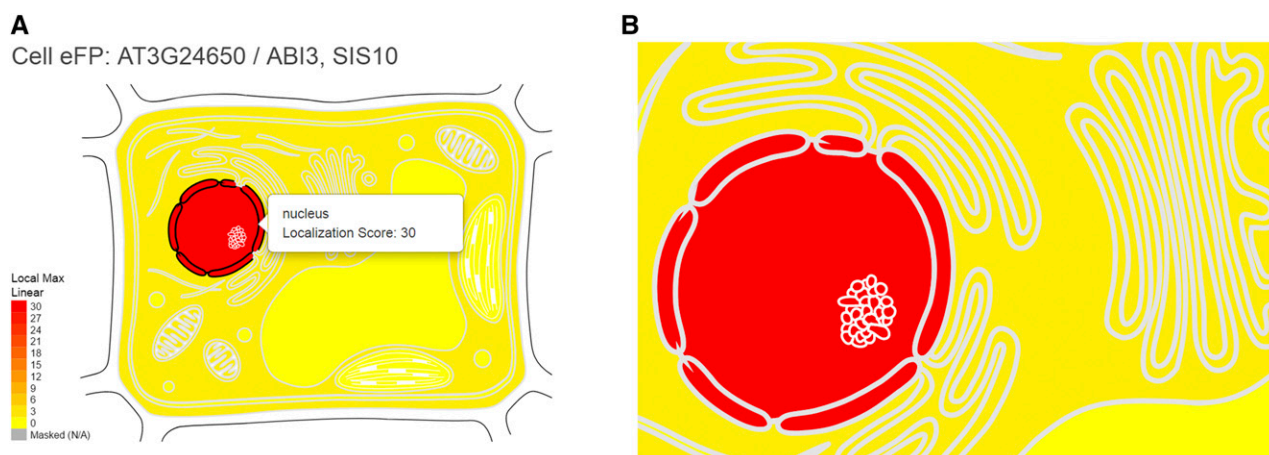


Figure 7. The Subcellular Localization eFP Viewer.

(A) ABI3 is mostly localized in the nucleus.

(B) An inset of a high-resolution version of the same image.

It is also possible to hide either experimentally determined or predicted PPIs and PDIs.

The module was built with JavaScript using Cytoscape.js (<http://js.cytoscape.org>), an open-source library for biological network analysis and visualization that is the successor to Cytoscape Web (Lopes et al., 2010). The PPI data come from a database of 70,944 predicted Arabidopsis interacting proteins generated by Geisler-Lee et al. (2007) and 36,306 confirmed interaction proteins from the Biomolecular Interaction Network Database (Bader et al., 2003), high-density Arabidopsis protein microarrays (Popescu et al., 2007, 2009), Braun et al.'s Arabidopsis Interactome (Arabidopsis Interactome Mapping Consortium, 2011), Wolf Frommer's Membrane protein Interactome Database MIND (Lalonde et al., 2010), and over 1190 other literature sources. The PDI data come from a database of 1784 confirmed interactions generated by Taylor-Teeple et al. (2015) and DAP-seq data generated by the Ecker lab (O'Malley et al., 2016), which the authors show to be quite similar to ChIP-seq data in terms of quality, while encompassing a far greater number of transcription factors (there are just ~200 Arabidopsis ChIP-seq experiments in GEO). All data are downloaded from a web service at the BAR. The interactions in BIND and from other sources were identified using several different methods, such as yeast two-hybrid screens, but also via traditional biochemical methods. Subcellular localization data is from SUBA, the Arabidopsis Subcellular Database (Tanz et al., 2013).

Figure 8A shows a diagram of protein and DNA interaction partners for At1g54330. This is a good example of how combining multiple levels of data into the same chart can improve systems biology workflows and deepen our understanding of biological functioning, especially in the area of gene regulatory networks.

Molecule Viewer

The Molecule Viewer maps information from four separate databases onto a 3D model of the selected protein's molecular structure. The 3D model (structure models have been computed

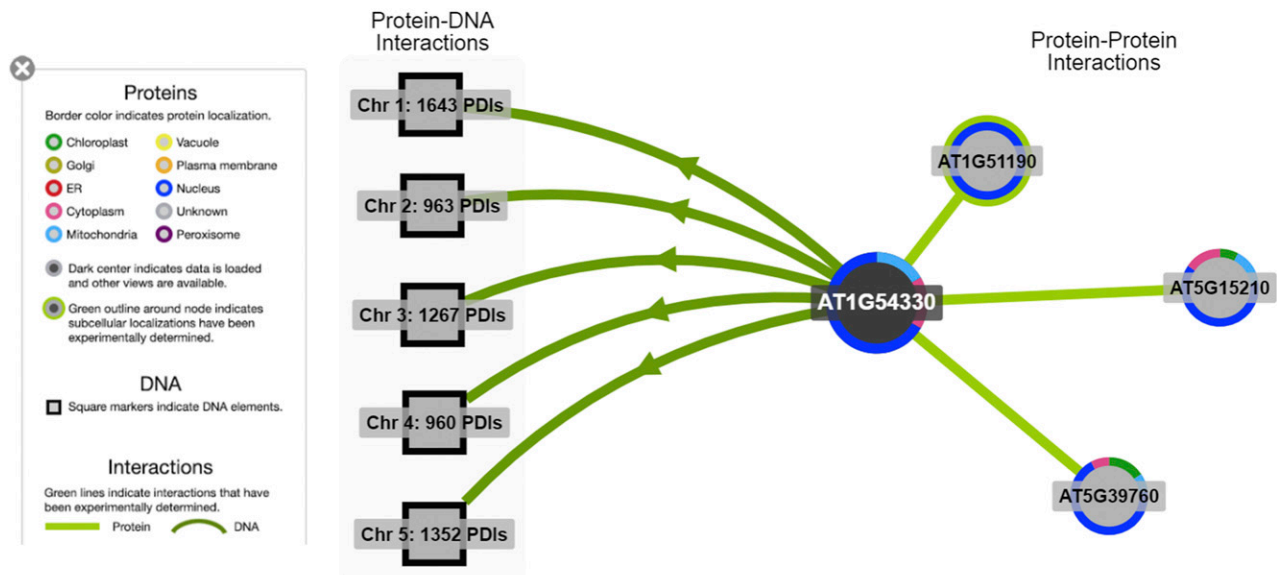
for 23,091 Arabidopsis gene products as part of the ePlant effort) comes from Phyre2 (Kelley et al., 2015) and data layers include (1) complete protein sequences from Araport (Krishnakumar et al., 2015); (2) nonsynonymous SNP locations in the underlying gene sequence from the 1001 Proteomes project (Joshi et al., 2011) with a list of which ecotypes in which they are found; (3) Pfam domains (Finn et al., 2014); and (4) CDD feature hits (Marchler-Bauer et al., 2015). Drawing this information onto the 3D molecular structure enables researchers to visualize exactly where in the physical model of the protein such features exist. This makes it possible to easily answer the question, "Is there a polymorphism causing a nonsynonymous amino acid change near the DNA binding site of my favorite transcription factor that might affect its binding to a *cis*-element?", as shown in Figure 8B.

The PDB file is displayed with JSmol (Hanson et al., 2013). The protein sequence is drawn on the bottom of the screen along with pin markers that indicate the position and frequency of SNP locations. A sliding window enables users to control which part of the sequence they are looking at. Hovering the mouse over the protein sequence highlights the associated location on the 3D model, and hovering the mouse over the 3D model highlights the associated location within the sequence. This enables users to quickly identify which parts of the protein sequence are exposed versus found in the interior parts of the model, just by moving the mouse over the content. The location of a non-synonymous SNP, Pfam domain, or CDD feature could have a very large influence on the behavior of the molecule, and this application of mouse-over indicators makes it very easy to find them.

Sequence Viewer

In many ways, sequence browsers were the first ZUIs for bioinformatics since they enable micro-to-macro level exploration of data, providing detail and overview level information at the same time. They also facilitate comparison of multiple levels of data from a variety of sources (e.g., methylation, phosphorylation, SNPs,

A
Interaction Viewer: AT1G54330 / ANAC020, NAC020
Recursive interactions not shown



B
Molecule Viewer: AT3G24650 / ABI3, SIS10

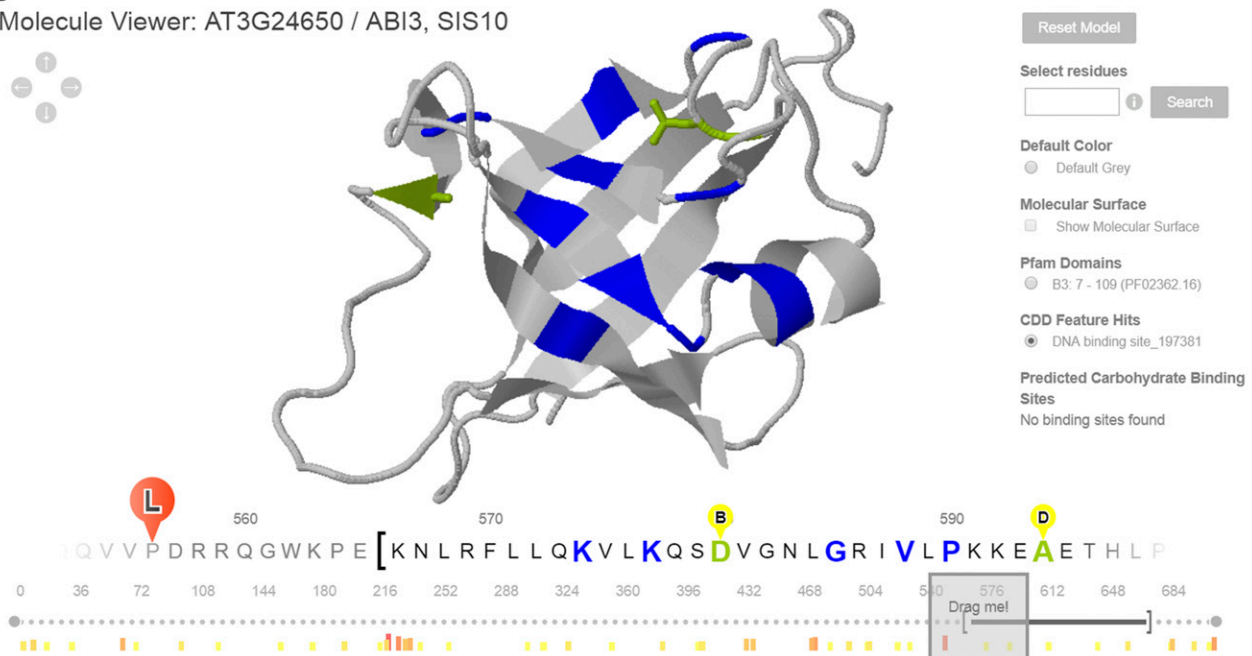


Figure 8. ePlant Interaction and Molecule Viewers.

(A) Protein and DNA Interaction Viewer showing interactions for At1g54330.
(B) Molecule Viewer showing the transcription factor ABI3's Phyre2-predicted partial 3D structure with its DNA binding site highlighted in blue and two nonsynonymous changes (from a web service provided by the 1001 Proteomes site) highlighted in green. Changes of higher frequency are denoted by larger, redder pins above the sequence below.

conserved noncoding regions, etc.) by mapping each layer onto the chart as a separate track.

This module (Figure 2H) uses an implementation of JBrowse (Skinner et al., 2009) using data provided by web services at Araport (Krishnakumar et al., 2015). Due to the complexity of the program, we did not apply the ePlant style guide to this tool so there is some perceptual difference between this and the other ePlant modules. To create more usable screen real estate, the gene panel that occupies the left side of the screen can be slid out of the way by clicking the triangular toggle button at the top of the navigation stack. The Sequence Viewer permits more than a dozen nucleotide-resolution data types (RNA-seq data, conserved noncoding regions, chromatin states, methylation data, noncoding RNAs, and others) to be further explored within ePlant.

Links to External Tools

There are many more data visualization modules we would have liked to include with this version of ePlant, but could not for various reasons. The Links to External Tools module contains a list of dynamic links to automatically open the ThaleMine at Araport, TAIR, GeneMANIA, Expressologs, and SeedNet pages for the currently selected gene. While this is not ideal from an integrative tool perspective, it does save many clicks and reduces the inconvenience of having to navigate between sites. This module is easy to update and we plan to add more links in the near future.

Additional ePlant Features

Expression Angler

Researchers might not come to ePlant with a priori knowledge of which genes they wish to explore. The Expression Angler tool, an implementation of the tool described by Austin et al. (2016), helps users identify and download Arabidopsis genes by their expression pattern instead of by name. It does this by calculating the correlation coefficients for expression for all gene expression vectors compared with an expression pattern that the user defines or to the expression pattern associated with a single AGI ID or gene name that a user enters (Toufighi et al., 2005). For example, researchers who are interested in exploring the mechanisms associated with seed development may use the Expression Angler to search for genes with high transcript levels in early stage seeds but not in any other tissue. Alternatively, they may use the Expression Angler to find the top 25 genes with similar expression patterns as *ABI3* (as depicted in Figure 5). The tool can be accessed via the button under the gene input box in the upper left corner of the screen.

Mutant Phenotype Gene Selector

This tool provides two approaches for helping users identify genes associated with loss-of-function mutant phenotypes in Arabidopsis: *Search by Classification* and *Search by Data Table*. It is based on a literature curation effort by Lloyd and Meinke (2012) that includes a database of 2400 Arabidopsis genes with a documented loss-of-function mutant phenotype as well as a proposed schema for categorizing them. The search by classification method

uses an interactive Reingold-Tilford tree selection method, implemented with d3.js (<https://d3js.org/>). The search by data table method was built with DataTables (<https://datatables.net/>).

Rapid Serial Visual Presentation

Identifying genes of interest from a large set of eFP images can be a daunting task. To succeed, researchers must find subtle differences between multiple nearly identical images. The RSVP display technique has the potential to improve the experience as it exploits our ability to recognize differences between images when they are displayed on a screen in a rapid and serial manner. The technique is known to be an efficient way to find the presence or absence of a specific item within a set of images (Spence, 2002), akin to flipping through a book to find a specific picture.

The “Slide Show” RSVP feature, accessed near the top of the gene panel, automatically advances the currently selected gene every 250 ms. Upon reaching the bottom of the list, it cycles back to the top. The “Hover” RSVP feature enables users to hover their mouse over the gene panel to adjust the currently selected gene. Moving the mouse up and down over the list produces a “user-controlled” RSVP effect. We have shown through controlled user testing that both of these methods are more efficient than “point and click” when it comes to selecting genes of interest from a set of eFP images (Waese et al., 2016).

DISCUSSION

ePlant permits researchers to easily see where and when a gene is “active,” how its protein product can fold into a molecular machine to do what it needs to do, and whether there are any natural genetic variants of the gene that might allow it to do it better. It is an open-source visual analytic platform that was designed to help plant researchers seamlessly explore data from different biological levels through a single window. It uses a ZUI that enables users to quickly transition from natural variation at the kilometer scale, through gene expression levels in tissues and cell types, sub-cellular localization of gene products, protein-protein and protein-DNA interactors, to protein tertiary structure and gene sequences at the nanometer scale.

Integrating data from different biological levels can allow novel hypotheses to be generated. By combining data from several biological levels of analysis into the same view, ePlant makes it possible to easily examine protein-protein interactions and ask whether these protein products are in the same compartment, what the tertiary structure of a protein product might be, and whether there are any polymorphisms that lie close to structurally important features, like DNA binding sites. Adams et al. (2017) have shown that structural clustering of variation can predict functional sites in proteins. Our lab is interested in natural variation in ABA signaling, and an analysis of ABA-related bZIP transcription factors (*ABSCISIC ACID-RESPONSIVE ELEMENT BINDING FACTORS1* through 4 and *ABI5*, collapsed to a consensus sequence) shows that there are a few frequent variants close to the DNA binding site (Supplemental Figure 1). The representation of variant frequency across ecotypes by pin size in the ePlant Molecule Viewer is also helping us prioritize on which variants to focus our analysis. There are many nonsynonymous

SNPs that occur in just one ecotype that we hypothesize to be of less functional importance than those that are found in several ecotypes. This sort of analysis was not possible with the prototypic ePlant interface released several years ago (Fucile et al., 2011). Not having to switch windows and contend with several different user interfaces to explore an idea from multiple perspectives should liberate researchers and make it easier to stay “on task” and make creative associations without distraction.

ePlant was developed with regard to best practices for user experience design and data visualization, as well as with feedback gathered from two rounds of user testing (see Methods). We have successfully deployed ePlant on the new international portal for Arabidopsis information, Araport.org (Krishnakumar et al., 2015). This is a large collaborative effort that demonstrates the power of a federated web service-based approach in integrating and visualizing data from multiple sources, as articulated by the International Arabidopsis Informatics Consortium (2012). We have made the project open source, such that other groups may develop modules for ePlant as new data types become available and new linkages between different levels of data are discovered.

We have received funding from Genome Canada to leverage the ePlant framework to create 15 ePlants for agronomically important species including tomato, maize, wheat, and soybean. Here, a novel “navigator” will be developed to readily permit the exploration of homologous sequences and their associated transcriptomic, proteomics, structural, and other data. This framework would be highly useful to improve crop species, and being able to efficiently query and visualize the huge amount of data generated in the past 5 years will be key to improving and managing these crops to feed, shelter, and power a world of 9 billion people by the year 2050. By adding multiple species to the framework (through a pipeline that is also being developed as part of this grant), it will be possible to see if nonsynonymous changes map to the same location in one protein’s structure as do nonsynonymous changes in another species for a homologous gene. If that is the case, then the likelihood of that polymorphism being biologically relevant would be high. Other powerful research-driven questions that would be possible to ask with this interface include, “Which homolog of my gene of interest in the species I work on has the same expression pattern in equivalent tissues as in the reference species?” These kinds of questions are relevant for translational biology, that is, for extending the information and knowledge derived from a reference species into agronomically important ones. In principle, the benefits of our systems approach extend to any species with available genomic sequences.

METHODS

System Architecture

The system architecture (Figure 3A) can be divided into five categories: (1) databases, (2) web services, (3) data processing functions; (4) ZUI framework, and (5) data visualization modules. ePlant aggregates data from numerous sources, most of which are stored in SQL databases on servers hosted by Araport, TAIR, or our own BAR (Toufighi et al., 2005). The actual data are accessed via web services (typically served up by Perl or Python CGI scripts) hosted on the same server as the databases. When a user selects a gene to download, ePlant sends a batch of queries to each of the web services associated with the various data visualization modules. They return several file formats depending on the nature of the data: JSON

objects, XML files, and prerendered PNG images (in the case of the Gene Cloud images; Krouk et al., 2015). Databases and web services (i.e., categories one and two in Figure 3A) are server side constructs and can be considered “back-end” components of ePlant.

Often, data must be reformatted and processed before they can be visualized. Hard coding the myriad number of data permutations (i.e., category three in Figure 3A) directly into the visualization modules would be difficult to maintain if the data format changed or if new data sources become available. Thus, although data processing functions are executed locally on the client machine, they are separate elements within the system architecture. These functions can be considered a “middle layer.”

On the “front-end,” ePlant maintains separate functions for ZUI management and data visualization (i.e., categories four and five in Figure 3A). To maximize interface responsiveness, these functions are executed locally on the client machine. The ZUI framework is responsible for drawing interface elements to the screen and triggering various functions in response to user input. The data visualization modules consist of separate programs that are initialized when data becomes available and then run in the background, waiting for the ZUI framework to make their screen visible.

For the data visualization modules that are based on HTML5 canvas, the ZUI framework treats each module as a separate <div> and simply animates the scale and visibility of that element. For the other modules, the zoom transitions are built into the views themselves. For example, the eFP viewers use CSS transitions defined within their own function scopes, and the Molecule Viewer calls the JSmol library to resize the 3D molecule model.

ePlant was written to be easily expandable. Adding new data visualization modules is a simple matter of adding the necessary data loading and visualization programs to the host directory, adding citation and data source information, and adding an icon to the ZUI navigation panel. The code is well documented and available as open source code on GitHub for anyone to explore or fork and modify (see below).

User Testing

To ensure the relevance and usefulness of ePlant for its intended users, we adopted an “agile” approach to software design (Highsmith and Cockburn, 2001), a process that includes frequent user testing, analysis of user needs, prototyping, and refinement. As part of that process, we conducted two rounds of user testing at the 2014 and 2015 International Conference on Arabidopsis Research (ICAR). Attendees were invited to follow a user testing protocol based on Nielsen’s guidelines for usability engineering (Nielsen, 1993) and (Hudson, 2014) that consisted of three phases: free exploration of the tool, completion of 10 sample tasks, and a Google Forms questionnaire. The protocol was approved by the University of Toronto Research Ethics Board (protocol no. 30490).

Participants were recorded with Screencast-O-Matic software as they interacted with the tool. All mouse clicks and verbal comments were recorded, and participants were asked to “think out loud” so we could collect qualitative feedback about interactions as they happened. Performance time was measured for typical tasks such as: In which tissue is ABI3 most strongly expressed? Where is ABI3 localized in the cell? Can you name an interaction partner for ABI3? In what part of the world does AT1G16850 show the most natural variation of expression? What is the annotation for ATAP3? The same tasks were presented in the same order across both user testing sessions. Tasks that could not be answered by the majority of participants in 20 seconds or less were flagged for additional development.

At the 2014 ICAR in Vancouver, 13 participants completed the study (four professors, two post docs, five PhD candidates, and two industry researchers). At the 2015 ICAR in Paris, 18 participants completed the study (eight professors, one post doc, three PhD candidates, three industry researchers, and three undeclared). This may not seem like a large number, but Nielsen and Landauer (1993) found that the typical user testing session will identify 31% of all the usability problems in a design and that 85% of a site’s problems can be found with as few as five participants.

After using the tool for about 10 min, participants were asked to complete a Google Forms questionnaire with a seven-point Likert scale according to the following: (1) Please rate the quality of ePlant's user interface; (2) Please rate how useful ePlant is for Arabidopsis researchers; (3) How would you describe the depth of information contained in ePlant?; (4) How would you compare ePlant against current methods for accessing the same information?; (5) How likely are you to use ePlant in a research project?; (6) How would you describe the depth of information contained in ePlant?; (7) How likely are you to use ePlant again?; and (8) Please rate your overall user experience of using ePlant.

Responses across both years were positive. In 2015, almost all participants responded with the most positive response for "How likely are you to use ePlant in a research project?" and "How likely are you to use ePlant again?" These are essentially the same question, and the response suggests that ePlant successfully delivers on the objective to build a research platform that plant biologists want to use.

Quantitative data provide a snapshot of the efficacy of the tool; however, the main value from user testing is found in the qualitative data that were collected. Notes taken while coding the screencasts produced a total of 88 feature requests, bug reports, interface modifications, and other suggestions on how to improve the final tool. These notes were entered into an issue-tracking platform called Pivotal Tracker that allows tasks to be sorted according to difficulty and assigned to individual programmers to work on. At this time, virtually all of the tasks have been addressed and/or implemented.

Implementation on Araport

ePlant was initially written as a standalone program for the Bio-Analytic Resource for Plant Biology (<http://bar.utoronto.ca/>). It has been deployed as a science app on Araport, accessible from Araport's front page at Araport.org. Using Araport's Yeoman-based application scaffold (called `aip-science-app`), ePlant front-end code was ported and integrated into the Araport science app framework. Two multipoint pass-through Araport Data And Microservices API (ADAMA) adapters, "eplant_service" and "expression_angler_service," were developed to retrieve data from the BAR web services. These adapters, hosted on a public GitHub repository (https://github.com/BioAnalyticResource/Araport_ePlant_ADAMA), were registered with Araport as community API adapters. Araport users may try out these adapters at <https://www.araport.org/api-explorer> after signing into Araport or via the command line using an Araport OAuth 2.0 access token (<https://www.araport.org/docs/building-community-apis-adama/getting-token>). The JavaScript code of ePlant was modified to load data using these ADAMA adapters. The modified code is hosted at this public GitHub repository: https://github.com/BioAnalyticResource/Araport_ePlant. Users may run ePlant on their computers using Araport's text environment built with Grunt and Node.js. They may also deploy the ePlant app into their own Araport workspaces.

Supplemental Data

Supplemental Figure 1. Analysis of nsSNPs near ABF1's DNA binding site performed using ePlant or across ABA-related bZIP transcription factors using data from the 1001 Proteomes site.

Supplemental Table 1. A list of web services used by ePlant to populate the data various data visualization modules.

Supplemental File 1. An example SVG file downloaded from the Plant eFP Viewer for *ABI3*.

ACKNOWLEDGMENTS

This project was funded by the Government of Canada through Genome Canada/Ontario Genomics (OGI-071 to N.J.P.). M.J.S. and L.A.K. were

supported UK BBSRC Grant BB/M011526/1. We thank the reviewers for helpful suggestions.

AUTHOR CONTRIBUTIONS

J.W. conceived, programmed, and coordinated development of ePlant. J.F. programmed ePlant. A.P. developed web services for ePlant and deployed ePlant on Araport. H.Y. programmed initial ePlant framework. G.F. developed components of ePlant's Molecule Viewer. R.S. developed ePlant's Interaction Viewer. M.C. performed nonsynonymous SNP analysis. L.A.K. and M.J.S. provided Phyre2 structure-ome. V.K., E.F., J.M., and C.T. developed Araport components for the ePlant app. W.S. provided input on ePlant user testing and user interface. N.J.P. conceived and oversaw overall ePlant project. J.W. and N.J.P. wrote most of the manuscript, with a small amount of additional material provided by some of the coauthors.

Received January 27, 2017; revised July 17, 2017; accepted August 9, 2017; published August 14, 2017.

REFERENCES

- Adams, J., Mansfield, M.J., Richard, D.J., and Doxey, A.C. (2017). Lineage-specific mutational clustering in protein structures predicts evolutionary shifts in function. *Bioinformatics* **33**: 1338–1345.
- Arabidopsis Interactome Mapping Consortium. (2011). Evidence for network evolution in an Arabidopsis interactome map. *Science* **333**: 601–607.
- Austin, R.S., et al. (2016). New BAR tools for mining expression data and exploring *cis*-elements in *Arabidopsis thaliana*. *Plant J.* **88**: 490–504.
- Bader, G.D., Betel, D., and Hogue, C.W.V. (2003). BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31**: 248–250.
- Bederson, B.B. (2010). The promise of zoomable user interfaces. In *Proceedings of the 3rd International Symposium on Visual Information Communication*. (New York: ACM), doi/10.1145/1865841.1865844.
- Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genes* **53**: 474–485.
- Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. *Science* **302**: 1956–1960.
- Böhmer, M., and Schroeder, J.I. (2011). Quantitative transcriptomic analysis of abscisic acid-induced and reactive oxygen species-dependent expression changes and proteomic profiling in Arabidopsis suspension cells. *Plant J.* **67**: 105–118.
- Casson, S., Spencer, M., Walker, K., and Lindsey, K. (2005). Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. *Plant J.* **42**: 111–123.
- Chae, L., Kim, T., Nilo-Poyanco, R., and Rhee, S.Y. (2014). Genomic signatures of specialized metabolism in plants. *Science* **344**: 510–513.
- Chatr-Aryamontri, A., et al. (2015). The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **43**: D470–D478.
- Couldridge, C., Newbury, H.J., Ford-Lloyd, B., Bale, J., and Pritchard, J. (2007). Exploring plant responses to aphid feeding using a full Arabidopsis microarray reveals a small number of genes with significantly altered expression. *Bull. Entomol. Res.* **97**: 523–532.
- Finn, R.D., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* **42**: D222–D230.

- Fucile, G., Di Biase, D., Nahal, H., La, G., Khodabandeh, S., Chen, Y., Easley, K., Christendat, D., Kelley, L., and Provart, N.J. (2011). ePlant and the 3D data display initiative: integrative systems biology on the world wide web. *PLoS One* **6**: e15237.
- Geisler-Lee, J., O'Toole, N., Ammar, R., Provart, N.J., Millar, A.H., and Geisler, M. (2007). A predicted interactome for Arabidopsis. *Plant Physiol.* **145**: 317–329.
- Gilding, E.K., and Marks, M.D. (2010). Analysis of purified glabra3-shapeshifter trichomes reveals a role for NOECK in regulating early trichome morphogenic events. *Plant J.* **64**: 304–317.
- Goda, H., et al. (2008). The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.* **55**: 526–542.
- Hanson, R.M., Prilusky, J., Renjian, Z., Nakane, T., and Sussman, J.L. (2013). JSmol and the Next-Generation Web-Based Representation of 3D Molecular Structure as Applied to Proteopedia. *Isr. J. Chem.* **53**: 207–216.
- Harris, I., Jones, P.d., Osborn, T.j., and Lister, D.h. (2014). Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**: 623–642.
- Healey, C.G., and Enns, J.T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Trans. Vis. Comput. Graph.* **18**: 1170–1188.
- Helt, G.A., Nicol, J.W., Erwin, E., Blossom, E., Blanchard, S.G., Jr., Chervitz, S.A., Harmon, C., and Loraine, A.E. (2009). Genoviz Software Development Kit: Java tool kit for building genomics visualization applications. *BMC Bioinformatics* **10**: 266.
- Highsmith, J., and Cockburn, A. (2001). Agile software development: the business of innovation. *Computer* **34**: 120–127.
- Hony, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. *Genome Biol.* **5**: R85.
- Hudson, W. (2014). Agile UX and UCD. In CHI '14 Extended Abstracts on Human Factors in Computing Systems. (New York: ACM), pp. 1009–1010.
- International Arabidopsis Informatics Consortium (2012). Taking the next step: building an Arabidopsis information portal. *Plant Cell* **24**: 2248–2256.
- Joshi, H.J., et al. (2011). MASCIP Gator: an aggregation portal for the visualization of Arabidopsis proteomics data. *Plant Physiol.* **155**: 259–270.
- Joshi-Tope, G., et al. (2005). Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**: D428–D432.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**: 845–858.
- Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* **50**: 347–363.
- Klein, C., and Bederson, B.B. (2005). Benefits of animated scrolling. In CHI '05 Extended Abstracts on Human Factors in Computing Systems. (New York: ACM), pp. 1965–1968.
- Krishnakumar, V., et al. (2015). Araport: the Arabidopsis information portal. *Nucleic Acids Res.* **43**: D1003–D1009.
- Krouk, G., Carré, C., Fizames, C., Gojon, A., Ruffel, S., and Lacombe, B. (2015). GeneCloud reveals semantic enrichment in lists of gene descriptions. *Mol. Plant* **8**: 971–973.
- Lalonde, S., et al. (2010). A membrane protein/signaling protein interaction network for Arabidopsis version AMPv2. *Front. Physiol.* **1**: 24.
- Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M., and Weigel, D. (2005). Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genet.* **1**: 109–118.
- Lloyd, J., and Meinke, D. (2012). A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant Physiol.* **158**: 1115–1129.
- Lopes, C.T., Franz, M., Kazi, F., Donaldson, S.L., Morris, Q., and Bader, G.D. (2010). Cytoscape Web: an interactive web-based network browser. *Bioinformatics* **26**: 2347–2348.
- Marchler-Bauer, A., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**: D222–D226.
- Marks, M.D., Wenger, J.P., Gilding, E., Jilk, R., and Dixon, R.A. (2009). Transcriptome analysis of Arabidopsis wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. *Mol. Plant* **2**: 803–822.
- Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.* **41**: 697–709.
- Nawy, T., Lee, J.-Y., Colinas, J., Wang, J.Y., Thongrod, S.C., Malamy, J.E., Birnbaum, K., and Benfey, P.N. (2005). Transcriptional profile of the Arabidopsis root quiescent center. *Plant Cell* **17**: 1908–1925.
- Nielsen, J. (1993). Usability Engineering. (San Francisco, CA: Morgan Kaufmann Publishers).
- Nielson, J., and Landauer, T. (1993). A mathematical model of the finding of usability problems. In Proceedings of the ACM INTERCHI93 Conference, April 24–29, 1993 (Amsterdam: IOS Press), pp. 206–213.
- O'Malley, R.C., Huang, S.S., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**: 1280–1292.
- Pandey, S., Wang, R.-S., Wilson, L., Li, S., Zhao, Z., Gookin, T.E., Assmann, S.M., and Albert, R. (2010). Boolean modeling of transcriptome data reveals novel modes of heterotrimeric G-protein action. *Mol. Syst. Biol.* **6**: 372.
- Pillitteri, L.J., Peterson, K.M., Horst, R.J., and Torii, K.U. (2011). Molecular profiling of stomatal meristems reveals new component of asymmetric cell division and commonalities among stem cell populations in Arabidopsis. *Plant Cell* **23**: 3260–3275.
- Popescu, S.C., Snyder, M., and Dinesh-Kumar, S. (2007). Arabidopsis protein microarrays for the high-throughput identification of protein-protein interactions. *Plant Signal. Behav.* **2**: 416–420.
- Popescu, S.C., Popescu, G.V., Bachan, S., Zhang, Z., Gerstein, M., Snyder, M., and Dinesh-Kumar, S.P. (2009). MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes Dev.* **23**: 80–92.
- Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M.A., and Palanivelu, R. (2009). Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet.* **5**: e1000621.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**: 501–506.
- Schneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In Proceedings of the 1996 IEEE Symposium on Visual Languages (Los Alamitos, CA: IEEE Computer Society Press), pp. 336–343.
- Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J., and Holmes, I.H. (2009). JBrowse: a next-generation genome browser. *Genome Res.* **19**: 1630–1638.
- Speier, C., Vessey, I., and Valacich, J.S. (2003). The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. *Decis. Sci.* **34**: 771–797.

- Spence, R.** (2002). Rapid, serial and visual: A presentation technique with potential. *Inf. Vis.* **1**: 13–19.
- Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F.** (2005). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. *Plant Physiol.* **139**: 1649–1665.
- Swanson, R., Clark, T., and Preuss, D.** (2005). Expression profiling of Arabidopsis stigma tissue identifies stigma-specific genes. *Sex. Plant Reprod.* **18**: 163–171.
- Tanz, S.K., Castleden, I., Hooper, C.M., Vacher, M., Small, I., and Millar, H.A.** (2013). SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis. *Nucleic Acids Res.* **41**: D1185–D1191.
- Taylor-Teeple, M., et al.** (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **517**: 571–575.
- Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provart, N.J.** (2005). The botany array resource: e-Northern, expression angling, and promoter analyses. *Plant J.* **43**: 153–163.
- Tufte, E.R.** (1990). *Envisioning Information*. (Cheshire, CT: Graphics Press).
- Tufte, E.R., and Graves-Morris, P.R.** (1983). *The Visual Display of Quantitative Information*. (Cheshire, CT: Graphics Press).
- Waese, J., Stuerzlinger, W., and Provart, N.J.** (2016). An evaluation of interaction methods for controlling RSVP displays in visual search tasks. In *Big Data Visual Analytics (BDVA) 2016* (Sydney, Australia: IEEE Xplore), doi/10.1109/BDVA.2016.7787041.
- Wang, W., Barnaby, J.Y., Tada, Y., Li, H., Tör, M., Caldelari, D., Lee, D.U., Fu, X.-D., and Dong, X.** (2011). Timing of plant immune responses by a central circadian regulator. *Nature* **470**: 110–114.
- Ware, C.** (2012). *Information Visualization: Perception for Design*. (Elsevier).
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J.** (2007). An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**: e718.
- Wisecaver, J.H., Borowsky, A.T., Tzin, V., Jander, G., Kliebenstein, D.J., and Rokas, A.** (2017). A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* **29**: 944–959.
- Yadav, R.K., Girke, T., Pasala, S., Xie, M., and Reddy, G.V.** (2009). Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. *Proc. Natl. Acad. Sci. USA* **106**: 4941–4946.
- Yang, Y., Costa, A., Leonhardt, N., Siegel, R.S., and Schroeder, J.I.** (2008). Isolation of a strong Arabidopsis guard cell promoter and its potential as a research tool. *Plant Methods* **4**: 6.