UNIT 1.11

Using The Arabidopsis Information Resource (TAIR) to Find Information About Arabidopsis Genes

Leonore Reiser¹, Shabari Subramaniam¹, Donghui Li¹, and Eva Huala¹
¹Phoenix Bioinformatics, Redwood City, CA USA

ABSTRACT

The Arabidopsis Information Resource (TAIR; http://arabidopsis.org) is a comprehensive Web resource of Arabidopsis biology for plant scientists. TAIR curates and integrates information about genes, proteins, gene function, orthologs gene expression, mutant phenotypes, biological materials such as clones and seed stocks, genetic markers, genetic and physical maps, genome organization, images of mutant plants, protein sub-cellular localizations, publications, and the research community. The various data types are extensively interconnected and can be accessed through a variety of Web-based search and display tools. This unit primarily focuses on some basic methods for searching, browsing, visualizing, and analyzing information about Arabidopsis genes and genome, Additionally we describe how members of the community can share data using TAIR’s Online Annotation Submission Tool (TOAST), in order to make their published research more accessible and visible.

Keywords: Arabidopsis ● databases ● bioinformatics ● data mining ● genomics

INTRODUCTION

The Arabidopsis Information Resource (TAIR; http://arabidopsis.org) is a comprehensive Web resource for the biology of Arabidopsis thaliana (Huala et al., 2001; Garcia-Hernandez et al., 2002; Rhee et al., 2003; Weems et al., 2004; Swarbreck et al., 2008, Lamesch, et al., 2010, Berardini et al., 2016). The TAIR database contains information about genes, proteins, gene expression, mutant phenotypes, germplasms, clones, genetic markers, genetic and physical maps, genome organization, publications, and the research community. In addition, seed and DNA stocks from the Arabidopsis Biological Resource Center (ABRC; Scholl et al., 2003) are integrated with genomic data, and can be ordered through TAIR.

TAIR is a curated database; data are processed by Ph.D.-level plant biologists who ensure their accuracy. Curation adds value to the large-scale genomic data by incorporating information from diverse sources and making accurate associations between related data. Data from manual literature curation, such as protein localization, biochemical function, gene expression, and phenotypes, are added to the corpus of knowledge presented for each locus in the genome. TAIR aims to produce a ‘gold standard’ functionally annotated plant genome that plant biologists can

This is the submitted version. For the final, edited version see:

use as a reference for understanding gene function in crop species and other plants of importance to humans (Berardini et al., 2016).

The database content and other information relevant to plant scientists can be accessed through dynamic Web interfaces and static hypertext (HTML) pages. Users can perform simple searches of much of the database using names or keywords. Advanced search forms for different data types are used for more complex or specialized queries. Genomic data can be accessed through text-based queries, via the graphical genome browsers (SeqViewer; GBrowse; see UNIT 9.9), and with a variety of sequence similarity tools such as BLAST and WU-BLAST (see UNIT 3.3, 3.4 & 3.11). Data from TAIR can also be obtained in bulk from selected query tools and downloaded from the Web site. TAIR provides an extensive set of links from the database and Web site to other sources of *Arabidopsis* genomic data around the world.

The data and services of TAIR are organized into eight categories, which appear on the navigation toolbar on all TAIR pages. Text-based query tools for performing simple and complex searches of specific types of data in TAIR, such as genes (see Basic Protocol 2), DNA, proteins, polymorphisms (including alleles), people, laboratories, and germplasms (see Basic Protocol 5) are found in the **Search** section. The **Browse** section allows the user to browse the ABRC stock catalog, the *Arabidopsis* transposon families, *Arabidopsis* gene families, as well as Gene and Plant Ontology terms (see Basic Protocol 4) and other data types. Within the **Tools** section are TAIR’s graphical genome browsers (SeqViewer, GBrowse; see Basic Protocol 3), MapViewer for aligning physical and genetic maps, sequence similarity software (NCBI BLAST and WuBLAST), Motif Analysis and Patmatch (see Basic Protocol 7), the TAIR synteny viewer GBrowse_syn (see Commentary), the literature full-text search tool Textpresso (see Commentary), an *Arabidopsis* chromosome map tool (see Commentary), among other data analysis and visualization tools. Under the **Tools** section, one will also find tools for downloading sets of sequences, protein data, Gene Ontology assignments (see Basic Protocol 4), and other curated data sets (see Basic Protocol 6 and Commentary) for a list of genes, as well as a GO term enrichment tool for Arabidopsis and other plant species (see Basic Protocol 4). The **ABRC Stocks** section contains links to the ABRC stock catalog (see Basic Protocol 5), DNA and germplasm searches (see Basic Protocol 5), and information about the stock center. The **Portals** section hosts pages with links to other databases and Web sites containing useful data and tools. **Portals** also contains comprehensive lists of community resources generated by large-scale functional genomics projects, general information about *Arabidopsis* biology and history, and educational resources. The **Download** directory contains several logically organized directories containing large data sets related to gene, sequence, microarray, Gene Ontology annotations and other data. The **Submit** section contains forms and documentation for submitting data to TAIR.

Users can contribute functional annotation for submitted papers using the web-based TAIR Online Annotation Submission Tool (TOAST; Basic Protocol 8) or by providing data in preformatted spreadsheets. TAIR also maintains the Gene Symbol Registry for Arabidopsis and registered users can submit Gene Symbols via the web (see Commentary). In the **News** section are links to the *Arabidopsis* community newsgroup, announcements from TAIR, meetings, and job postings.

For *Arabidopsis* to be effectively used as a reference plant species, it is essential that researchers know what data are available and how to use the information they obtain. This unit includes several basic protocols for accessing the wealth of information about *Arabidopsis* genes that has been generated by the research community and made available through TAIR. The types of data and tools at TAIR are diverse and cannot all be described in a single unit. Therefore, this unit

---

focuses on the data and tools that are related to retrieving, mining, and visualizing information about Arabidopsis genes. These protocols are based upon data and tools available as of August, 2017. As with any actively updated Web-based informatics resource, the data and tools will change over time.

**BASIC PROTOCOL 1**

**TAIR HOMEPAGE, SITEMAP, AND NAVIGATION**

The TAIR home page (http://arabidopsis.org) is the main entry point to the database and Web site (Fig. 1.11.1). To facilitate navigation of the TAIR Web site, a navigation toolbar is located at the top of all TAIR pages containing headings such as Tools, Search, and Portals. When mousing over each item in the tool bar, a drop-down menu appears with clickable submenus that lead to a variety of datasets, tools, and external links. Several additional buttons are located above the main toolbar, including items such as Help, About Us, Subscribe, Register and Login. The Help section of the Web site (http://arabidopsis.org/help/) provides a quick guide to new users, frequently asked questions, a glossary of terms used on the Web site, tutorials, a search help function, and user guides for database searches, specific tools, and registration. Registered users can click on Login to order stocks, submit data and update personal information. The About Us section has information about the project, its goals, and its deliverables. The home page also includes quick links to connect with TAIR via social media (Facebook and Twitter) and through a YouTube channel where users can view video tutorials.

**Necessary Resources**

**Hardware**

Computer with Internet access

**Software**

Up-to-date Web browser. The browser must have cookies enabled to log in and process stock orders. TAIR makes extensive use of JavaScript; this feature must also be enabled. See http://www.arabidopsis.org/help/index.jsp for information on properly configuring one’s browser.

**Performing a quick search**

1. Go to the TAIR home page (http://www.arabidopsis.org). Type the search term into the text box in the upper right corner of the page and choose a category from the drop-down menu (see Fig. 1.11.1). Click the Search button.

   The quick search performs a name search for most of the objects in the database (e.g., Genes, Clones, ESTs or BAC ends, People/Labs, Polymorphisms/Alleles, Germplasms, Ecotypes, Keywords, Genetic Markers, Proteins, Seed and DNA Stocks by stock name, and Vectors). By default, this is a “contains” search (a...
search for aba1 retrieves both ABA1 and ATRABA1A). It is also important to be aware that this search is not limited to the name field. For example, if the gene category is chosen, the gene description and keywords fields will be searched as well as the name. This is done to avoid missing any potentially relevant results, but may produce a large number of results.

2. A list of all matching records is displayed for the data type chosen. Click on each record to access full details for that object, or download the current page of results using the download button at the top of the page. For gene search results, the additional option “download all” provides a way to download the entire result set at once, and “get all sequences” provides an option to download sequences for all the genes in the result set.

3. Alternatively, to search for any data type in TAIR by name, choose “Exact name search” from the drop-down menu to the right of the box where the search term was typed in step 1. The query will return a summary (TAIR Search Result) page listing all data types with matching records and the number of records for each data type. Click on any item in the list to display a summary of all the records retrieved for that data type. In this example, clicking on Proteins displays a list of the two ABA1 proteins encoded by different splice forms of the ABA1 gene.

4. In the event that a general query returns too many results, try an Advanced Search for the specific data type (see Basic Protocol 2 for an example of an advanced search for Genes). The advanced search parameters can be used to narrow down an overly broad query.

BASIC PROTOCOL 2

FINDING COMPREHENSIVE INFORMATION ABOUT ARABIDOPSIS GENES

The locus detail pages represent the most comprehensive starting point for a researcher interested in finding out what is known about a gene. The physical location of an annotated gene on the genome is called a locus in TAIR. The locus serves as a useful concept for grouping genes with other objects having the same genomic location. For convenience, genetically defined genes (i.e., those identified by linkage studies but which are not yet associated with a genomic sequence) are also included as loci that have a genetic, but no physical location. Each locus is associated with at least one gene model, which can be thought of as a version of a gene. Several gene models (labeled as splice variants in TAIR) can be associated to a gene locus based on the existence of predicted or verified alternative transcripts. Every sequenced locus is assigned a unique identifier, the Arabidopsis Genome Initiative (AGI) locus identifier. This has the format AT (for Arabidopsis thaliana) X (where X is either a number from 1-5 corresponding to one of the 5 nuclear chromosomes or C for chloroplast or M for mitochondrion) NNNNN (a 5 digit number). The locus detail page collects information such as gene symbols and full names, experimentally determined or predicted function, gene expression data, mutant phenotypes, associated germplasms, polymorphisms, clones, and publications. Because data in TAIR are highly integrated, it is possible to access the locus detail page from detail pages of almost every other
type of object in the database. This protocol illustrates a commonly used way of finding genes using the Advanced Gene Search form.

**Necessary Resources**

See Basic Protocol 1

**Searching for information about a specific gene or set of genes**

1. Go to the TAIR home page (http://www.arabidopsis.org). In the top navigation bar click on the Search header (see Fig. 1.11.1) and select the Genes link to go to the TAIR Gene Search page (http://www.arabidopsis.org/servlets/Search?action=new_search&type=gene).

2. To search by name, choose “Gene name” as the option from the Search Name drop-down menu (the options include “Gene name,” “description,” “phenotype,” “GenBank accession,” “GenBank gi,” “Locus TAIR object ID” or “Gene TAIR object ID”). Using the drop-down menu to the right of this, set the search to an exact match or an inexact match (the options are “contains,” “starts with,” “ends with,” or “exactly”) and type the name in the text box on the right-hand side of the same line. For example, to find a set of related genes sharing a gene symbol, such as ARF for Auxin Response Factor family members (Hagen and Guilfoyle, 2002), type in ARF as the name term and choose the “starts with” option to the left of this. Click the “submit query” button.

   *Gene names include systematic names assigned based on chromosomal location (so called ‘AGI locus identifiers’ such as AT1G01010) or gene symbols. For more information about Arabidopsis gene nomenclature, see the Arabidopsis Gene Nomenclature Guidelines (http://www.arabidopsis.org/portals/nomenclature/guidelines.jsp).*

3. All of the loci that match the query term will be displayed in a list of results (on a page titled TAIR Gene Search Results). Click on the locus name to view the locus detail page. A sample locus detail page obtained by using the search name ABA1, and then selecting the AT5G67030 locus from the TAIR Gene Search results page, is shown in Figure 1.11.2.

   *The default search only retrieves genes that are active in the database. Checking the “include obsoleted genes” check box will retrieve both active and obsoleted genes, along with the history of their status in the database. Genes may become obsolete if they are merged with other genes—or if improved genome annotation methods find inadequate evidence for their existence. TAIR retains information about obsolete genes in order to maintain a record of their histories and associations.*

**Using the detail pages to find information about a locus**

4. On a locus detail page (Fig. 1.11.2), data sections are displayed in alternating background color bands; related data are generally grouped together. The following annotations (the red lettered items on the left side in Fig. 1.11.2) summarize the typical information displayed on a locus detail page. Definitions of each data type can be obtained by clicking on the adjacent question mark image to display a pop-up.

   This is the submitted version. For the final, edited version see:

definition window.

a. Representative gene model and summary information (Fig. 1.11.2A, a items). The Representative gene model for a protein coding gene is the gene model with the longest coding sequence (CDS); for other gene types, the representative model is set as default to the .1 model.

Data in this section includes Gene Model Type, Other Names and Summary. Example gene model types are protein coding, pseudogene, non coding RNA, among others. Other names include gene symbols and full names curated from the literature or provided by researchers via the Gene Symbol Registry. The Description field is a short summary of the gene’s function either manually composed by a curator or computationally generated. The latter is only shown if the locus has not yet been curated manually. Descriptions from Araport 11 were computationally generated (Cheng, et al., 2017).

b. Other Gene Models/Map Image (Fig. 1.11.2A, b items). Links to other gene models (termed splice variants in TAIR) are displayed below the representative gene model information. Clicking the gene model name will open a new window displaying the gene model detail page. View this page to see gene model specific data such as gene features in a tabular format and annotations that are specific to individual gene models. The Map detail image is a graphical display of the exon-intron boundaries of all the gene models of a locus. Clicking on the image directs the user to GBrowse (see Basic Protocol 3).

c. Gene function, biological role, and localization (Fig. 1.11.2A, c item).

The Annotations section contains all of the controlled vocabulary terms that have been assigned to describe the molecular function, biological role, subcellular localization, and expression of the gene product. The annotations are grouped according to the type of vocabulary and summarized on the locus page. Click on the Annotation Detail link (located at the bottom right of the Annotations section) to display the full annotation details, which include the type of evidence supporting the annotation and the corresponding reference that is the source of the data supporting the annotation.

d. Sequences (Fig. 1.11.2A, d item).

Links to genomic sequence, full-length CDS, full-length cDNA, and protein sequence are located in the Sequence section. Clicking on the sequence name will display a new window containing the sequence, which can be uploaded directly into TAIR’s WU-BLAST tool. In addition to WU-BLAST, TAIR also hosts a version of NCBI BLAST (Altschul et al., 1990). These tools are available with some specialized Arabidopsis sequence data sets such as intergenic regions, upstream and downstream sequences, and UTRs.

This is the submitted version. For the final, edited version see:

These tools can be accessed from the TAIR homepage under the **Tools** section.

e. Gene expression (Fig. 1.11.2A, e item).

Information about the expression of the gene can be found in the **RNA Data** section and the lower part of the **Annotations** section. In the RNA Data section, array elements from one-channel and/or two-channel experiments that map to the locus are listed. Array element names are linked to detail pages. Note that TAIR stopped integrating and updating microarray data in 2005, see Commentary for more current datasets and tools. Lists of full-length cDNAs and expressed sequence tags (ESTs) can be found in the **Associated Transcripts** subsection within the RNA Data section. Click on the number next to the type name to see a list of all the clone records. The clone records are linked to GenBank, where information about the cDNA libraries (and therefore expression) can be found. Finally, information about gene expression, curated from the literature, is shown in the **Annotations** band along with the Plant Ontology associations (see Fig. 1.11.2A, section “c”: “expressed during”, “expressed in”).

f. Protein data (Fig. 1.11.2A, f item).

Structural and physical characteristics of the protein encoded by the reference gene model, including molecular weight, conserved domains, and isoelectric point, are displayed in this section. Click on the AGI name in the protein section to open a new window displaying more detailed information and the amino acid sequence itself.

g. Gene families, PANTHER tree viewer, PANTHER plant homologs (Fig. 1.11.2A, g item).

The **Gene Family** data section displays information about proteins that are evolutionarily related to the locus. The gene families section provides direct links to orthologs in external resources such as Ensembl Plants (Bolser et al., 2016), PLAZA (Proost et al., 2015) and Phytozome (Goodstein, et al., 2012), among others. Clicking on the PANTHER tree view glyph will open PANTHER’s phylogenetic tree viewer (Mi, et al., 2017) in a new window displaying the phylogenetic tree of the corresponding multispecies PANTHER gene family, centered on the locus. The PANTHER Plant Homologs section provides a list of plant homologs from the PANTHER resource (Mi et al. 2017) as well as links to their corresponding entries in UniProt (to view the protein record), and the EBI QuickGO (to obtain GO annotations). This information makes it easier for users to find orthologs and analyze gene families to facilitate comparison of protein function between Arabidopsis and other species.

h. Map locations (Fig. 1.11.2A, h item).

The **Map Locations** section displays the chromosome and coordinates of the
locus for the maps on which it is found. The gene can be viewed in a whole-genome context by clicking on one of the three map options (Map Viewer, Sequence Viewer, GBrowse) in the Map Links section (See Basic Protocol 3).

i. Alleles and polymorphisms (Fig. 1.11.2A, i item).

All of the polymorphisms that map within the locus are shown in the Polymorphisms section, along with the type of variation. This section includes natural variations found in different ecotypes and induced mutations (e.g., T-DNA insertions) that have been mapped by sequence identity and alleles that have been curated from the literature. To find detailed information about a polymorphism, click on the name of the polymorphism.

j. Germplasm information (Fig. 1.11.2A, j item).

The Germplasm section provides information on all germplasms available for a locus, including phenotype descriptions and images of plants (if available).

k. Clones (Fig. 1.11.2B, k item).

Clones linked to a locus may include vectors, BACS, clone ends (ESTs) that contain sequences from the locus of interest. If the clone is an ABRC stock, that information will be displayed along with an option to select stocks to order (see Basic Protocol 5).

l. External links (Fig. 1.11.2B, l item).

There are other Web sites that provide either alternate views or different information about a locus (see Commentary). In order to provide access to as much information about a locus as possible, TAIR provides links to the corresponding locus pages in other databases and Web sites. Types of external links include other Arabidopsis genome annotation databases, gene expression databases, and functional genomics sites, as well as links to tools for further analysis. For example, all sequenced loci are linked to other Arabidopsis annotation databases including Araport, NCBI, and MIPS. Links are grouped by data types such as: Genomics, Expression/Localization, or Interactions. TAIR also provides links to UniProt and NCBI Reference genome from the protein detail pages.

m. Community Comments (Fig. 1.11.2B, m item).

Comments may contain additional data contributed by registered TAIR users, and are included in the display for nearly all of the TAIR detail pages. This function can be used to report new data, as well as errors or omissions related to the displayed object (see Basic Protocol 8 and http://www.arabidopsis.org/help/helppages/addcomment.jsp)

n. Publications (Fig. 1.11.2B, n item).
Papers and conference abstracts are shown at the bottom of the detail page in the section marked Publications. Publications include published literature imported from PubMed, Agricola, and BIOSIS, along with abstracts from the International Conference on Arabidopsis Research. Only the most recent ten papers are listed on the detail page; to retrieve the complete list, click on the View Complete List link. Clicking on the title of the publication opens a new link to the detailed record where one can read the abstract, link to the PubMed citation, associated loci and annotations, and find authors among TAIR’s community.

- **Update History** (Fig 1.11.2B, o item).

  TAIR maintains a history of changes to a locus for the purposes of tracking. Changes may include merges, splits or insertions.

**Saving the results of a search to a file**

5. Return to the list of results obtained by the query submitted in step 2 (page titled TAIR Gene Search Results). Check the box to the far left of the results summary. Each page of results must be saved separately. Only those results that are selected will be saved. Use the Check All function to save all of the results displayed on the page.

  Before downloading a large set of results, use the browser to go back to the Advanced Search page, make sure the number of records per page of results is set to the maximum (usually 200 records/page), and resubmit the query.

6. After selecting all of the desired results on a page, click on the Download Checked button (or Download All if you wish to export all results) in the upper right corner of the TAIR Gene Search Results page. The checked results will then be displayed in the browser window as tab-delimited text file. Use the Save As function under the File menu in the browser toolbar to save the results in a file on the local computer. This process must be repeated for each page of results.

7. In order to retrieve sequences for the selected results, click on the Get Checked Sequences button (or Get All Sequences if you wish to retrieve sequences for all results) on top of the TAIR Gene Search Results page. This will bring you to the Sequence Bulk Download and Analysis page from where you can retrieve different types of sequences for your list of genes. For more information about that tool, see Basic Protocol 6.

  The download feature is found on all of the search results pages. Each set of results includes different information in the downloadable file. See the help documents for the specific search to view a listing and description of the downloaded fields. The files contain tab-delimited text that can be opened using a text editor or spreadsheet software such as Microsoft Excel. The download sequence option is only available on the Gene Search Results page.

This is the submitted version. For the final, edited version see:

BASIC PROTOCOL 3

USING THE ARABIDOPSIS GENOME BROWSERS (SeqViewer AND GBrowse)

TAIR provides two alternative Web applications (SeqViewer and GBrowse) that allow users to explore the annotated Arabidopsis genome sequence. SeqViewer is a graphical genome browser developed by TAIR while GBrowse (Stein et al., 2002) was developed by the Generic Model Organism Database project (GMOD; www.gmod.org). Both tools allow the user to search for and display various sequence features such as genes, polymorphisms, T-DNA insertions, and transcripts (ESTs/cDNAs), provide a mechanism for navigating around the genome, and allow individual users to customize the type of data displayed. These tools are useful for a wide variety of tasks including positional cloning, identifying mutants in a gene of interest, finding cDNA and ESTs for a gene of interest, and finding and displaying the distribution of sequence features (e.g., polymorphisms, T-DNA insertions) in a whole-genome context. While both tools share some functionality, each tool has its own specific set of features. Additionally, GBrowse contains many data types not represented in SeqViewer.

Necessary Resources

See Basic Protocol 1

Exploring SeqViewer

Displaying a defined region of the genome

1. Go to the TAIR home page (http://www.arabidopsis.org). In the Tools drop-down menu of that page, click on the link to SeqViewer. Alternatively, go directly to the URL https://seqviewer.arabidopsis.org/.

   This will invoke the SeqViewer home page, which shows the five chromosomes of the nuclear genome sequence represented as five green bars, one for each chromosome. When using SeqViewer, it is a good idea to note the version number/date shown below the chromosome bars. Genome annotation changes over time; the versioning/time stamp provides a way of tracking annotations that may change or become obsolete.

2. To search by name, make sure that the “name” radio button is selected, then enter (by typing) or upload (using the Browse button) a file of up to 250 names into the text input box in the lower right section of the home page. For example, to search for the gene AT1G077810, enter AT1G07810 in the text input box and choose “gene” as the name search option from the drop-down menu to the right of the “name” radio button. Submit the query by clicking the Submit button. The results of the search are shown in “whole-genome view” (Fig. 1.11.3). Matches to the genome are displayed as tick marks (red on screen) on the green chromosome bars in the whole-genome context.

This is the submitted version. For the final, edited version see:

3. The genome browser can also be searched using short sequences such as PCR primers used for genetic markers or highly conserved sequence motifs such as miRNA core binding sites. Paste in or upload up to four nucleotide sequences (each between 15 and 150 nucleotides long) in FASTA format and choose the radio button to search by “sequence.” The sequence search finds only exact sequence matches; ambiguous matches are not allowed. Hits to the genome are displayed as red tick marks on the whole-genome view and in a special Match track in the Close-up view (Fig. 1.11.4).

**Displaying a close-up view of a genomic region**

4. To display an enlarged view of the genomic region centered on the gene or other object found in the search, click on the red tick mark corresponding to the match. This opens up a Close-up view of a 200-kb region, approximately centered on the gene highlighted in yellow (and similar to the view shown in Fig. 1.11.4).

   *Click on any object to retrieve its detail page from the database. Mousing over the data will display a brief summary in a small pop-up window (Fig 1.11.4, e).*

   *In the example shown in Figure 1.11.4, the highlighted locus is displayed at a 10-kb resolution and centered in view. To obtain a similar view using the Close-up view controller, first zoom to 10 kb by selecting this option from the drop down menu (shown in the close up view section of Fig. 1.11.4), then enter the name of the locus in the text entry box next to the Find button then click on the Find button (below the “Zoom to” controls in Fig. 1.11.4). The display will now show a 10-kb window centered on the selected locus, which is highlighted in yellow, as shown in Figure 1.11.4.*

   *Alternatively, use the cursor to center the view. Move the cursor along the centering bar at the top of the Close-up view between the left and right scroll arrows (letter “c” in Fig. 1.11.4). A yellow bar will appear above the cursor, indicating which region to select. When the yellow bar is over the desired region, click once in the centering box.*

5. Create a custom view of any region of the genome by clicking on the appropriate chromosome in the whole-genome view (i.e., the screen illustrated in Fig. 1.11.3). A new Close-up display will appear centered on the selected region of the chromosome. In the Close-up view control panel (Fig. 1.11.4) enter the left and right coordinates into the Select Range box at the left of the screen and click the Go button.

   *Each chromosome corresponds to a pseudomolecule that is a composite of all linked BAC sequences in the genome tiling path. BAC sequences may be trimmed or extended in regions of overlap to ensure a continuous sequence. Coordinates for each base pair are indicated by the following convention: numbers start from the top of the upper chromosome arm (to the left of the centromere on the SeqViewer Whole Genome View) and end at the bottom of the lower chromosome arm. When selecting coordinates to input for generating a custom view, the quickest way to find coordinates for an object is to obtain this data from the detail page. For example, to create a custom view between two loci, go to each locus detail page and find the coordinates shown for the AGI map in the Map.*

This is the submitted version. For the final, edited version see:

Locations band. Pay close attention to the orientation of the locus: if the two objects are on opposite strands, the starting coordinate of one sequence will be flipped, relative to the other.

The custom view feature is useful for positional cloning. A genetically defined region between two markers can be displayed to find new markers or polymorphisms in a region for fine mapping. Alternatively, when searching for candidate genes within the interval, a downloadable summary of the genes located in the displayed region can be obtained by clicking on the List Genes in Range button.

Displaying selected bands of data in the Close-up view

6. The global controller on the left-hand side of the whole-genome view page can be used to select sequence features to display (letter “c” in Fig. 1.11.3). Each sequence feature (genes, transcripts, polymorphisms, T-DNA/Transposons, genetic markers, and annotation units) can be removed or added to the display in the Close-up view by checking or unchecking the box next to the feature name. These check boxes are also found on the Close-up view page, and therefore can be used before or after zooming in to a Close-up view (letter “a” in Fig. 1.11.4); the selection will affect all open Close-up views. To see a complete explanation of the bands and graphics used in the display, click “show legend” located just above the “zoom to” box in the Close-up view control panel (left side of Fig. 1.11.4).

Displaying all rows of data in the Close-up view

7. In order to simplify the display, the default Close-up view shows only three rows of data for all sequence features. Items not displayed are indicated by black tick marks; displayed items are indicated with red tick marks. To display all of the data for each sequence feature in the Close-up view, zoom to between 10 kb and 200 kb and click the radio button to display all data in the box below the green chromosome bars (indicated by the letter “b” in Fig. 1.11.4). This will expand the rows for all of the selected data types to show all objects in the Close-up view. Before showing all rows it is a good idea to zoom in to a fairly high level of resolution, as large amounts of data can result in a very long page. Zoom levels of 1 Mb and higher will only display a maximum of six rows for each type of data. Clicking on any of the data types will open the corresponding detail page in TAIR. The six types of data available are as follows.

a. Markers.

Genetic markers shown in SeqViewer are mapped based on sequence identity and include the following types: simple sequence length polymorphisms (SSLPs); cleaved amplified polymorphisms (CAPS); amplified fragment length polymorphisms (AFLPs); restriction fragment length polymorphisms (RFLPs); and other markers detected by hybridization such as single nucleotide polymorphisms (SNPs) used on whole-genome mapping arrays (Borevitz and Nordborg, 2003). The marker type is indicated in the mouse-over pop-up window. Markers can be used for a variety of purposes such as genetic mapping or positional cloning and as linked markers for tracking specific alleles.

This is the submitted version. For the final, edited version see:

b. Polymorphisms.

Polymorphisms include substitutions, small insertions (less than 20 bp), deletions, and combination insertion-deletions (INDELs). All polymorphisms are mapped relative to the reference (Col-0) genome, although TAIR includes polymorphisms between a wide variety of other natural variants (ecotypes). Sequence variations between natural populations can be used as a starting point for generating genetic markers for mapping, designing allele specific primers for a given locus, quantitative trait analysis, and linkage disequilibrium studies. TAIR has incorporated and mapped hundreds of thousands of polymorphisms from several large-scale SNP identification projects including the Perlegen project (Clark et al., 2007), the Multinational Arabidopsis Steering Committee SNP Database (https://www.genomforschung.uni-bielefeld.de/GF-dataresources/masc/search_masc_snps.php), Nordberg Lab Genomic Survey and Linkage Disequilibrium project, and the Stanford Genome Center (SGC). Cereon has also made a list of over 50,000 polymorphisms between the two most common laboratory strains (Columbia and Landsberg erecta) freely available to academic researchers and nonprofit institutions. The Cereon dataset has not been incorporated into the SeqViewer, but can be downloaded from the TAIR Web site (http://arabidopsis.org/Cereon/index.jsp).

c. T-DNAs/Transposons.

Plant genomic sequences flanking T-DNA or transposon insertions are used to map approximate insertion sites onto the genome. The arrowhead indicates the direction of the sequence and points away from the insertion site that lies at the opposite end. Thick lines indicate regions of the insertion flanking sequence that match to the genome and thinner lines indicate regions that do not match the genomic sequence. The mouse-over pop-up window shows the name of the insertion, the start position, and length of the flanking sequence. The positions of insertions are rough estimates and should be confirmed by amplifying the product using genomic and insertion sequence primers and resequencing the ends. The validated sequences can be used to update the records in TAIR (see Basic Protocol 8, Community Comments).

d. Gene models.

For each locus, the representative gene model and all splice variants are displayed in this band. Each splice variant is indicated by a suffix following the locus identifier (e.g., .1, .2, .3). Arrowheads indicate the direction of transcription. Exons are shown in yellow, introns are white, and UTRs (if known) are shown in red.

e. Transcripts.

Full-length or partial cDNA transcripts are indicated in blue, and ESTs are gray. The exon-spanning regions are indicated with solid boxes and the introns by thin lines. Arrowheads indicate the direction of transcription. For genes without full-length cDNA support, the transcripts can be used to verify the gene model structure as well as to identify misannotated or unannotated genes. Some transcripts may map to intergenic regions and may indicate the presence of a
gene that has not yet been annotated. Other transcripts may indicate the presence of alternatively spliced forms, or genes for which de novo methods of detection predicted incorrect products.

f. Annotation units.

Annotation units are units of sequence derived from large-insert clones that comprise the backbone of the whole-chromosome assembly. To simplify construction of the tiling path and annotation of genes in regions of overlap, some of the original genomic clone sequences were trimmed and others were extended based on neighboring clone sequences. Therefore, the annotation unit sequences no longer represent the original clone sequences, and the coordinates of genes and other features mapped on annotation units differ from the coordinates on the corresponding clone sequences found in GenBank. Genomic sequence corrections applied for TAIR genome releases require recalculation of the chromosome coordinates and assembled sequence. TAIR maintains a list (http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/agicomplete.jsp) of incompletely sequenced BACs and known gaps remaining from the genome sequencing project.

The Arabidopsis genome assembly did not change between TIGR5 and TAIR8, was updated for TAIR9 (June 19, 2009). The assembly did not change with the update to TAIR10 (Nov 17, 2010) or Araport 11 (Cheng, et al., 2017). Details on the TAIR9 assembly update can be found in the Genome Annotation section of the TAIR website (http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp).

TAIR provides a script that allows users to convert coordinates from TIGR5, TAIR6, TAIR7, or TAIR8 to the coordinates on the updated TAIR9 assembly. This script, and instructions on how to run it, can be found in the Software section of the TAIR Downloads directory (http://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FSoftware%2FUpdateCoord).

**Using the SeqViewer Nucleotide View to view genome annotations**

From the SeqViewer Close-up view, there are several ways to drill down to the 10 kb Nucleotide View.

8. From the Close-up view, use the sequence ruler (Fig. 1.11.4) to select a region of the genome to view. Point the cursor to the desired area in the ruler and click. A SeqViewer Nucleotide View window appears as shown in Figure 1.11.5.

Alternatively, to view a 10-kb region centered on a specific object, position the cursor over the object (such as the locus AT1G01200.10) and click on the link to “nucleotide seq view” in the pop-up window that appears. A third option is to go directly to the Nucleotide View from the list of matches to the genome. After submitting a query, click on the link to the list of matches above the whole-genome view in the SeqViewer. Click on the coordinates in the last column of the list (Location) to display the Nucleotide View.
Displaying selected features in the Nucleotide View

The Nucleotide View (Fig. 1.11.5) shows 10 kb of sequence at a time. The view can be scrolled 5 kb upstream or downstream using the arrows at the top and bottom of the view. The location of genes is shown on the far right; the direction of the arrow indicates the direction of transcription. The display can be set to show specific sequence features singly or in combination, on either DNA strand. An explanation of the display is shown in the legend at the top of the page. The procedure below describes how to view the location of the T-DNA insertion in the Nucleotide View.

9. In the drop-down menu selector labeled Choose Objects to be Highlighted (located in the upper right corner of the Nucleotide View, shown in Fig. 1.11.5) select Genes/T-DNA/Tn Insertions. The display shows the translation start and stop points of various T-DNA/Tn insertions (1.11.5) in blue highlighting; the UTRs are red, exons are uppercase and yellow, and introns are lowercase and purple. The positions matching insertion flanking sequences are shown underneath the corresponding genomic sequence. The matching regions are represented by a double dashed line (==); nonmatching regions are represented by a single dashed line (—), and the approximate point of insertion at the 5’ end of the flanking sequence is shown as a vertical line (|). The arrowhead (3’ end of the flanking sequence) shows the orientation of the flanking sequence relative to the chromosome.

One of the many nice features of SeqViewer is the ability to copy and paste sequences directly from the nucleotide sequence view where the upper/lowercase formatting is retained. This can be useful when exporting sequences to primer design programs and selecting primers that span introns.

Exploring GBrowse

Viewing a gene or region of interest in GBrowse

10. Go to the TAIR home page (http://www.arabidopsis.org). In the Tools section of the menu bar, click on the link to GBrowse. Alternatively, go directly to the URL https://gbrowse.arabidopsis.org/cgi-bin/gb2/gbrowse/arabidopsis/

11. The GBrowse (Fig. 1.11.6) display is divided into the following sections:
   a. GBrowse menu, the default is the browser view, (a item) other options are Select Tracks, Snapshots, Custom Tracks, and Preferences
   b. Search section with ‘Landmark or Region’ (b item) that allows you to input your query and sample data entry points, under ‘Examples’ with examples of GBrowse search queries.
   c. Data source drop down menu (c item), used to select the data source (genome of choice).
   d. Reports and Analysis selector (d item). ‘Configure’ button to select ways to visualize data, in tabular or graphic format; ‘Save Snapshot’ to create a snapshot of the current viewing panel or ‘Load Snapshot’ to upload a previously saved snapshot.
   e. Scroll/Zoom (e item) controls for GBrowse window size and zoom.
f. Overview panel (f item), which shows a graphical representation of the selected area on the entire chromosome.

g. Region (g item), shows the currently displayed subset of the chromosome.

h. Detail view, showing tracks (h item, individual tracks marked with *). Each track is displayed as a pictorial representation of the genomic features for the selected region (e.g. Locus, ‘Annotation Units’ in Figure 1.11.6h).

12. The names and position of genomic features such as genes or genetic markers can be entered in the search box (Fig. 1.11.6, item b). For genes, either the AGI code (e.g., AT1G05460) or gene symbol (e.g., SDE3) is a valid search query. Nucleotide ranges can also be entered to allow specific regions of interest to be displayed. The chromosome and start and end coordinate of the desired region must be entered in the following format Chr1:1504365..1514364.

If a query returns multiple hits, GBrowse will display these as distinct rows with the position of each feature shown. Clicking the desired hyperlink will open the detail display for the selected region.

13. The assembly version or build can be selected from the Data Source drop-down menu (Fig 1.11.6 c). By default, the most recent Arabidopsis version is displayed.

Altering the Data Source changes the chromosome sequence, gene models, and other tracks to those of the specified release, allowing alternative versions to be compared. Note that gene models or other features present in a later release may potentially be absent, located at a different position, or otherwise altered relative to an earlier release. A description of the latest genome release can be found at http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/genome_annotation.jsp, details of earlier releases can be found on the TAIR site (http://www.arabidopsis.org/download/index-auto.jsp?dir=/download_files Genes).

14. Entering a feature name or region and clicking Search (Fig 1.11.6 b) will update the overview and details display. The overview map shows the position of the region displayed in the detail view relative to the rest of the chromosome. The size of this region is shown in the Scroll/Zoom drop-down. As default, certain annotated data is displayed in the details panel; this includes gene models, annotation units (BACs), Arabidopsis cDNAs, and polymorphisms.

If a specific feature was searched for by name (e.g., AT1G05460.1), the feature is highlighted in yellow. This highlighting provides a convenient way of maintaining position of the feature when the view is expanded to display a larger region. Highlighting can be turned off by clicking the Clear highlighting button directly below the details display panel.

15. The zoom feature (Fig 1.11.6.e) can be used to adjust the viewing dimensions in order to display a larger-scale view of the genome. Select the desired region size from the drop-down menu and the display will automatically reload to a new detail map covering the region.

This is the submitted version. For the final, edited version see:

The change in view scale is reflected in the increased size and position of the purple box in the Region panel and the change in genomic coordinates in the search box.

16. To move along the chromosome, click on the grey arrows (Fig 1.11.6 e) to shift the display to the left or right. The entire display of the chromosome can be flipped by checking the Flip box. Flipping the display may be useful when viewing a gene located on the minus strand.

_Moving the cursor over a feature will bring up a pop-up box that displays additional information specific to the feature. For genes, this includes known symbols or the confidence ranking for that gene model. Every feature is also hyperlinked; clicking on a feature will open a new data page specific to the feature. For example, clicking on a protein coding gene model will open the TAIR gene page, whereas clicking on a Brassica EST transcript will link out to the relevant GenBank entry at NCBI._

Customizing the GBrowse display

17. To customize the tracks within the GBrowse view, click on the ‘Select Tracks’ tab in the GBrowse menu (Fig 1.11.6.a) or the ‘Select Tracks’ button at the bottom of the GBrowse viewer page. TAIR GBrowse has 12 track categories: **Araport11**, Assembly, Community annotation, DNA, Expression, Gene, Genomic Features, Methylation and Phosphorylation, Orthologs and Gene families, Sequence similarity, Variation, and Analysis. Each track category has multiple check boxes for different types of data (Fig. 1.11.7).

_Further information about the track can be obtained by moving the cursor over the track name. In addition, clicking the track name opens a separate page describing all available tracks._

18. To add or remove tracks from the detail display simply check or uncheck the required tracks and click the Back to Browser link. The track order can be adjusted by clicking the track title in the details panel and dragging the track up or down to a new position.

19. Configure individual tracks by clicking the wrench icon next to the track title. This allows the user to choose the shape and color of the glyphs, put a limit on the number of features displayed in any one region, and set preferences if a text label is displayed.

_Tracks can also be customized by clicking on the Configure tracks button. To revert to the default settings click Revert to default._

20. The Preferences (Fig 1.11.6 item a) panel can be used to change certain features of the details view such as the image width, position of the key, tool tip display, or how the tracks are listed. The highlight feature boxes can be used to highlight specific features or regions which may be useful when giving presentations or showing images in publications.

This is the submitted version. For the final, edited version see:

Visualizing private data in GBrowse

21. To upload your own annotation data to GBrowse go to the “Custom Tracks” tab part of the GBrowse menu (Fig. 1.11.6 item a). This feature allows you to view your own annotations such as primers and cDNA clones in the context of the Arabidopsis genome. GBrowse allows users to enter the features to view by (a) directly pasting the feature information on the text box, (b) uploading a track from a URL or, (c) uploading a file from a local computer in an acceptable format (e.g. GFF3 format; See the GBrowse help documents for information on accepted formats (https://gbrowse.arabidopsis.org/gbrowse2/annotation_help.html)

22. To add enter annotations manually, choose the option “from a file” from the Add custom tracks. Type in or copy the data. The browser accepts the annotation file format described in the tutorial document available by clicking on the ‘Help with uploading custom tracks’ hyperlink in the “Add your own tracks” panel (https://gbrowse.arabidopsis.org/gbrowse2/annotation_help.html).

23. To upload annotations from a local file choose the “from a file” option in the Add custom tracks menu. Locate the file by clicking on the Browse button; when found, upload the file. GBrowse will automatically incorporate the annotations into a new track in the details view (Fig. 1.11.6 item h). The new tracks can be configured in the same manner as tracks provided by TAIR. The annotation data text file can be edited or deleted by clicking on the respective buttons in the Add Your Own Tracks Panel (Fig 1.11.7B)

Uploaded annotations will persist until you delete them; these annotations are private and will not be seen by other individuals.

24. To upload annotations to GBrowse from a URL, paste the Web address into the Fetch track file from this URL box. This feature allows you to view annotations created by other groups in your own GBrowse. In addition, users with access to a Web server, can publish their own tracks to make them available to colleagues or collaborators. Further details about this process can be found on in the GBrowse help document (https://gbrowse.arabidopsis.org/gbrowse2/annotation_help.html).

Using the decorated FASTA function

GBrowse allows you to download a decorated FASTA file. This option allows you to extract the sequence in a particular region and highlight specific features of interest. For example, coding regions can be marked in a different colored font and polymorphisms shown in bold or underlined, allowing you to easily identify which polymorphisms lie in coding regions.

25. Go to the Reports and Analysis features box and select Download Decorated FASTA file from the menu options (Fig. 1.11.6, d item). Clicking Configure opens the feature configuration page.

26. From the configuration page you can select which features you wish to highlight on the FASTA sequence file and choose from a variety of markup options such as caps, italics, bold, and alternative font and background colors.

27. Once satisfied with your selection click Go. The new Web page will display the

This is the submitted version. For the final, edited version see:

FASTA sequence for the region displayed in the detail view with the selected features highlighted.

*On the configuration page there is an additional ‘Configure’ button. Clicking this will save the settings during your session so that any other downloads can be launched with the same settings by clicking the Go button next to Reports and Analysis features box (Fig. 1.11.6 c)*

**BASIC PROTOCOL 4**

**USING THE GENE ONTOLOGY ANNOTATIONS FOR GENE DISCOVERY AND GENE FUNCTION ANALYSIS**

Annotations (associations of controlled vocabularies or keywords to data objects) provide a richer, more complex picture of a gene that is also more computationally accessible for the purpose of querying, classification, and making correlations among seemingly unrelated data. TAIR makes extensive use of controlled vocabularies for describing data in the database. The controlled vocabularies (ontologies) that are used by TAIR are also used by other model organism databases, thereby facilitating cross-species comparisons. All of the ontologies used by TAIR are included in the Open Biological Ontologies Project (http://www.obofoundry.org/) where they are freely accessible.

TAIR is member of the Gene Ontology (GO) Consortium (http://www.geneontology.org) and participates by developing and refining the ontologies and annotating Arabidopsis gene products (The Gene Ontology Consortium, 2010). The GO controlled vocabularies describe three aspects of gene products: molecular function, biological process, and subcellular location. TAIR also imports manual and computational annotations for Arabidopsis made by other groups including UniProt, BioGrid, JCVI (formerly TIGR) and others (Wortman et al., 2003; Berardini et al., 2004). These annotations are contributed independently by each organization, to the GO database, where they are accessible through the AmiGO query tool for making cross-species queries (http://amigo.geneontology.org/amigo). The other main ontology used at TAIR is the Plant Ontologies developed by the Plant Ontology Consortium (POC; http://www.plantontology.org). The POC has used the GO model to develop controlled vocabularies for plant structures and developmental stages (Jaiswal, et al, 2005). In TAIR, both of these ontologies are used to annotate many additional types of data such as microarray experiments, gene expression, phenotypes, and publications. TAIR also collects and displays annotations contributed by members of the community who can use a simple web tool (TAIR Online Annotation Submission tool, aka TOAST) to provide GO and PO annotations for genes based on published works (see BASIC PROTOCOL 8).

**Necessary Resources**

See Basic Protocol 1

**Files**

WRKYFamily.txt

This is the submitted version. For the final, edited version see:

Using the Keyword Browser to find candidate genes

For researchers, finding candidate genes involved in a particular pathway typically involves a fishing expedition using a variety of genetic, molecular, and biochemical assays. The GO annotations can be useful in making educated guesses about what genes may act in a pathway or are members of transcriptional/signaling cascades. Because TAIR and its community contributors have focused on GO curation from the literature, Arabidopsis is the most well annotated plant genome, with a large number of experiment-based annotations. Thus Arabidopsis GO annotations can be particularly useful for being able to infer gene function for unknown genes in other plant species based on sequence similarity or evolutionary relatedness. Another common use of GO annotations is to identify sets of genes associated with a given function or process in Arabidopsis as a starting point to identify genes with similar functions in other species.

1. Go to the TAIR home page (http://www.arabidopsis.org), click Search in the toolbar (Fig. 1.11.1), and select Keywords from the drop-down menu that appears. The page shown in Figure 1.11.8A is returned (TAIR Keyword Search and Browse; can be directly accessed at http://www.arabidopsis.org/servlets/Search?action=new_search&type=keyword ). Enter term (keyword) “root development” in the text box and choose “contains” (an inexact search) from the drop-down menu to the left of the text box. From the group of check boxes for restricting the search, choose GO Biological Process as the keyword type and click the Submit Query button.

Many of the terms in GO exist as complex phrases. TAIR searches take the entire entered term or phrase as a complete phrase rather than a set of words. Consequently, an “exact match” search will often not retrieve any entries. Therefore, the authors recommend using the “contains” option for keyword searches.

2. On the Keyword Search Results page (Fig. 1.11.7A), each controlled vocabulary term is displayed along with a count of all data objects (e.g., loci, publications, annotations) annotated to that term. Click “loci” to display the genes annotated to “root development.” The results are displayed as a Gene Search Result page (see Basic Protocol 2) where all of the genes associated to the term “root development” or its children, are displayed. Click on the locus name to view the locus details or save the list as a text file (see Basic Protocol 2).

Finding genes annotated to related functions

3. On the Keyword Search Results page, find the listing for “root development,” and click on the “treeview” link. This will open a window displaying the term in a hierarchical tree view (Fig. 1.11.8B).

In the Gene Ontology, terms have a parent-child relationship to one another. Parent terms are less specific than their child terms. A child term may be a part of the parent (as thylakoid is part of chloroplast) or a type of the parent (as chloroplast is a type of plastid). In contrast to simple hierarchies, a child term may have more than one parent. The ontologies are intended to be as biologically accurate as possible. Terms and their relationships are defined by what is known about the biology of the process, function, or cellular component.

This is the submitted version. For the final, edited version see:

By examining the structure of the ontology to find related terms, related gene products can also be found via their annotations to the terms.

4. Click on the plus sign next to the parent term (“root development”) to expand the node and display all of the child terms.

5. To display genes annotated to each of the parent and child terms, select the “loci” radio button from the top of the tree view page (Fig. 1.11.8B), then click the Display button. The display will be redrawn to show a count of the number of loci annotated to each term and the number of loci annotated to the children of each term. Click on the link to list loci annotated to the term “regulation of root development” to find all loci that are annotated to this term.

Retrieving GO annotations for sets of genes

GO Annotations can also be used to rapidly classify sets of genes such as gene families or co-clustered genes revealed by analysis of high throughput expression data.

6. Go to the TAIR home page (http://www.arabidopsis.org), click Search in the toolbar (Fig. 1.11.1), and select Gene Ontology Annotations from the drop-down menu that appears. Alternatively, go to the URL http://www.arabidopsis.org/tools/bulk/go/index.jsp.

7. Upload a list of AGI locus identifiers using the sample data file WRKYFamily.txt. This file contains a list of 74 loci all belonging to the WRKY transcription factor family (Eulgem et al., 2000; http://www.arabidopsis.org/browse/genefamily/WRKY-Som.jsp). Select the Text radio button under “Select output type”; to results locally in a table format. Click on the “Get all GO Annotations” button. The output file contains a list of all the specified loci and their annotations to all three aspects of the GO ontology. The annotations include the evidence code and reference for the data supporting the annotation. The file can be saved onto a local computer as a tab-delimited text file. If the HTML option is chosen, the results are hyperlinked to TAIR detail pages for loci, keywords, and publications. The Web output also has links to the corresponding keyword entry in the GO database, where one can find annotations to genes from other organisms.

Classifying sets of genes into functional categories

8. Alternatively, instead of getting a list of all annotations, the genes can be grouped into broader categories based on their annotations. After uploading the gene list (step 7 above), choose “HTML output” and click the Functional Categorization button.

For each aspect of the GO ontologies, a subset of terms have been selected to represent 10 to 20 major categories, called GO Slim categories. If a gene is annotated to a child term of one of the GO Slim terms, it is included in the category. The GO Slim is less specific, but presents a simpler classification. The results include gene annotations that are both experimentally supported and computationally predicted. To find sets of annotated genes based on evidence codes, use Search by Associated Keyword on the Gene Search page

This is the submitted version. For the final, edited version see:

GO Slim assignments are also included in the detailed GO annotation output (from step 7). See http://arabidopsis.org/help/helppages/go.slim_help.jsp for a list of all GO Slim terms and their definitions.

9. The database will return a functional categorization list showing all categories represented in the genes from the input file, along with the frequency of distribution of the genes within the set (Fig. 1.11.9A). To view a list of genes in each category, click on the number in the “Gene count” column.

Only the categories represented by the genes in the list are included; the absence of any of the GO Slim categories means that there are no genes in the list that fall into that particular group. The default option displays the list grouped by keyword type and then by categories sorted by the number of annotations in each category. The table can be re-sorted to list by gene count. Frequency refers to the number of occurrences of a gene-keyword pair in the list. Multiple annotations to the same term are essentially compressed in this view, in contrast to the Get all GO annotations option. Genes that are annotated to multiple terms that fall into different categories will be included in each of the GO Slim bins. Therefore, the total number of annotations to each aspect of the GO ontologies may be greater than the total number of genes in the query list.

**Displaying the functional classification as a chart**

10. The distribution of functional categories can be displayed graphically as either an annotation pie chart or gene bar chart. To display as a pie chart, above the Functional Category column (Fig. 1.11.9A), select ‘Annotation Pie Chart’ and click on the ‘Draw’ button.” This will create a new page showing three separate pie charts, one for each aspect of the Gene Ontology (Fig. 1.11.9B). Depending on how the results are sorted, the sections can be displayed from most to least frequent category, or by related categories. The percentage of the total is shown in the color key for each graph.

11. To save the graph images, hold down the Ctrl key while clicking on the image, or right click the mouse if using a PC, and save the image to the clipboard or to a file. The images are in Graphic Interchange Format (GIF), which can be opened using a variety of graphics software.

**Downloading the entire set of Arabidopsis GO annotations**

In some cases it may be useful to download the set of Arabidopsis GO annotations for the entire genome. For example, a common use of TAIR’s curated annotations is as a reference for annotation of other species using sequence similarity or homology based methods. In such cases it may be useful to import the Arabidopsis annotations into an analysis tool.


This is the submitted version. For the final, edited version see:

13. Navigate to the ATH.GO_SLIM.txt file. This is a text document is a tab-delimited file containing GO annotations for Arabidopsis genes annotated by TAIR, community, UniProt, the GO Consortium, IntAct, TIGR, and others. The file contains annotations to the narrow (granular) GO term as well as a column that maps the annotations to the corresponding GO Slim category. Users should consult the README file (http://www.arabidopsis.org/download_files/GO_and_PO_Annotations/Gene_Ontology_Annotations/ATH.GO.README.txt) for information on each of the data fields.

**GO Term Enrichment/ Statistical over-underrepresentation test.**

In addition to GO functional categorization, for any given set of genes users may also wish to determine if there are terms that are over or underrepresented in that set as a means to generate hypotheses about gene function or biological events. TAIR uses a web service, provided by PANTHER DB to facilitate GO Term statistical enrichment tests for Arabidopsis and other plants represented in the PANTHER database (http://www.arabidopsis.org/tools/go_term_enrichment.jsp; Mi, et al., 2013). Users can enter a list of locus identifiers, choose the appropriate species, and select the GO aspect (biological process, cellular component or molecular function). PANTHER’s tool accesses a comprehensive list of GO annotations from the GO Consortium as well as a recent version of the ontology itself, both of which are updated monthly. Because annotations are constantly being updated as new information is obtained, the monthly updating schedule ensures that analyses done using the PANTHER tool rely on the most current annotation data.

14. Go to the TAIR home page (http://www.arabidopsis.org), click Tools in the upper menu bar (Fig. 1.11.1), and select GO Term Enrichment from the drop-down menu that appears. Alternatively, go to the URL http://www.arabidopsis.org/tools/go_term_enrichment.jsp

15. Enter in a list of gene identifiers such as AGI Locus IDs (e.g. AT5G61160), UniProt IDs (e.g. Q9FNP9) or NCBI Entrez GeneIDs (e.g. Gene: 836237), separated by newlines or commas.

16. Choose the appropriate plant species from the drop down menu.

The web service implemented at TAIR can be used to analyze Arabidopsis as well as other plant species included in the PANTHER database including: rice, soybean, tomato, moss, Chlamydomonas, poplar, sorghum, grape and Brachypodium,

17. Select the ontology aspect that you wish to analyze. The options are ‘biological process’, ‘molecular function’, and ‘cellular component.’

18. Click Submit, to send the data to PANTHER.

**Evaluating the results**

This is the submitted version. For the final, edited version see:

The web service sends the data to PANTHER and the results are returned in a new window on the PANTHER website (Figure 1.11.10).

19. The analysis summary box (Figure 1.11.10A) displays the analysis type (PANTHER can do several types of gene list analysis), annotation version and annotation dataset. This information is important to record and report in your publications, as the same analysis performed with different software versions and different annotation releases may yield different results.

20. ID mapping results. Uploaded IDs are mapped to the reference proteome set in PANTHER. Click on the number to review each list to see the details.
   a. Unmapped IDs are those that could not be mapped to a corresponding UniProt reference genome protein record in the PANTHER. This list would include any non-protein coding loci.
   b. Multiple IDs. PANTHER also provides a list of IDs where multiple IDs mapped to the same PANTHER protein entry. Typically this occurs because more than one gene produces the same amino acid sequence.

21. Term Enrichment Results (Figure 1.11.10B). The results are displayed in a table.
   a. Term list. The first column displays the over/underrepresented GO terms. By default only results with a p value of less than 0.05 are displayed. The terms are presented in a hierarchical format where related terms are grouped by background color, with the most granular term at the top. Invert the sort order by clicking the arrow next to the term ‘Hierarchy’ in the last column header. To view as a simple list, click ‘Hierarchy’.
   b. The second column shows the number of genes (#) in the reference genome dataset that map to the terms. This is the background frequency.
   c. The third column shows the number of genes (#) in the sample gene set that map to the GO term. This is the sample frequency.
   d. The fourth column displays the number of genes mapped to the term that would be expected based on the whole genome representation. For example if 113/27,352 genes in the reference set mapped to cytosolic large ribosomal subunit, then the expected frequency (0.0041) to map to that term in in the sample set (0.0041 X 247=1.02). Clicking on the number will retrieve a list of the genes that map to the term.
   e. The fifth and sixth columns show the fold enrichment and a sign to show increase (+) or decrease (-). Fold change is calculated by dividing the observed by expected results.
   f. The seventh column shows the p-value. The lower the p-value, the less likely the obtained result can be explained by random distribution.

This is the submitted version. For the final, edited version see:

In an effort to define the function of all *Arabidopsis* genes, the research community, supported by funding agencies, has invested heavily in generating populations of *Arabidopsis* mutants. Different types of mutations provide different information about a gene’s function. TAIR’s database includes mutants generated by chemical mutagenesis and insertional mutagenesis using T-DNA or transposons to generate knockouts and enhancer trap/gene trap constructs that can both disrupt gene function and reveal expression patterns. Other insertional mutagenesis strategies produce overexpression or ectopic expression phenotypes, which are useful when the loss of gene function does not reveal any overt phenotype. TAIR’s Germplasm search can be used to find *Arabidopsis* strains containing mutations in a gene of interest. TAIR works with the ABRC to integrate the ABRC seed and DNA catalogue and purchasing functions.

To order stocks from the ABRC, one must be registered at TAIR and be affiliated with a laboratory as a member or principal investigator. Instructions on how to register online and links to video tutorials can be found in the help documents linked to the online community search and registration forms (http://www.arabidopsis.org/help/helppages/commreg.jsp).

**Necessary Resources**

See Basic Protocol 1

**Finding mutants in a gene**

1. Go to the TAIR home page (http://www.arabidopsis.org), click Search in the upper menu bar (Fig. 1.11.1), and select Seed/Germplasm from the drop-down menu that appears. Alternatively, go directly to the URL http://www.arabidopsis.org/servlets/Search?action=new_search&type=germplasm.

2. The resulting Germplasm Search page offers various search options. In the first section the options include restricting by species (the default option is *Arabidopsis* thaliana, but ABRC has stocks from other species) and then searching by gene name, phenotype, germplasm/stock number, donor last name, construct promoter name, construct reporter name or all. To find mutants in a specific gene, type in the name of the gene (for example, PIN1). Choose to search by “gene name” (the default setting) and for an exact match by selecting “exactly.” Submit the query by clicking the Submit Query button.

   The search allows for up to three name fields to be selected, and each box has a drop-down menu of attributes that can be chosen. When multiple terms are entered, the query is treated as an AND search. Searching with “gene name” set

This is the submitted version. For the final, edited version see:

to PIN1 and “description” set to meristem will search for all Germplasms with a
gene name of PIN1 that contain the term “meristem” in the description. A
common reason why searches fail is because too many restrictive options are
selected. Click on Help next to the Germplasm Search page title to learn more
about all of the search parameters.

3. The resulting page displays a summary of the germplasms matching the query
including a thumbnail image, if one exists. Detailed information about the stock,
polymorphism, and germplasm can be found by clicking on the respective names of
the data in the results summary. If the germplasm is available as a stock from the
ABRC, a check box will appear at the far right of each record.

Another way to find mutants is to use the Polymorphism/Allele search
(http://www.arabidopsis.org/servlets/Search?action=new_search&type=polyalle
le). The results may differ because there are many Polymorphisms without
corresponding Germplasm entries. For example, polymorphisms due to natural
allelic variations (e.g., ecotype differences), published alleles that have curated
by TAIR from the literature, and T-DNA/transposon insertions are loaded into
TAIR and associated to loci, but may not be linked to Germplasm entries. This is
generally the case for T-DNA/Transposon insertion lines that are not available
from the ABRC. Other ways of finding mapped alleles and insertions is to use a
genome browser (i.e. SeqViewer or GBrowse; see Basic Protocol 3) or BLAST
to search the Insertion Flanking sequence data set.

Ordering stocks

4. Examine the germplasm records; click on the name of the germplasm to view specific
details such as phenotype and polymorphism data. If the germplasm is an ABRC
stock, the stock information will be included in the detail page. Select each stock to
order by clicking on the Order from ABRC button.

Users in Europe should order seeds from the Nottingham Arabidopsis Stock
Center (NASC, www.arabidopsis.info), and users not in North America or
Europe may order seeds from either ABRC or NASC. The NASC stock number
can be found in the ‘other names’ field on the germplasm detail page. Use that
identifier to query NASC for the corresponding stock if it exists.

5. Alternatively, stocks can be added directly to one’s “ABRC cart” by clicking the
Order check box for each stock on the results summary page. Once all of the selected
stocks on a page have been chosen, click the “Order Checked Stock” button in the
upper right corner of the Germplasm search results page.

If one is not logged in, one will be prompted to log in, and if one is not affiliated
with a laboratory, one will be asked to affiliate oneself to a laboratory. A
laboratory affiliation is required for placing stock orders. If one does not have a
login (i.e., is not registered), one will be prompted to register. Since registration
typically takes less than 24 hrs. to activate, one should download and save one’s
search results and try placing the order again.
Special considerations for choosing SALK T-DNA stocks

Thousands of laboratories all over the world have availed themselves of the extensive collection of T-DNA insertion lines (Alonso, et al, 2003) generated by the Ecker lab, now at The Salk Institute. The initial insertion lines from the SALK T-DNA insertion collection were identified by recovering flanking sequences as clones, sequencing the plant DNA and mapping those flank sequence tags (FST) to the genome. By convention, those flanking sequences (which are designated as polymorphisms/alleles in TAIR) include the name of the parent strain from which they were isolated. For example the SALK_062236.55.25.x insertion was isolated from the SALK_062236 parent strain. Due to the way that the insertion lines are generated, for any given plant, there may be more than one insertion in the genome, potentially resulting in independent insertions in more than one gene.

Recently, additional insertions have been identified using next generation sequencing and these bear names beginning with SALKseq (e.g. SALKseq_062236.4). To add to the confusion, in some cases, the name of the polymorphism may be the same as the stock. For example SALK_062236 refers to both a germplasm/stock (http://www.arabidopsis.org/servlets/TairObject?type=germplasm&id=4664824; Figure 1.11.12) as well as a polymorphism (http://www.arabidopsis.org/servlets/TairObject?type=polyallele&id=43842). The SALK_062236 polymorphism maps within the AT5G09430 locus. The SALK_062236 stock includes this insertion as well as 7 other T-DNA inserts (SALKseq_062236.0-7 see Fig 1.11.12, boxed area). For example, the SALKseq_062236.4 insertion maps within the coding region of AT2G03450. The progeny line, SALK_062236C, is a purified line that has been selected for the insertion into the AT5G09430 locus and may or may not contain additional insertions Therefore, to obtain the insertion in AT2G03450 the proper stock to order is SALK_062236 not SALK_062236C.

Finding full-length cDNA or EST clones for a gene of interest

TAIR’s database contains records for hundreds of thousands of clones, including full-length cDNA and expressed sequence tag (EST) clones generated by functional genomics projects, many of which are available to the research community from the ABRC or other sources. ESTs generated by sequencing cDNA clone ends may also include the entire coding sequence and can be useful as probes for Northern blotting or in situ hybridization.

6. Go to the TAIR home page (http://www.arabidopsis.org), click Search in the upper menu bar (Fig. 1.11.1), and select DNA/Clones from the drop-down menu that appears. Alternatively, go directly to the URL http://www.arabidopsis.org/servlets/Search?action=new_search&type=dna.

7. From the drop-down menu in the Output Options section, choose “clone” for searching full-length cDNA clones or “clone end” for searching ESTs. In the “Search by” section, choose “GenBank accession” and “exactly” from the respective drop-down menus, and type the accession number (for example, AY040062) in the text box to the right. Click the “Submit Query” button.

The available search options depend on the type of DNA being searched (e.g., clone, clone end, pooled genomic DNA, library, vector, filter, stock, or host strain). The search options for clones and clone ends are the same, but the

This is the submitted version. For the final, edited version see:

results differ. Consult the DNA search help pages
(http://arabidopsis.org/help/helppages/dnasearch.jsp) for information on the
specific parameters and how to use them.

8. If the clone is available as a stock from the ABRC, check the order box in the last
column and then click on the Order Checked Stock. If the clone is not available as a
stock or is not in stock, click on the link from GenBank accession number to go to
the GenBank record. This record should have information for the donor of the
sequence, who can be contacted to obtain the clone directly.

If not sure of the exact stock to look for, or simply to find out what is available,
one can browse the ABRC stock catalog
(http://www.arabidopsis.org/servlets/Order?state=catalog). The catalog is
organized into different categories of DNA and Seed stocks. Catalog entries are
linked either to the relevant stock pages or to information about the class of
stocks and how to order them.

BASIC PROTOCOL 6

USING GENE LISTS TO DOWNLOAD BULK DATASETS

TAIR provides a number of tools for obtaining data in bulk for sets of genes such as gene descriptions or
sequences (http://www.arabidopsis.org/tools/bulk/index.jsp). While the gene search and locus pages can
provide comprehensive information on a gene by gene basis (see Basic Protocol 2), it is often desirable to
obtain specific data for a large number of genes. TAIR’s bulk download tools can be used to take a set of
AGI locus identifiers as an input and obtain gene descriptions
(http://www.arabidopsis.org/tools/bulk/genes/index.jsp), GO annotations (see Basic Protocol 4) and PO
annotations (http://www.arabidopsis.org/tools/bulk/po/index.jsp), sequences
(http://www.arabidopsis.org/tools/bulk/sequences/index.jsp), protein properties
(http://www.arabidopsis.org/tools/bulk/protein/index.jsp), microarray elements
(http://www.arabidopsis.org/tools/bulk/microarray/index.jsp) and locus histories

Necessary Resources

See Basic Protocol 1

Downloading Gene Description/Summaries

1. On the TAIR home page (http://www.arabidopsis.org) select Bulk Downloads from
the Tools drop-down menu. Alternatively, go directly to the URL
3. Enter in or upload a list of AGI locus identifiers or gene model identifiers.
4. Choose which data set to search against to retrieve matching records. To obtain all

This is the submitted version. For the final, edited version see:

information about Arabidopsis genes. Current Protocols in Bioinformatics, 60, 1.11.1–1.11.45. doi: 10.1002/cpbi.36
descriptions for a locus, choose ‘get descriptions for all gene models/splice forms.

5. Choose how you want your results returned, to the browser or in a file.

**Guidelines for understanding the results**

6. The results will include the locus identifier, gene model name(s), description, primary gene symbol and other gene symbols. *Each locus may be associated to one or more gene models, and each model may have distinct descriptive information that is unique for that gene product. For example the locus AT2G42810 (http://www.arabidopsis.org/servlets/TairObject?id=33349&type=locus), encoding Protein Phosphatase5 (PP5) has a total of 5 gene models which represent different splice variants. The AT2G42810.2, or reference gene model, is an integral membrane protein whereas AT2G42810.1 does not contain the membrane domains and is localized to the cytoplasm.*

The gene description will either be a short, computationally derived description statement attributed to Araport 11, or a curated summary written by TAIR curators.

**Downloading whole genome annotations in bulk**

In some cases, it may be desirable to obtain data sets for the entire Arabidopsis genome such as sequences or functional annotations. TAIR provides access to curated data (e.g. PO annotations, phenotypes, gene summaries, gene aliases, etc.) after the data have been in TAIR for one year. Year old data is released on a quarterly basis and can be found in the Download section (see Introduction) under Public Data Releases. Subscribers can access more recent data sets from the Download section (Subscriber Data Releases). Sequences from TAIR10 and Araport11 are available as BLAST data files (http://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FSequences).

---

**BASIC PROTOCOL 7**

**USING TAIR’S ANALYSIS TOOLS TO FIND SHORT SEQUENCES AND MOTIFS**

**Using the Motif Analysis Tool for Identifying potential cis-regulatory motifs in upstream sequences**

The Motif Finder identifies six-oligomer nucleotide sequences that are statistically over-represented in a set of input sequences when compared to the whole genome. The most common application of this tool is for identifying potential cis-regulatory elements in genes whose

This is the submitted version. For the final, edited version see:

expression patterns correlate into a cluster. Consensus sequences for putative transcription factor binding sites can be used to identify additional genes having the element in the promoter using the PatMatch program (see Commentary).

Necessary Resources

See Basic Protocol 1

Entering the search parameters


2. On the data entry form enter the locus identifiers of your genes of interest. This can be done manually or by uploading a list of Arabidopsis AGI identifiers from a file. In the example shown in Figure 1.11.12A, we queried for motifs in the 500 bp upstream region of 15 genes co-expressed genes. Note that a minimum number of 3 locus identifiers has to be entered.

3. Select length (500, 1000, or 3000 bp) of upstream sequence to be queried. Submit the query.

   *The sequence data sets are either 500-, 1000-, or 3000-base-pair sequences upstream of the translation start site of each gene in the genome. The program will search for 6-mer words that are overrepresented in the upstream regions of the set of queried genes compared to upstream sequences in the entire genome. Both forward and reverse strands are queried.*

Evaluating the results

4. The results are displayed in a table as shown in Figure 1.11.12B. The columns, denoted “a” through “g” in Figure 1.11.12B, are as follows.

   a. Oligomer.

      *Each over-represented six-oligomer sequence is listed in the first column of the results table.*

   b. Absolute number of oligos in the query set.

      *Number of times the oligo appears in the upstream regions (of chosen length) of the query genes. This number can be higher than the number of query sequences, as some sequences contain multiple occurrences of the motif.*

   c. Absolute number of oligos in the genomic set.

      *Number of times the oligo appears in the upstream sequences (of chosen length) of all genes in the genome.*

   d. Number of sequences in query set containing oligomer.

      *Shows the ratio of the number of queried sequences containing the oligomer over the total number of queried sequences.*

This is the submitted version. For the final, edited version see:

e. Number of sequences (out of 33,602 in genomic set) containing oligomer.

*Shows the ratio of the number of genome sequences containing the oligomer over the total number of sequences in the genome.*

f. \( p \)-value.

*This score reflects the probability of the six-oligomer sequence occurring in the selected query set by chance. The lower the score (closer to zero) the greater the likelihood the match is significant.*

g. Query sequences containing this oligomer.

*All of the query genes containing the oligomer are listed here. The PatMatch tool (see next section) can be used to locate other genes that contain the oligomer in the upstream sequence.*

**Using the PatMatch Tool to find short sequence patterns in DNA and protein sequences**

PatMatch (Yan et al., 2005) was designed for identifying patterns in a selected TAIR dataset (e.g., genes, proteins, upstream sequences, etc.) that match regular expressions. PatMatch can be useful for finding short nucleotide patterns such as *cis*-elements, Massively Parallel Signature Sequence (MPSS), Serial Analysis of Gene Expression (SAGE) tags, or small RNA binding sites. Patmatch can also be used to search for motifs in protein sequences. Other options of this tool include the selection of a target data set, strand to be queried (in case of nucleotide search), and number of results to retrieve. If one needs to process large amounts of data or increase the number of results to be included, it users can download the PatMatch1.1 program ([http://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FSoftware%2FPatmatch/](http://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FSoftware%2FPatmatch/)) and run it locally on a Unix-based system. The BLAST data sets used by PatMatch can also be downloaded from the TAIR site ([http://www.arabidopsis.org/download/index-auto.jsp?dir=/download_files/Sequences](http://www.arabidopsis.org/download/index-auto.jsp?dir=/download_files/Sequences)).

**Necessary Resources**

See Basic Protocol 1

**Entering the search Parameters**

1. From the TAIR home page ([www.arabidopsis.org](http://www.arabidopsis.org)) select Patmatch from the Tools drop down menu. Alternatively go directly to [https://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl](https://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl).

2. Enter in a query pattern and the appropriate option for a DNA or Protein search from the drop down menu. Acceptable inputs include regular expressions that include mismatches, insertions, and deletions, and apply standard IUPAC notation to indicate ambiguous sequences. Supported syntax formats are displayed on the bottom of the data entry page.

3. Choose a sequence dataset to search. The program uses the same target data sets as TAIR’s BLAST software.

**Evaluating the results**

This is the submitted version. For the final, edited version see:

4. Patmatch does not generate alignments or provide scores for best hits. The results are displayed in a table format that includes the following information.
   a. **Sequence name:** Name of the gene or sequence for which a hit was found.
   b. **# of hits:** Number of times the query pattern was found in that specific sequence.
   c. **Hit pattern:** Pattern used for the query.
   d. **Matching positions:** Start and end position of the hit. These coordinates are always relative to the sequence (e.g., gene, upstream region, intergenic region).
   e. **Hit sequence:** Hyperlink to the sequence for which a hit was found. The pattern match is highlighted in red letters. For nucleotide searches, coordinates shown here are always relative to the chromosome.

**BASIC PROTOCOL 8**

**USING THE TAIR ONLINE ANNOTATION SUBMISSION TOOL (TOAST) TO SUBMIT FUNCTIONALANNOTATIONS FOR ARABIDOPSIS GENES**

In order to maximize the capture of experimental information about gene function from the literature and from our expert community, TAIR developed TOAST to provide a quick and easy way for authors or readers of an article to contribute information about Arabidopsis gene function from the published literature (Berardini, et al, 2012). The tool allows registered users to submit their own GO and PO annotations, and comments. The software is ‘article based’ meaning it is designed to annotate genes with experimental data from a specific paper, rather than a tool to annotate a specific gene. Users are not restricted to annotating their own papers, any published article can be curated in TOAST. Alternatively users can download a preformatted Excel spreadsheet and email annotations to TAIR.

**Necessary Resources**

See Basic Protocol 1

**Submitting Annotations**


2. Scroll to the center of the page and click the button to “Fill Online Form”. Alternatively go to [https://toast.arabidopsis.org/](https://toast.arabidopsis.org/).

   Before entering data, ensure that there is ample time to complete and submit the form. TOAST does not save work in progress. If it is necessary to stop or close the browser, it is best to submit whatever annotations have been completed to avoid losing the data.

This is the submitted version. For the final, edited version see:

3. Enter your TAIR username and password. Users must be registered with TAIR in order to submit data via TOAST.

4. On the resulting data entry form (Fig 1.11.12) enter the PubMed ID or DOI for the article to annotate.

   TAIR only displays annotations from published works. Users may submit annotations for articles that have been accepted for publication that have received a temporary DOI but TAIR does not accept annotations for unpublished work.

5. Enter the locus identifier of the first gene to annotate. Start by typing in the name and the autocomplete function will suggest possible completions. At this point enter in any symbolic name(s) to associate to the locus.

6. Enter the annotations. The form contains separate sections for each type of annotation (GO Molecular Function, GO Process, Expression). At least one annotation must be entered order to be able to submit (otherwise the Submit button remains grayed out).

   a. Enter the term in the left column. The auto-suggest function will offer a list of suggested terms. Choose one of the suggested terms, or if none is appropriate, enter a new term. A TAIR curator reviews all the contributions and will approve the annotation, update to find the best term that matches, or determine if a new ontology term needs to be added.

   b. Enter the supporting evidence in the right column. All annotations must be backed up by evidence. Choose the evidence type from the dropdown menu that most closely fits the experimental method.

7. To add additional annotations for any category click on the small green plus sign to the right. For example to add an additional molecular function annotation click the plus sign and a new data entry row will appear. To delete an annotation, click on the red X to the right of the annotation row (Fig. 1.11.12a).

8. Enter comments. At the bottom of the form there is a section to enter comments that may include information that cannot be captured in a GO or PO annotation. This section is optional.

9. Submit the annotations or add another gene. Once all of the annotations for the gene entered in step 5 are done, either submit the annotations or annotate another gene described in the same paper.

   a. To add another gene, go to the top of the form and click on the plus sign in the upper right corner (Figure 1.11.12b). This will append a new entry form to the bottom of the page.

   b. To submit annotations, click on the Submit Annotations button on the lower left side of the page.

10. Once the data is submitted a curator will review the submission. If there are any questions, a curator will contact the submitter. It can take a week or two before the data is visible in TAIR.

This is the submitted version. For the final, edited version see:

Other ways to submit data/corrections to TAIR

One of the most fundamental aspects of science is sharing data and results with the research community. The fruits of research drive new areas of discovery, and funding agencies, such as the National Science Foundation (NSF), have invested heavily in developing community resources. Web sites and databases such as TAIR make these data accessible to anyone connected to the Internet. The long-term sustainability of databases will increasingly rely upon contributions by the research community (Reiser et al., 2016; Leonelli, et al, 2017).

TAIR encourages feedback and data submission and provides several ways for researchers to contribute their expertise and data. Instructions for submitting various types of data including gene function, interaction partners, expression patterns, markers, phenotypes, and several others, are available on the Submit Overview page (http://arabidopsis.org/submit/index.jsp), accessible from the Submit drop-down menu in the top navigation bar. Users can prepare data formatted according to the guidelines or download and use the preformatted Excel spreadsheets. The spreadsheets may contain macros that ensure that the proper data formats are used. To use the spreadsheets, macros must be enabled. TAIR will also accept direct submissions by email to curator@arabidopsis.org for small datasets and corrections to existing data, as well as very large datasets and those requiring special formats. Please contact us with any questions about data submission.

In addition, each data detail page includes a Community Comments section where community members can add additional information; click on the comment text to view the entire comment. Registered users can submit comments that are then immediately displayed in the Comments section of the detail page. On-line instructions for submitting comments are available at http://arabidopsis.org/help/helppages/addcomment.jsp.

BASIC PROTOCOL 9

USING TAIR TO BROWSE ARABIDOPSIS LITERATURE

TAIR provides a number of ways for researchers to keep abreast of the literature. In addition to the curated links between genes and articles that are displayed on the locus detail pages (see Basic Protocol 2), the entire corpus of publications in TAIR (including abstracts and conference proceedings) can be searched using the Publication Search (http://www.arabidopsis.org/servlets/Search?action=new_search&type=publication) or Keyword browser (http://www.arabidopsis.org/servlets/Search?action=new_search&type=keyword). For users wanting to keep up with the latest Arabidopsis research, TAIR developed an additional tool for browsing recently added literature.

Necessary Resources

See Basic Protocol 1
Browsing Recently Added Literature


2. The page will display a list of the research articles downloaded from PubMed and entered into TAIR database during the time period specified in the header. Typically this is a two-week period. Each article is displayed in a separate band of alternating background color. The contents of each band include basic citation information and links to associated resources.

   The default display is sorted alphabetically by journal name. Choose author name from the drop down selector in the upper right corner to display the results in alphabetical order by the last name of the first author.

a. Citation. The citation includes the authors, title, journal name, and publication year.

b. Associated genes (may be empty). These are manually curated links to genes described in the paper. Clicking on the Gene name will display the corresponding locus detail page in TAIR where you can find more information about the locus (see Basic Protocol 2).

c. Associated Keywords (may be empty). Keywords are generated by automatic text matching of GO terms to the text and are not curated. To find other objects in TAIR associated to that keyword, click on the term to display the keyword detail page (see Basic Protocol 4).

d. Article views. Three options are provided that offer different views of the article.
   i. Click on the Journal link to read the article on the journal’s website.
   ii. Click on PubMed link to view the corresponding record and abstract at NCBI’s PubMed site.
   iii. Click on the TAIR link to view the publication detail page in TAIR. The TAIR publication page displays the citation, which may include the abstract, as well as associated keywords, loci and GO and PO annotations. The annotations and linked loci are manually curated.

This is the submitted version. For the final, edited version see:

GUIDELINES FOR UNDERSTANDING RESULTS

General Considerations for Using TAIR

As with any Web-based resource, some general guidelines should be observed when interpreting results. Databases are constantly changing; new information is incorporated and interfaces can also change from the time of publication of this unit.

Revisions to the Data in the Database

Over the course of genome annotation, many new genes have been added and existing genes have been made obsolete or updated (split or merged) to reflect new information (Haas et al., 2003; Swarbreck et al., 2008, Lamesch, et al., 2010; Cheng, et al., 2017). In 2004, TAIR inherited the responsibility of maintaining the genome sequence and annotation from the former Institute for Genomic Research (TIGR), now the J. Craig Ventner Institute (JCVI), which provided the genome sequence and annotation from 2000 to 2004. TAIR produced five genome releases culminating in TAIR 10 (Lamesch, et al., 2010). The Arabidopsis Information Portal (Araport), which subsequently took on the responsibility of genome annotation, released Araport 11 in 2016 (Chang et al., 2017). The naming convention agreed upon by the AGI for adding new loci and updating existing loci (http://www.arabidopsis.org/portals/nomenclature/guidelines.jsp) is continued in TAIR and Araport releases. Users are encouraged to submit structural annotation updates to Araport and to provide the relevant sequence to GenBank (http://www.arabidopsis.org/submit/gene_annotation.submission.jsp). Changes in sequence annotation may affect the association of genes to related data such as protein domains, polymorphisms, and homologies. For example, domains associated to a locus that was subsequently split may then be associated to only one of the two resulting loci. The locus history, shown on the bottom of the locus page (Fig. 1.11.2, o item), summarizes all of the changes that had been made to the locus. The locus history can also be searched independently by locus name using the Locus History Search (http://www.arabidopsis.org/tools/bulk/locushistory/index.jsp). For many data sets in downloads section of the website, TAIR maintains older versions of the data. Users should always note the date or version information associated with any data files, such as BLAST data sets or GO annotations.

Evidence Codes in GO Annotations

When interpreting Gene Ontology annotations, it is essential to understand the process of annotation and the importance of evidence codes in interpreting the annotations. The GO Consortium has developed a set of evidence codes (The Gene Ontology Consortium, 2010) as a way of quickly assessing the basis for the assertion made in the annotation. In TAIR, annotations include an evidence description, in addition to the evidence code (Berardini et al., 2004). The evidence description is a set of controlled vocabularies that describe the type of experimental or computational evidence in greater detail. For example, an annotation having the evidence code “inferred from mutant phenotype” (IMP) may be further specified by including more specific information about the type of experiment done such as “RNAi experiments.” Since more than one gene may be affected by RNA interference, the phenotype may be due to changes in expression...
of multiple loci. Thus the GO annotation should be viewed with the understanding that the phenotype may be due to the loss of function of more than one homologous locus. When no information is found in the available published literature, annotations are made to the root terms “biological_process,” “molecular_function,” or “cellular_component.” Such “root” annotations indicate that at the time of annotation, no information for a more specific assignment was available for the associated gene. In contrast, a gene lacking annotations altogether might have available data but has not yet been curated. At TAIR, GO annotation is an ongoing process; annotations are updated as new information about genes is published (Berardini et al., 2004). Each annotation has an associated date, which refers to the date the annotation was made. Users should carefully evaluate the source of any tools utilizing GO annotations (e.g. Term Enrichment) to ensure that the underlying annotations are current.

COMMENTARY

Background Information

TAIR was originally a collaborative project between biologists at the Carnegie Institution, Department of Plant Biology, and computer scientists at the National Center for Genome Resources, initiated in 1999. TAIR is the third incarnation of an Arabidopsis community database after AAtDB (An Arabidopsis thaliana Database, which continued from 1991 to 1994) and AtDB (Arabidopsis thaliana Database, which continued from 1994 to 1999; Flanders et al., 1998; Rhee et al., 1999). TAIR arose out of the need to accommodate genomic data such as the genome sequence, gene annotations, and integration of physical and genetic maps, in the context of the experimentally verified data in the literature. From its inception until early 2014, the National Science Foundation (NSF) funded TAIR. In late 2013, anticipating the end of NSF funding, four TAIR staff members founded the nonprofit organization Phoenix Bioinformatics (www.phoenixbioinformatics.org) and transitioned TAIR to a new, sustainable user fee model (Reiser, et al., 2016). The user fee structure was carefully crafted to distribute the costs equitably among the widespread and varied user community. With the support of the research community, TAIR continues to provide up to date, continuously curated data to its global users. Curated data is updated weekly. Data that have been in TAIR for one year are released on a quarterly basis and available for download and reuse (http://www.arabidopsis.org/download/index-auto.jsp?dir=/download_files/Public_Data_Releases). Users without a subscription can access a limited number of page views per month. Unlimited access to all TAIR pages and quarterly releases of recent data requires a subscription. ABRC stock detail and ordering pages are available free of charge. The data release policy was designed encourage and support data reuse, while still proving an incentive to subscribe. More information on how institutions or individuals can help support this nonprofit effort is available by clicking on the Subscribe button on the TAIR home page.

Design principles and current limitations

TAIR has been designed and built as a Web tool to allow researchers to access all of the data housed in TAIR using a standard Web browser such as Google Chrome or Mozilla Firefox. It is built upon industry standards for database management systems, software architecture, and software design (Weems et al., 2004). TAIR is primarily designed by biologists and, although the interfaces were created with biologists in mind, it has not always been possible to arrive at solutions that meet every user’s requirements. A certain amount of familiarity with Arabidopsis

This is the submitted version. For the final, edited version see:

and with basic concepts of molecular genetics and plant biology is assumed. Consequently, the breadth of information on the home page and myriad options on the search interfaces can be daunting to a novice user. More experienced users and developers may be frustrated by the difficulty in obtaining the entire database for retrieving specialized, custom data sets. Users are encouraged to contact us via email (curator@arabidopsis.org) for assistance in using any of the tools or in accessing large or specialized datasets.

**Keeping up to date with TAIR and Arabidopsis research**

There are two ways to keep updated on TAIR and news about the *Arabidopsis* research community. Registered users can choose to receive a occasional e-mail newsletters from TAIR that describes significant new or updated data and tools. In addition, researchers that use *Arabidopsis* or *Arabidopsis* data frequently can subscribe to the newsgroup (http://arabidopsis.org/news/newsgroup.jsp), which is a moderated mailing list through which meetings, jobs, and new data/tools are announced and where researchers post problems and get feedback.

Users can also stay connected with TAIR by becoming a fan of the *Arabidopsis* Information Resource on Facebook (http://www.facebook.com), or receive tair_news twitter feeds (http://twitter.com/tair_news) or YouTube channel alerts (https://www.youtube.com/user/TAIRinfo). TAIR News and Job Postings are relayed through the TAIR Twitter feed.

**Additional tools at TAIR**

In addition to the tools discussed in the protocols, TAIR hosts several other useful analysis tools. Some of these are briefly described below.

**Synteny viewer: GBrowse_Syn**

GBrowse_syn is a GBrowse-based synteny browser designed to display multiple genomes, with a central reference species compared to several additional species. It is included with the standard GBrowse package (version 1.69 and later). GBrowse_syn was built to help researchers study and analyze syntetic regions, homologous genes, and other conserved elements between sequences. It can also be used to study genome duplication and evolution. By comparing newly sequenced or less studied genomes to the well-annotated *Arabidopsis* genome in GBrowse_syn (http://gbrowse.arabidopsis.org/cgi-bin/gbrowse_syn/arabidopsis/) scientists can identify novel genes and putative regulatory elements.

The current version of the GBrowse_syn tool at TAIR includes the genomes of *A. thaliana*, *A. lyrata*, and *P. trichocarpa*. The *A. lyrata* and *P. trichocarpa* alignment data were provided to us by Pedro Pattyn from the University of Ghent.

**Textpresso**

Textpresso is an information extracting and processing package for biological literature. Textpresso for *Arabidopsis* (http://www.textpresso.org/arabidopsis) allows users to search all abstracts and over 40,527 full-text publications in TAIR. Keyword searches can be narrowed by searching in specific categories. Textpresso was initially developed by Hans-Michael Muller, Eimear Kenny, and Paul W. Sternberg, with contributions from JuanCarlos Chan and David

This is the submitted version. For the final, edited version see:

Chen. The most recent version, Textpresso 2.0, was developed by Hans-Michael Muller with contributions from Arun Rangarajan and Tracy K. Teal.

Chromosome map tool

This tool (http://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp) allows the user to map genes on top of the five Arabidopsis chromosomes using a list of locus names (e.g., At1g01010). The list should contain one locus name per line. To display an alternate name, append the symbol after the locus identifier in the same row (e.g. AT1g01010 ANAC001). The resulting image, which displays the location of the queried list of genes on the five chromosomes, can be saved in a variety of formats.

Critical Parameters and Troubleshooting

No data found

A frequently reported problem is that searches do not retrieve any results. In some cases the data sought are not in the database, but in other cases the data are in TAIR but are not found because of problems arising from poorly formed queries or improper use of the search forms. The temptation to fill out all of the optional fields in the advanced searches can generate too many restrictions that limit the scope of the data retrieved. This can be overcome by using fewer, rather than more options. Another reason why searches fail is that the data are not accessible through the existing search interfaces. The categories under the Advanced Search section of the Web site (http://arabidopsis.org/servlets/Search?type=general&action=new_search) list data types that can be searched. To obtain that are included in the TAIR database but are not easily accessible through any of the advanced searches, please send an e-mail to the curators to request the data.

Too much data found

While “no data found” is probably the most common problem encountered, retrieving too many results can also be a problem. There are two ways to handle this problem: (1) using the advanced searches and restricting parameters to retrieve a subset of the results, or (2) manipulating the results set to select a subset of data. Restricting the search parameters can be done on all the Advanced Search pages and detailed help on using these parameters is available (http://arabidopsis.org/help/helpcontents.jsp). Large results sets can be downloaded and reformatted to explore the data more efficiently. All of the search results can be downloaded as tab-delimited text files (see Basic Protocol 2). The results can be imported into software like Excel or Google Spreadsheets that allows manipulations such as sorting, reordering, reformatting columns, and graphing the results.

Layers of connected data that are hidden

TAIR’s database structure exploits the relational database design and each data type has a high degree of association to other data types. This network of associated data is not easily represented in a two-dimensional, tabular format via hyperlinks. Consequently, associated data may be separated by one or more hyperlinks. For example, all gene models are associated to a given locus, but to view information for a specific gene model, such as a list of gene features (e.g. introns, exons, UTRs) and coordinates, it is necessary to click on the link to the individual gene model.

This is the submitted version. For the final, edited version see:

**Reporting problems and requests to curators**

Perhaps the most important thing to know about troubleshooting problems with TAIR is that users are encouraged to e-mail curators (curator@arabidopsis.org) to report problems, ask for help or request data. Users that want a particular set of customized data are welcome to contact the curators, who can then generate the requested file. Reporting problems may also lead to improvements to TAIR’s data display or addition of new data or tools that benefit the whole community.

**Advanced parameters**

Despite the extensive content of this unit, it still does not cover all of the functionalities of the searches and tools that are offered at TAIR. Users familiar with the basic functionalities and who are interested in more complicated queries or specialized views are encouraged to review the help documents or contact the curators.

**Suggestions for Further Analysis**

While there are no complete alternatives to TAIR, there are other Web sites that provide a significant amount of *Arabidopsis* data and alternative ways of to view, manipulate and analyze the data. All of these sites are linked extensively from TAIR, whereby the sites in the former category are linked from each locus detail pages and sites in both categories are listed and updated in the TAIR Portal pages (http://www.arabidopsis.org/portals/index.jsp). Many of these resources integrate TAIR curated functional annotations (e.g. gene summaries, names, literature) however, in accordance with TAIR’s data release policy, the data on these sites will be at least one year out of date.

**Arabidopsis genome annotation resources**

Araport (www.araport.org) is an online resource for Arabidopsis and has taken over the responsibility for genome re-annotation from TAIR (Krishnakumar, et al., 2016; Chang et al, 2017). Users can search, download and analyze Arabidopsis genome data with Araport’s InterMine instance called ThaleMine and browse the Araport11 annotated genome in JBrowse. Another view of the Arabidopsis genome can be found as one of the databases in Ensembl Plants (http://plants.ensembl.org/Arabopsis_thaliana/Info/Index). Ensembl provides its own genome browser for visualization, as well as plant gene families generated using Compara. Users can access variant data for Arabidopsis ecotypes generated by the 1001 Genomes project (http://1001genomes.org/) within Ensembl. Experienced users will find it useful to be able to generate their own custom datasets using Ensembl’s BioMart or Araport’s ThaleMine. SIGnAL (http://signal.salk.edu/) from the Salk Institute offers a genome viewer (T-DNA Express, http://signal.salk.edu/cgi-bin/tdnaexpress) that is decorated with all of the T-DNA and transcript data that are generated from Salk and other laboratories around the world. Often, SIGnAL displays data that are not yet displayed at TAIR; therefore, it is a good idea to check this site to get the latest mapping of T-DNA insertions and cDNA clones. AtGDB (*Arabidopsis thaliana* Genome Database, http://www.plantgdb.org/AtGDB/) offers another view of the genome that has been annotated with their gene-prediction algorithms. For most genes, the description and exon-intron structures of genes in these sites are identical. However, in a small number of cases, there are genes that have different descriptions and/or structures because of the differences in the methods of annotation and interpretation of the evidence. Users should pay attention to the

This is the submitted version. For the final, edited version see:

Figure 1.11.12  Motif Finder tool. (A) Users can type in or upload a list of genes and select the promoter length to be analyzed. (B) The resulting motifs are listed with the corresponding genes in which they are found.

Figure 1.11.13 TOAST data entry page. Users can add additional fields for each aspect by clicking the plus sign (box a) or add another gene (box b).
evidence shown for the gene structures and make their own interpretation of the structure and function of the genes. There are also many sites that provide detailed information about a subset of genes of Arabidopsis such as chromatin remodeling factors, transcription factors, and small RNAs. TAIR tries to maintain up-to-date links to these resources from the TAIR Portal pages. Please contact TAIR by e-mail (curator@arabidopsis.org) if there are missing or nonfunctional links.

**Arabidopsis Microarrays and other Gene Expression Resources**

There are a number of databases and tools that have been developed for storing, accessing and analyzing public Arabidopsis expression data. TAIR stopped accepting microarray data in 2005 as ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) and the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) emerged as centralized community repositories. TAIR still provides access to the data via the Microarray Experiment (http://www.arabidopsis.org/servlets/Search?type=expr&search_action=new_search) and Microarray Expression searches (http://www.arabidopsis.org/servlets/Search?action=new_search&type=expression) for archival purposes. In addition to the data and tools available at ArrayExpress and GEO, The BioAnalytic Resource for Plant Biology (http://bar.utoronto.ca/) hosts a number of user friendly tools for visualizing and analyzing Arabidopsis expression data, motif analysis and more. It provides a virtual graphical gene expression map (eFP browser;http://bar.utoronto.ca/eFP/cgi-bin/eFPWeb.cgi) an Expression Angler tool which can be used to find similarly expressed genes (http://bar.utoronto.ca/ExpressionAngler/) and Expressolog TreeViewer (http://bar.utoronto.ca/expressolog_treeviewer/cgi-bin/expressolog_treeviewer.cgi) for finding expression orthologs. Another popular tool is GENEVESTIGATOR (https://genevestigator.com/), which contains most of the publicly available high-density array data from AtGenExpress (http://arabidopsis.org/info/expression/ATGenExpress.jsp) and other laboratories, and allows searching and displaying of the data (Zimmermann et al., 2004). Academic users must create a basic account, after which they can search for genes that are expressed in specific conditions, growth stages, or organs, or for genes of particular interest to them, and get a comprehensive view of the expression profiles in the different environmental conditions, growth stages, and organs. GENEVESTIGATOR requires subscriptions to access additional data and tools.

**Arabidopsis Metabolic Pathways**

There are several excellent resources for visualizing, analyzing and accessing information about biochemical pathways in Arabidopsis and other species. AraCyc is a curated metabolic pathway database specifically for Arabidopsis thaliana and is included in the Plant Metabolic Network (PMN, www.plantcyc.org) database. PMN includes over 350 plant species. The AraCyc database was initially built using the Pathologic module in the Pathway Tools software developed for MetaCyc (Karp et al., 2002; Mueller et al., 2003). Pathologic predicts possible metabolic pathways based upon the set of annotated enzymes available for a particular species. Following the initial computational build of AraCyc, pathways were manually validated and some were supplemented with additional experimental evidence. AraCyc includes tools to search and browse metabolic pathways drilling down to individual reactions and products, ways to visualize gene expression data superimposed on a global pathway map, and options to save data in Smart Tables. Unification links to MetaCyc and PlantCyc facilitate comparison of pathways from different organisms. TAIR includes extensive links out to AraCyc from the locus pages (see Basic Protocol

This is the submitted version. For the final, edited version see:

Another resource is the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/kegg2.html) that includes pathways, reactions, enzymes, genome and other information for Arabidopsis and many other plant species. Plant Reactome (http://plantreactome.gramene.org; Naithani, et al., 2017) contains information about biochemical, genetic and other pathways, and tools for data visualization and analysis. Reactome curates pathways for the reference genome Oryza sativa, which is presented along with data for many other species, including Arabidopsis.

Acknowledgments

The authors of this unit are grateful for the continued support of members of the research community who share their expertise, ideas, data and criticisms, all of which improve TAIR immensely. We thank Dr. Tanya Berardini for helpful comments, along with all of the curators and programmers past and present who helped make TAIR such a valued resource. This work is supported by individual, institutional, corporate and government subscriptions. TAIR is a project of Phoenix Bioinformatics (www.phoenixbioinformatics.org), which is supported, in part, by a grant from the Alfred P. Sloan Foundation.

Literature Cited


This is the submitted version. For the final, edited version see:


This is the submitted version. For the final, edited version see:

This is the submitted version. For the final, edited version see:


**Figure Legends**

**Figure 1.11.1** TAIR’s home page ([http://arabidopsis.org](http://arabidopsis.org)) is the main entry point to the database and Web site.

**Figure 1.11.2** A sample of a locus page from TAIR showing the major data included in the detail page. A portion of the germplasm section has been deleted for simplicity. Each of the data types displayed in the alternating colored bands can be grouped into one or more the following categories: (A) (a) general descriptive locus information, (b) gene model information, (c) functional annotations, (d) nucleotide and protein sequences (e) gene expression data,(f) protein data, (g) gene family data, (h) mapping data, (i) polymorphisms and alleles, (j) germplasm information;(B) (k) clones, (l) links to resources outside of TAIR, (m) community comments about the locus, (n) papers and abstracts and (o) locus history.

**Figure 1.11.3** SeqViewer home page after submitting the gene name AT1G07810 as a query term. The five nuclear chromosomes are shown as green lines with blue boxes indicating the location of the centromeres (a). A few markers are included as landmarks for orientation. Queries can be typed, pasted, or uploaded into the text input box (b). The available options include searches by name or sequence. The number of matches is displayed above the chromosomes (in this example this number is 1) and is hyperlinked to a list of results. Each match to the genome is indicated with a red tick mark on the chromosomes; clicking on the mark will open a detailed Close-up view. The Close-up View options (c) are used to select the zoom level and types of objects to display in the detailed view.

**Figure 1.11.4** A 10-kb region of chromosome 1 centered on the AT1G07810.1 gene, which is highlighted in yellow. (a) The area of the genome shown in the Close-up view is indicated by the This is the submitted version. For the final, edited version see:
numbered box in the whole genome view. (b) The radio button for selecting three or all rows of data to display in the Close-up view. (c) The gray re-centering bar. (d) The gray bar between the T-DNA and gene bands is used for selecting a 10-kb region to display in the nucleotide sequence view. (e) A pop up window displaying summary information for the selected data object.

Figure 1.11.5 A nucleotide sequence view centered on AT1G07810.1 showing annotated genes and T-DNA/transposon insertion flanking sequences. The drop-down menu (shown in upper right corner) was used to select the items to display in the nucleotide sequence view.

Figure 1.11.6 Overview of the GBrowse tool. The upper panel of this tool has many options and controls to allow the users to (a) set preferences and tracks, (b), search for landmarks in the genome, (c) choose a data source, (d) configure displays and, (e) scroll and zoom in and out of a specific genomic region. The Details view (items f-h) orients the viewer within the whole chromosome context and displays a series of data tracks (*) such as genes, cDNAs, polymorphisms, the VISTA sequence similarity track and many more.

Figure 1.11.7 Gbrowse Tracks Selection View. (A) Using the Select Tracks menu, the user can choose which tracks to display by selecting either whole data categories or specific types of data. (B) The Custom Tracks landing page where users specify custom tracks to display in GBrowse.

Figure 1.11.8 Keyword results and detail page. (A) Keyword search results after querying for the GO Biological Process terms containing the words “root development”. (B) A tree view of the term “root development” and associated gene annotations.

Figure 1.11.9 (A) Results display for functional categorization of WRKY genes. The members of this family fall into 24 different GO Slim categories based on their annotations to more granular GO terms. The list can be re-sorted by choosing Gene count from the “re-sort by:” drop-down menu and clicking on the “re-sort” button. The list of 24 categories is shown grouped by keyword category. The frequency of annotations to each category is listed in the last column; the number is linked to a list of genes annotated to the terms that are children of that category. (B) The drop down menu is used to select a graphical output format showing the distribution and frequency of annotations to each of the GO slim terms as either a bar graph or pie chart. A different graph/chart is created for each aspect of the GO ontologies.

Figure 1.11.10 GO Term Enrichment Tool. Results display for sample data from query input form, after it has been received and processed via the PANTHER web service. Results display (A) query parameters and identifier mapping, and (B) hierarchcal ordered table of term enrichment results.

Figure 1.11.11 Stock/Germplasm Detail. Known insertions/polymorphisms in the Associated Polymorphism section. Click on the individual SALKseq insertions links to find information about the insertion including affected loci and insertion sites. Related stocks such as parents and progeny are displayed in the Pedigree section.

This is the submitted version. For the final, edited version see:

The Arabidopsis Information Resource

Featured Paper

The F-box Protein KIB1 Mediates Brassinosteroid-Induced Inactivation and Degradation of GSK3-like Kinases in Arabidopsis

J. Zhu, Y. Li, D. Cao, H. Yang, E. Oh, Y. Bi, S. Zhu, Z. Wang

About TAIR

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.

The Arabidopsis Biological Resource Center at The Ohio State University collects, reproduces, preserves and distributes seed and DNA resources of Arabidopsis thaliana and related species. Stock information and ordering for the ABRC are fully integrated into TAIR.

TAIR is located at Phoenix Bioinformatics and funded by subscriptions.

Full access to TAIR requires a subscription. ABRC catalog, stock and ordering pages hosted at TAIR will remain freely accessible. Please see our subscription page for further details.

Note: This site has been tested with Internet Explorer 10, Chrome 39, Safari 1.0 and Firefox/Mozilla browsers. Some pages may not work as expected if you are using older browsers. For best results, update your browser and enable Javascript and cookies (see help).

Scheduled Maintenance: This site may be down for maintenance on any Saturday from 8 am to 10 am PDT.
TAIR SeqViewer Whole Genome View

1 hit, click here for a list.

Click on green chromosome to open a closeup view

Closeup View Options:
- Zoom level: [200 Kb]
- Display:
  - Markers
  - Polymorphisms
  - T-DNA/Tn
  - Gene Models
  - Transcripts
  - Annotation Units

Whole Genome View Options:
- Search
  - name gene
  - sequence (15-150 nt)
- Paste in name(s) (up to 250) or sequence(s) (up to 4): AT1G7810
- or upload a file of names or sequences:

Version: TAIR 10.0 genome sequence, released November 2010
TAIR Keyword Search Results

Your query for keywords where contains root development resulted in 11 matches.

Displaying 1 - 11 of 11 records on page 1 of 1 pages.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Keyword Category</th>
<th>Tree View</th>
<th>Associated Data (to this term and to children terms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>root development</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>515 loci, 1788 publications, 677 annotations</td>
</tr>
<tr>
<td>post-embryonic root development</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>135 loci, 648 publications, 152 annotations</td>
</tr>
<tr>
<td>lateral root development</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>118 loci, 625 publications, 135 annotations</td>
</tr>
<tr>
<td>adventitious root development</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>6 loci, 13 publications, 6 annotations</td>
</tr>
<tr>
<td>primary root development</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>30 loci, 27 publications, 30 annotations</td>
</tr>
<tr>
<td>regulation of lateral root development</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>10 loci, 10 publications, 11 annotations</td>
</tr>
<tr>
<td>regulation of post-embryonic root development</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>12 loci, 11 publications, 13 annotations</td>
</tr>
<tr>
<td>regulation of root development</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>71 loci, 54 publications, 86 annotations</td>
</tr>
<tr>
<td>negative regulation of lateral root develop</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>6 loci, 3 publications, 6 annotations</td>
</tr>
<tr>
<td>positive regulation of lateral root develop</td>
<td>GO Biological Process</td>
<td>treview</td>
<td>2 loci, 1 publications, 2 annotations</td>
</tr>
<tr>
<td>cell wall polysaccharide catabolic process involved in lateral root develop</td>
<td>GO Biological Process</td>
<td>treview</td>
<td></td>
</tr>
</tbody>
</table>

Display:
- loci
- publications
- annotations
- microarray experiments

Check the box and click the display button to see numbers of associated data.

Keyword: root development
ID: GO:00488364
Definition: The process whose specific outcome is the progression of the root over time, from its formation to the mature structure. The root is the water- and mineral-absorbing part of a plant which is usually underground, does not bear leaves, tends to grow downwards and is typically derived from the radicle of the embryo.

Keyword Categories - Click on the link to generate a treview for the category.
- GO Cellular Component
- GO Biological Process
- Plant Growth and Developmental Stages
- GO Molecular Function
- Plant Anatomical Entity
- Experimental Method

<table>
<thead>
<tr>
<th>I = 'is a' relationship</th>
<th>O = 'part of' relationship</th>
<th>P = 'positively regulates' relationship</th>
<th>N = 'negatively regulates' relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>root system development (6 loci to term + 515 loci to children)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>root development (163 loci to term + 395 loci to children)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>root morphogenesis (20 loci to term + 251 loci to children)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post-embryonic root development (14 loci to term + 121 loci to children)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adventitious root development (6 loci to term)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>root cap development (15 loci to term)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>root meristem growth (14 loci to term + 27 loci to children)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stele development (1 loci to term)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>primary root development (30 loci to term)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>root radial pattern formation (3 loci to term)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regulation of root development (25 loci to term + 51 loci to children)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Functional Categorization

<table>
<thead>
<tr>
<th>Keyword Category</th>
<th>Functional Category</th>
<th>Annotation Count</th>
<th>Gene Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO Cellular Component</td>
<td>nucleus</td>
<td>72</td>
<td>21</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>cytoplasmic components</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>intracellular</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>extracellular</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>nuclear</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Cellular Component</td>
<td>organelle</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>DNA or RNA binding</td>
<td>90</td>
<td>71</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>transcription factor activity</td>
<td>71</td>
<td>71</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>protein binding</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>nucleic acid binding</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>other binding</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>nucleotide binding</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>kinase activity</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>transferase activity</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>hydrolase activity</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>other enzyme activity</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>transporter activity</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>receptor binding or activity</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>other molecular functions</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>structural molecule activity</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Molecular Function</td>
<td>unknown molecular functions</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>transcription-DNA-dependent</td>
<td>155</td>
<td>21</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>other cellular processes</td>
<td>114</td>
<td>21</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>other metabolic processes</td>
<td>92</td>
<td>21</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>response to stress</td>
<td>71</td>
<td>31</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>response to abiotic or biotic stimulus</td>
<td>53</td>
<td>21</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>other biological processes</td>
<td>39</td>
<td>23</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>developmental processes</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>signal transduction</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>protein metabolism</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>transport</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>DNA or RNA metabolism</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>cell organization and biogenesis</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>unknown biological processes</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>electron transport or energy pathways</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Charts for Functional Categorization**

**GO Cellular Component**
- Nucleus: 60.747% (raw value = 72)
- Other cytoplasmic components: 4.819% (raw value = 4)
- Other intracellular components: 4.507% (raw value = 3)
- Extracellular: 2.411% (raw value = 2)
- Plasma membrane: 0.019% (raw value = 1)
- Other: 1.205% (raw value = 1)

**GO Biological Process**
- Transcription/DNA-dependent: 29.493% (raw value = 15)
- Cellular processes: 25.566% (raw value = 14)
- Other metabolic processes: 10.012% (raw value = 10)
- Response to stress: 8.363% (raw value = 7)
- Signal transduction: 2.206% (raw value = 12)
- Protein metabolism: 0.144% (raw value = 1)
- Transport: 0.004% (raw value = 1)

**GO Molecular Function**
- DNA or RNA binding: 40.309% (raw value = 90)
- Transcription factor activity: 31.020% (raw value = 15)
- Protein binding: 20.920% (raw value = 12)
- Other binding: 2.009% (raw value = 6)
- Nucleic acid binding: 1.997% (raw value = 4)
- Tissue activity: 1.346% (raw value = 3)
- Transferase activity: 0.899% (raw value = 2)
### PANTHER Analysis Summary

**Analysis Type:** PANTHER Overrepresentation Test (release 20170413)

**Annotation Version and Release Date:** GO Ontology database Released 2017-06-29

**Analysed List:** upload_1 (Arabidopsis thaliana)

**Reference List:** Arabidopsis thaliana (all genes in database)

**Annotation Data Set:** GO biological process complete

**Use the Bonferroni correction for multiple testing.**

### Results

<table>
<thead>
<tr>
<th>Mapped IDs:</th>
<th>Reference list</th>
<th>upload_1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27060 out of 27060</td>
<td>2467 out of 2461</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unmapped IDs:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiple mapping information:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

| Bonferroni count: | 2096 |

### Displaying only results with P<0.05: [click here to display all results]

<table>
<thead>
<tr>
<th>GO biological process complete</th>
<th>Arabidopsis thaliana (REF)</th>
<th>upload_1 (H: Hierarchy, NEW?)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td># expected</td>
</tr>
<tr>
<td>response to chitin</td>
<td>104</td>
<td>57</td>
</tr>
<tr>
<td>response to oxygen-containing compound</td>
<td>1975</td>
<td>212</td>
</tr>
<tr>
<td>response to chemical</td>
<td>1974</td>
<td>313</td>
</tr>
<tr>
<td>response to stimulus</td>
<td>4024</td>
<td>611</td>
</tr>
<tr>
<td>response to organonitrogen compound</td>
<td>144</td>
<td>60</td>
</tr>
<tr>
<td>response to nitrogen compound</td>
<td>211</td>
<td>66</td>
</tr>
<tr>
<td>response to organic substance</td>
<td>1134</td>
<td>217</td>
</tr>
<tr>
<td>benzene-containing compound metabolic process</td>
<td>73</td>
<td>11</td>
</tr>
<tr>
<td>organic substance metabolic process</td>
<td>6614</td>
<td>691</td>
</tr>
<tr>
<td>metabolic process</td>
<td>7800</td>
<td>843</td>
</tr>
<tr>
<td>cellular metabolic process</td>
<td>6176</td>
<td>666</td>
</tr>
<tr>
<td>cellular process</td>
<td>8237</td>
<td>915</td>
</tr>
<tr>
<td>response to decreased oxygen levels</td>
<td>57</td>
<td>19</td>
</tr>
<tr>
<td>response to oxygen levels</td>
<td>58</td>
<td>19</td>
</tr>
<tr>
<td>response to abiotic stimulus</td>
<td>1397</td>
<td>211</td>
</tr>
<tr>
<td>toxin metabolic process</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>single-organism cellular process</td>
<td>3369</td>
<td>384</td>
</tr>
<tr>
<td>single-organism process</td>
<td>6538</td>
<td>763</td>
</tr>
<tr>
<td>secondary metabolic process</td>
<td>351</td>
<td>68</td>
</tr>
<tr>
<td>single-organism metabolic process</td>
<td>2870</td>
<td>345</td>
</tr>
<tr>
<td>response to wounding</td>
<td>169</td>
<td>49</td>
</tr>
<tr>
<td>response to stress</td>
<td>2497</td>
<td>373</td>
</tr>
<tr>
<td>response to karrin</td>
<td>96</td>
<td>24</td>
</tr>
<tr>
<td>innate immune response</td>
<td>211</td>
<td>51</td>
</tr>
<tr>
<td>immune response</td>
<td>216</td>
<td>51</td>
</tr>
<tr>
<td>immune system process</td>
<td>245</td>
<td>57</td>
</tr>
<tr>
<td>defense response</td>
<td>1177</td>
<td>192</td>
</tr>
<tr>
<td>regulation of defense response</td>
<td>168</td>
<td>39</td>
</tr>
<tr>
<td>regulation of response to stress</td>
<td>217</td>
<td>45</td>
</tr>
<tr>
<td>regulation of response to stimulus</td>
<td>387</td>
<td>65</td>
</tr>
<tr>
<td>regulation of biological process</td>
<td>3799</td>
<td>423</td>
</tr>
<tr>
<td>biological regulation</td>
<td>4106</td>
<td>471</td>
</tr>
<tr>
<td>response to salicylic acid</td>
<td>144</td>
<td>33</td>
</tr>
<tr>
<td>response to acid chemical</td>
<td>837</td>
<td>153</td>
</tr>
<tr>
<td>response to jasmonic acid</td>
<td>164</td>
<td>37</td>
</tr>
<tr>
<td>response to hormone</td>
<td>1133</td>
<td>161</td>
</tr>
<tr>
<td>response to endogenous stimulus</td>
<td>1140</td>
<td>165</td>
</tr>
<tr>
<td>response to oxidative stress</td>
<td>28</td>
<td>57</td>
</tr>
</tbody>
</table>
## Associated Polymorphisms

<table>
<thead>
<tr>
<th>Name</th>
<th>Locus</th>
<th>Gene Name(s) and/or Equivalents</th>
<th>Allele Mutagen</th>
</tr>
</thead>
<tbody>
<tr>
<td>SALKSEQ_062236.0</td>
<td>AT5G09430</td>
<td>AT5G09430.1</td>
<td>T-DNA insertion</td>
</tr>
<tr>
<td>SALKSEQ_062236.1</td>
<td></td>
<td></td>
<td>T-DNA insertion</td>
</tr>
<tr>
<td>SALKSEQ_062236.2</td>
<td></td>
<td></td>
<td>T-DNA insertion</td>
</tr>
<tr>
<td>SALKSEQ_062236.3</td>
<td></td>
<td></td>
<td>T-DNA insertion</td>
</tr>
<tr>
<td>SALKSEQ_062236.4</td>
<td></td>
<td></td>
<td>T-DNA insertion</td>
</tr>
<tr>
<td>SALKSEQ_062236.5</td>
<td></td>
<td></td>
<td>T-DNA insertion</td>
</tr>
<tr>
<td>SALKSEQ_062236.6</td>
<td></td>
<td></td>
<td>T-DNA insertion</td>
</tr>
<tr>
<td>SALKSEQ_062236.7</td>
<td></td>
<td></td>
<td>T-DNA insertion</td>
</tr>
<tr>
<td>SALK_062236</td>
<td>AT5G09430</td>
<td></td>
<td>T-DNA insertion</td>
</tr>
</tbody>
</table>

**NOTE:** kanamycin resistance gene may be silenced; PCR- or hybridization-based segregation analysis is required to confirm presence of insertion; may be segregating for phenotypes that are not linked to the insertion; may have additional insertions potentially segregating.
**Statistical Motif Analysis in Promoter or Upstream Gene Sequences**

The program compares the frequencies of 6-mer "words" in your query set of sequences (on both strands) with the frequencies of the words in the current genomic sequence set of 33518 sequences, each containing 500 (or 1000) bp upstream of the start codon of each gene. You can type in sets of AGI locus identifiers (e.g. At1g0130) or sets of fasta sequences. Make sure each fasta header is formatted as such, fasta symbol (>), immediately followed by a unique ID, a space, then all other descriptions (e.g. >ABC1D1.1 my gene). Ensure that there are no sequences appearing more than once in your query set.

- At1g69930
- At3g46230
- At5g12020
- At4g10250
- At5g12030
- At1g69920
- At5g52760
- At2g26150
- At1g59860
- At2g28210
- At1g56090
- At3g02840
- At3g54150
- At1g53540
- At5g42380

Upload file: [Choose File]  No file chosen

Dataset:
- 500 bp upstream  
- 1000 bp upstream  
- 3000 bp upstream

Output type:
- HTML
- Text

---

**Motif Analysis in Promoter/Upstream Sequences**

Only oligos occurring in 3 or more of sequences in the query set are reported, and are sorted by p-value. Columns are as follows (left to right):

- oligoMer
- Absolute number of this oligoMer in query set
- Absolute number in genomic set
- Number of sequences in query set containing oligoMer
- Number of sequences (out of 33602 in genomic set) containing oligoMer
- p-value from binomial distribution
- Query sequences containing this oligoMer

<table>
<thead>
<tr>
<th>oligoMer</th>
<th>query count</th>
<th>genomic count</th>
<th>query hits</th>
<th>genomic hits</th>
<th>p-value</th>
<th>query sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTTTTT</td>
<td>19</td>
<td>519880791154815552</td>
<td>8/15</td>
<td>5104188064021423616/33802</td>
<td>-3.40e+216</td>
<td>AT1G69930 AT4G10250 AT5G52760 AT2G26150 AT1G66090 AT3G02840 AT3G54150 AT1G53540</td>
</tr>
<tr>
<td>CTTTTT</td>
<td>10</td>
<td>5140642372042998784</td>
<td>8/15</td>
<td>5101168805088446484/33802</td>
<td>-3.37e+216</td>
<td>AT1G69930 AT3G46230 AT5G12030 AT5G52760 AT1G59860 AT1G66090 AT3G54150 AT5G42380</td>
</tr>
</tbody>
</table>
### Molecular Function Annotations (1)

<table>
<thead>
<tr>
<th>Molecular Function</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein kinase activity</td>
<td>enzyme assays</td>
</tr>
<tr>
<td>phosphoribosyltransferase activity</td>
<td>Enzyme assays</td>
</tr>
</tbody>
</table>

### Biological Process Annotations

<table>
<thead>
<tr>
<th>Biological Process</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>seed development</td>
<td>biochemical analysis</td>
</tr>
</tbody>
</table>

### Subcellular Localization Annotations

<table>
<thead>
<tr>
<th>Subcellular Localization</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>plasma membrane</td>
<td>localization of GFP/YFP fusion protein</td>
</tr>
</tbody>
</table>

### Expression Annotations

<table>
<thead>
<tr>
<th>Expression</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypooxyl</td>
<td>transcript levels (e.g., RT-PCR)</td>
</tr>
</tbody>
</table>

### Interacting Partner Annotations

<table>
<thead>
<tr>
<th>Interacting Partner</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g01040</td>
<td>enzyme assays</td>
</tr>
</tbody>
</table>

### Other Comments

Enter anything you like here.