ARABIDOPSIS: BIOINFORMATIC RESOURCES

Dr. Suvanish Kumar V.S

Dated: 15 January, 2023
Keywords: Arabidopsis, database, genome, transcriptome, proteome, metabolome

INTRODUCTION

Arabidopsis thaliana, also known as Arabidopsis, was the first plant to have its genome fully sequenced. Despite many other plants now having their genomes sequenced, Arabidopsis remains a premier reference for plant biology research in various fields including molecular mechanisms and global ecology. Over 30,000 researchers worldwide use Arabidopsis or data derived from it to inform their research. Data exchange and sharing are crucial for the success of Arabidopsis given its importance in the plant research ecosystem. The worldwide Arabidopsis community had come together to form various organizations to aid future needs of Arabidopsis research. One such example includes Arabidopsis Information Resource (TAIR) which was expanded as International Arabidopsis Informatics Consortium (IAIC) due to lack of funds. The goal of the IAIC is to develop a distributed system of data, tools, and resources through collaborative teams and international funding. The IAIC also developed Araport, a bioinformatics portal to replace and augment TAIR. Araport serves as the underlying infrastructure of Arabidopsis informatics, linking to resources and data sets worldwide and allowing for community-generated modules to be linked in a federated approach. While the federated approach was less successful than intended, Araport and TAIR now co-exist and offer complementary functionalities. The Arabidopsis Information Resource (TAIR), the National Center for Genome Resources (NCGR), and the Bio-Analytic Resource for Plant Biology (BAR) have taken over the operation of IAIC and expanded the functionalities that were previously available on Araport. These tools include BAR thalemine (https://bar.utoronto.ca/thalemine), TAIR Jbrowse (https://www.arabidopsis.org/servlets/jbrowse/arabidopsis?default=true) , NCGR GCV (https://gcv-arabidopsis.ncgr.org/).

Similarly, the Multinational Arabidopsis Steering Committee (MASC) was formed in the 1990s as an ad-hoc committee of scientists from the United States, Europe, Japan, and Australia. The main goal of the committee was, and still is, to promote large-scale studies in Arabidopsis thaliana, with a focus on strengthening international collaboration and coordination. The idea behind this was to reduce redundancy and help guide the community in making progress on projects that can only be successful through combined international efforts. This philosophy of combined and coordinated effort, along with a policy of open data sharing, has proven to be successful and has led to Arabidopsis thaliana being established as a reference plant, with the Arabidopsis community being one of the most active research communities.

Over the past 40 years, Arabidopsis researchers have made significant progress, including publishing the first fully sequenced plant genome, functionally annotating most genes, and developing a plethora of techniques, tools, and resources. In the last 20 years, MASC has released two road maps that have guided Arabidopsis research. The first, "The Multinational Coordinated Arabidopsis thaliana - Genome Research Project" was released in the 1990s, followed by "The Multinational Coordinated Arabidopsis thaliana Functional Genomics Project" in the 2000s. The third road map, "The Multinational Arabidopsis Steering Committee - From bench to bountiful harvests" was released in 2012 (Lavagi et al., 2012, Plant Cell, 24:2240-2247).

The main objective of the Arabidopsis community stated in the third road map is to gain in-depth knowledge of how the genome is translated into a continuum of processes, from the single molecule to cells and tissues, the whole plant, plant populations, and fields of plants. This will allow researchers to build a predictive model of an Arabidopsis plant. This is accompanied by the development of big data management systems. Additionally, Arabidopsis researchers have put increased effort into outreach to other plant communities and into translational approaches to allow effective exchange of information.

Online resources:

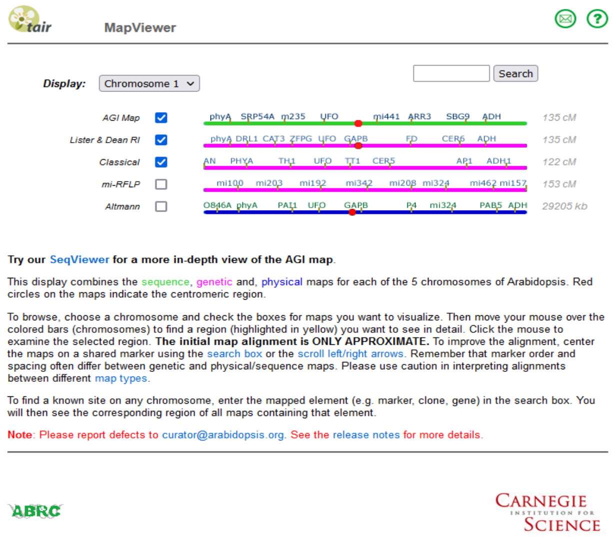
The availability of huge research data on Arabidopsis has led to generation of large database resources with information on genome, proteome and metabolomics. Following are the details on online repositories and resources exclusively for Arabidopsis thaliana:

1. **The Arabidopsis Information Resource (TAIR) https://www.arabidopsis.org/index.jsp**

The Arabidopsis Information Resource (TAIR) is a comprehensive database that serves the needs of the Arabidopsis community and the larger biological research community by providing easy access to a variety of complex data types related to the model plant Arabidopsis thaliana. With the completion and full annotation of the Arabidopsis genome sequence in 2000, the need for a reliable and user-friendly database has become crucial. TAIR's goal is to allow users to efficiently query, browse, visualize, and download information about genes, clones, sequences, markers, mutants, seed stocks, researchers, and research papers. The TAIR team works to maintain data integrity by associating data with researchers, references, and methods and continuously updating and adding new data types as they become available. The transition from the previous Arabidopsis database, AtDB, to TAIR was made with minimal interruption of service, and legacy data from AtDB was also accommodated.

The TAIR website is a go-to resource for those studying the model plant Arabidopsis thaliana. The website offers a variety of tools developed through the project and serves as an easy-to-use interface for accessing the database. The website is divided into six main sections: TAIR DB, Tools, Arabidopsis Information, News, External Links, and an FTP directory.

Visualisation tool: Map Viewer


tair
MapViewer
Display:
Chromosome 1
Search
AGI Map ☑
phyA SRP54A m235 UFO mi441 ARR3 SBG9 ADH
135 cM
Lister & Dean RI ☑
phyA DRL1 CAT3 ZFPG UFO GAPB FD CER6 ADH
135 cM
Classical ☑
AN PHYA TH1 UFO TT1 CER5 AP1 ADH1
122 cM
mi-RFLP □
mi100 mi203 mi192 mi342 mi208 mi324 mi462 mi157
153 cM
Altmann □
O846A phyA PAI1 UFO GAPB P4 mi324 PAB5 ADH
29205 kb
ABRC
CARNEGIE INSTITUTION FOR SCIENCE

Try our SeqViewer for a more in-depth view of the AGI map.

This display combines the sequence, genetic and, physical maps for each of the 5 chromosomes of Arabidopsis. Red circles on the maps indicate the centromeric region.

To browse, choose a chromosome and check the boxes for maps you want to visualize. Then move your mouse over the colored bars (chromosomes) to find a region (highlighted in yellow) you want to see in detail. Click the mouse to examine the selected region. **The initial map alignment is ONLY APPROXIMATE.** To improve the alignment, center the maps on a shared marker using the search box or the scroll left/right arrows. Remember that marker order and spacing often differ between genetic and physical/sequence maps. Please use caution in interpreting alignments between different map types.

To find a known site on any chromosome, enter the mapped element (e.g. marker, clone, gene) in the search box. You will then see the corresponding region of all maps containing that element.

Note: Please report defects to curator@arabidopsis.org. See the release notes for more details.

TAIR's MapViewer

Map viewer is a tool for viewing genetic, physical, and sequence maps for each Arabidopsis chromosome. It allows users to search, browse, align, zoom, scroll, and print maps and mapped objects in TAIR's database. The control panel at the top allows for simultaneous scrolling, zooming, and searching of all open maps, while individual controls on the left provide the same functions for a single map. The chromosome bar for each map shows the current location on the chromosome and allows for easy access to other regions. Each entity on the maps is hyperlinked to an output page from the database, displaying all information about the entity, including associations to other data types, attribution, history, and comments. An extensive help page (http://arabidopsis.org/mapViewer/help/tairmapa.htm) is available for interpretation of the data and navigation of the tool.

TAIR's SeqViewer

SeqViewer is a tool provided by TAIR that allows users to view and analyze the Arabidopsis genome and its associated annotations. It offers a range of features including the ability to view gene annotations, BAC clones, transcripts, polymorphisms, T-DNA/Transposon insertions, and markers. Users can search using names or short nucleotide sequences and can visualize the location of their search results on the genome at various levels of zoom, from 50 megabases to 10 kilobases. Additionally, users can click on displayed objects to access detailed information about them in the TAIR database.

TAIR's GO Term Enrichment for Plants

The Gene Ontology (GO) Term Enrichment tool allows you to identify over or under-represented GO terms in a set of genes, such as those from co-expression or RNAseq data. The tool sends the data to the PANTHER Classification System, which contains current GO annotation data for Arabidopsis and other plant species. The advanced setting allows you to adjust parameters and access other tools provided by PANTHER for analyzing sets of genes.

TAIR's JBrowse

JBrowse is a GMOD (Generic Model Organism Database) project, funded by the National Human Genome Research Institute (NHGRI). JBrowse is a web-based genome browser that provides a user-friendly interface for visualizing and exploring the Arabidopsis genome. It is built using JavaScript and HTML5, which allows for a fast and responsive user experience, even when working with large data sets.

JBrowse allows users to search or browse a map of the Arabidopsis genome, providing detailed information on various features of the genome such as genes, cDNAs, ESTs, insertion mutants, SNPs, markers, and BACs. The genome map can be navigated by scrolling and zooming, and users can easily jump to specific regions of interest by searching for a gene, marker or other feature of the genome.

The JBrowse interface also allows users to customize the tracks they view on the genome map, by selecting which features they want to display and adjusting the track height and colour. Additionally, users can upload their own annotation tracks to view alongside the standard tracks.

TAIR's SyntenyViewer

SyntenyViewer is a tool designed to allow users to view and analyze precomputed syntenic regions between A.thaliana (Arabidopsis thaliana) and related genomes. Synteny refers to the presence of similar genetic material, such as genes or chromosomal regions, in different species that have been conserved through evolutionary time.

The syntenic regions that are displayed in SyntenyViewer have been precomputed using the SynMap tool, which is available at genomevolution.org. This tool uses a computational approach to identify and compare syntenic regions across different genomes. It compares the genomic sequences of related species and identifies regions that have high levels of similarity, indicating that they have been conserved through evolution.

Once the syntenic regions have been identified, they are then visualized using the GeVo tool, which is also provided by genomevolution.org. This tool allows users to view the syntenic regions in a graphical format, making it easy to understand the similarities and differences between the genomes. The GeVo tool provides various features such as zooming, panning, and navigating through the genome.

TAIR's Aracyc Pathway

AraCyc Pathways is a web-based tool designed to help researchers understand and explore all the known biochemical pathways of this model plant species. This tool is based on the AraCyc database, which is a comprehensive resource with information on the enzymatic reactions, metabolic pathways, and genes of A. thaliana. This database is manually curated by experts in the field and is updated regularly to ensure accuracy.

The AraCyc Pathways tool allows users to search for specific pathways or enzymes, view the metabolic pathways in a graphical format, and explore the relationships between different pathways. Users can also view the genes and enzymes associated with each pathway, as well as the metabolic intermediates and products. One of the key features of the tool is its ability to display the pathways in a hierarchical format, which makes it easy for users to understand the relationships between different pathways. This tools also provides access to external databases such as KEGG, PubChem, and UniProt.

TAIR’s BLAST

BLAST allows users to search for their nucleotide or peptide sequences against all public Arabidopsis sequences, subsets of them, or all higher plant sequences from GenBank.

TAIR’s Pattern Matching

TAIR's Pattern Matching tool is a web-based tool based on regular expressions that allows users to search for short (<20 residues) nucleotide or peptide sequences, or ambiguous/degenerate patterns in TAIR's Arabidopsis dataset. Users can specify complex search criteria and find matches even in cases of ambiguous or degenerate sequences. This tool is particularly useful for researchers who are looking for specific sequences or patterns in the Arabidopsis genome or associated annotations. The tool returns a list of matches, which can be viewed in a tabular format and sorted by various criteria such as the chromosome location or gene name and allows to download and upload data in .fasta format.

TAIR’s Motif Analysis tool

Motif Analysis tool allows finding overrepresented 6-mer oligos in upstream regions of genes. It allows users to search for and analyse cis-regulatory motifs (sequences of nucleotides that control the expression of genes) in the Arabidopsis genome. Users can search for known motifs or upload their own sequences to search for potential matches. The tool also provides information about the location and frequency of the motifs in the genome, as well as information about the genes that are regulated by the motifs.

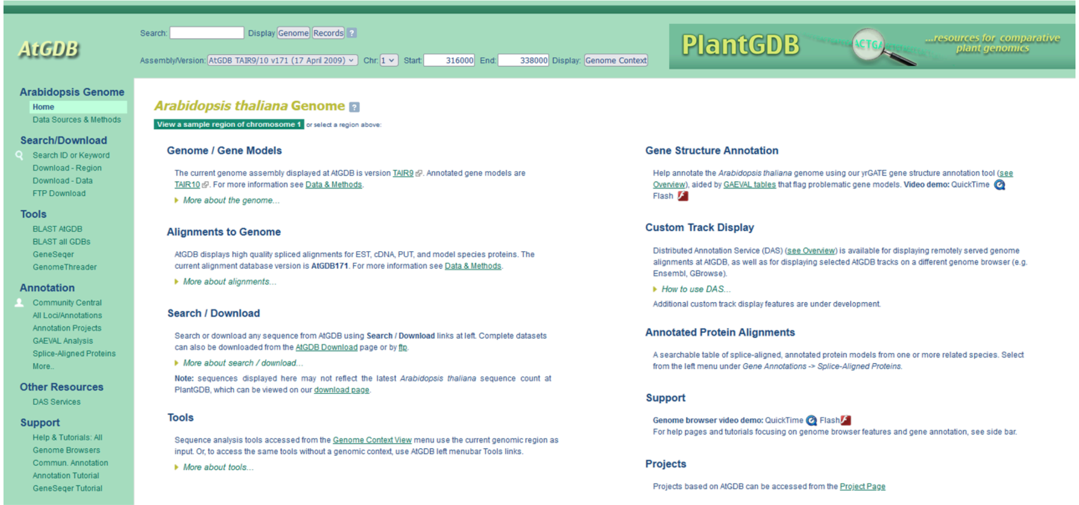
VxInsight®

VxInsight® is a data-mining software developed by Sandia National Laboratories. It is designed to map gene expression data in an intuitive 3D format, represented as a mountain terrain map. This allows for easy exploration and understanding of the data. The program can be used for a wide range of gene expression datasets, but it is particularly useful for analyzing data from The Arabidopsis Information Resource (TAIR). One of the key features of VxInsight® is its ability to identify patterns and relationships within large sets of gene expression data. It can identify clusters of genes that have similar patterns of expression, as well as genes that are differentially expressed between different conditions or treatments. This allows researchers to quickly identify genes that may be involved in specific biological processes or pathways.

TAIR's Chromosome map tool

Chromosome map tool allows users to view the physical and genetic map of the Arabidopsis genome, including information on genes, markers, and other features of the chromosomes. The tool allows users to navigate the genome using a variety of views including the ideograms, cytological, and genetic maps.

2. The Arabidopsis Genome Database (AtGDB) (https://www.plantgdb.org/AtGDB/)



Arabidopsis thaliana **Genome**

Genome / Gene Models

The current genome assembly displayed at AtGDB is version TAIR9. Annotated gene models are TAIR10. For more information see Data & Methods.

▸ *More about the genome...*

Alignments to Genome

AtGDB displays high quality spliced alignments for EST, cDNA, PUT, and model species proteins. The current alignment database version is **AtGDB171**. For more information see Data & Methods.

▸ *More about alignments...*

Search / Download

Search or download any sequence from AtGDB using **Search / Download** links at left. Complete datasets can also be downloaded from the AtGDB Download page or by ftp.

▸ *More about search / download...*

Note: sequences displayed here may not reflect the latest *Arabidopsis thaliana* sequence count at PlantGDB, which can be viewed on our download page.

Tools

Sequence analysis tools accessed from the Genome Context View menu use the current genomic region as input. Or, to access the same tools without a genomic context, use AtGDB left menubar Tools links.

▸ *More about tools...*

Gene Structure Annotation

Help annotate the *Arabidopsis thaliana* genome using our yrGATE gene structure annotation tool (see Overview), aided by GAEVAL tables that flag problematic gene models. **Video demo:** QuickTime Flash

Custom Track Display

Distributed Annotation Service (DAS) (see Overview) is available for displaying remotely served genome alignments at AtGDB, as well as for displaying selected AtGDB tracks on a different genome browser (e.g. Ensembl, GBrowse).

▸ *How to use DAS...*

Additional custom track display features are under development.

Annotated Protein Alignments

A searchable table of splice-aligned, annotated protein models from one or more related species. Select from the left menu under *Gene Annotations -> Splice-Aligned Proteins.*

Support

Genome browser video demo: QuickTime Flash
For help pages and tutorials focusing on genome browser features and gene annotation, see side bar.

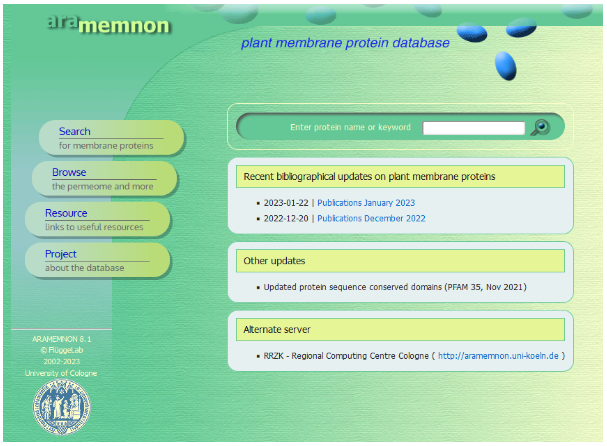
Projects

Projects based on AtGDB can be accessed from the Project Page

The Arabidopsis Genome Database (AtGDB) is a bioinformatic resource that provides a wide range of information and tools for studying the genetics, genomics, and biology of the model plant Arabidopsis thaliana. It includes data for genomic sequences, gene expression, genetic and physical maps, and a wide range of other related information.

AtGDB provides a user-friendly web interface that allows users to browse, search and download data, and several tools and resources for data analysis, such as gene annotation, functional analysis, and gene expression visualization. It allows downloading of specific sequence or a complete dataset using ID or keyword from any genomic region in FASTA format. Downloading of aligned/computed transcripts or proteins or all genome annotations in genbank, GFF3 or EMBL format is made available.4

3. Plant Membrane Protein Database (PMPD) http://aramemnon.botanik.uni-koeln.de/



Recent bibliographical updates on plant membrane proteins

- 2023-01-22 | Publications January 2023
- 2022-12-20 | Publications December 2022

Other updates

- Updated protein sequence conserved domains (PFAM 35, Nov 2021)

Alternate server

- RRZK - Regional Computing Centre Cologne (http://aramemnon.uni-koeln.de)

The Plant Membrane Protein Database (PMPD) is a database that contains information on various types of membrane proteins from different plant species including Arabidopsis thaliana. The database provides detailed information on the structure, function, and localization of these proteins, as well as their evolutionary relationships which can be useful in understanding the origin and diversification of these proteins. The PMPD also provides a platform for the visualization and analysis of the data.

The data in PMPD is curated from a variety of sources, such as literature, experimental data, and computational predictions. The database is regularly updated with new data and is freely accessible to the scientific community.

4. *Arabidopsis* Small RNA Project (ASRP)

https://asrp.danforthcenter.org/

The ASRP is a research project and website that provides access to data and resources from the Carrington laboratory. This website is aimed at understanding the role of small RNAs in the regulation of gene expression in the model plant Arabidopsis thaliana. Small RNAs are a class of non-coding RNAs that are typically 20-24 nucleotides in length and play important roles in various cellular processes, including gene silencing, chromatin remodelling, and RNA decay.

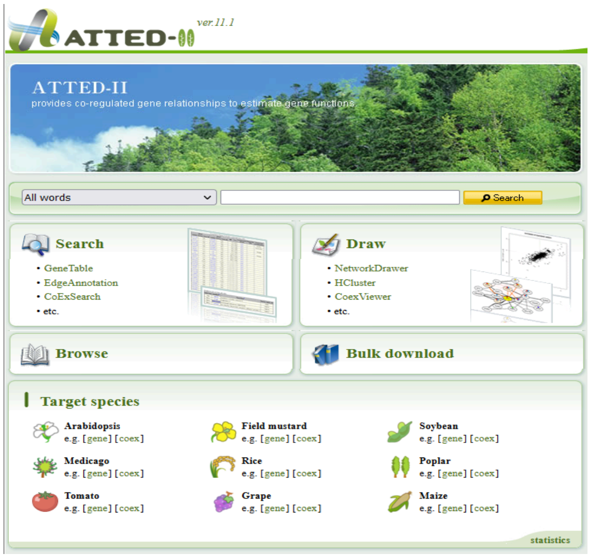
The project involves the use of high-throughput sequencing techniques to map and analyze the small RNA transcriptome of Arabidopsis. This includes the identification and characterization of small RNA species, their genomic locations, and the genes and pathways that they regulate.

The project has generated a large amount of data on small RNAs in Arabidopsis, including the identification of many novel small RNA species and the characterization of their biogenesis and function. The data generated by this project has provided insights into the mechanisms of small RNA-mediated gene regulation in plants and has revealed a complex network of small RNAs that play important roles in various physiological processes.

5. Arabidopsis thaliana Transcriptional Expression Database ATTED-II

https://atted.jp/

The Arabidopsis thaliana Transcriptional Expression Database (ATTED) is a database that contains information on the transcriptional expression patterns of genes including gene expression levels, tissue-specific expression patterns, and the effects of environmental and genetic perturbations on gene expression in the model plant Arabidopsis thaliana. The database is based on the results of large-scale microarray and RNA sequencing experiments and provides

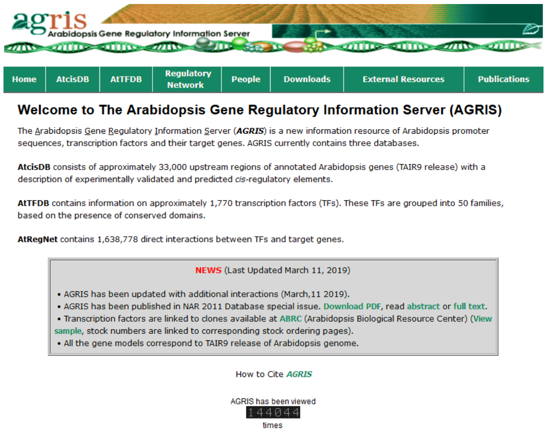


ATTED-II provides co-regulated gene relationships to estimate gene functions.

a comprehensive view of the expression of Arabidopsis genes under various conditions and in different tissues and developmental stages. ATTED database integrates data from multiple experiments, which allows researchers to obtain a more complete and accurate view of gene expression patterns. The database also includes data from different platforms, such as Affymetrix and Illumina, which allows for a more accurate and comprehensive view of the transcriptome. The database is continuously updated with new data and is freely accessible to the scientific community.

6. The *Arabidopsis* Gene Regulatory Information Server (AGRIS) https://agris-knowledgebase.org/

AGRIS (International System for Agricultural Science and Technology) is a global database of agricultural and related information. It is maintained by the Food and Agriculture Organization of the United Nations (FAO) and is a bibliographic information system that provides access to research and technical literature in the field of agriculture, forestry, and fisheries. The database contains over 3 million bibliographic records and covers publications from over 150 countries. It is available online and can be searched by keyword, author, title, and other criteria. AGRIS aims to improve access to information for researchers, policymakers, and other stakeholders in the agricultural sector.



Welcome to The Arabidopsis Gene Regulatory Information Server (AGRIS)

The Arabidopsis Gene Regulatory Information Server (***AGRIS***) is a new information resource of Arabidopsis promoter sequences, transcription factors and their target genes. AGRIS currently contains three databases.

AtcisDB consists of approximately 33,000 upstream regions of annotated Arabidopsis genes (TAIR9 release) with a description of experimentally validated and predicted *cis*-regulatory elements.

AtTFDB contains information on approximately 1,770 transcription factors (TFs). These TFs are grouped into 50 families, based on the presence of conserved domains.

AtRegNet contains 1,638,778 direct interactions between TFs and target genes.

NEWS (Last Updated March 11, 2019)

- AGRIS has been updated with additional interactions (March,11 2019).
- AGRIS has been published in NAR 2011 Database special issue. Download PDF, read abstract or full text.
- Transcription factors are linked to clones available at ABRC (Arabidopsis Biological Resource Center) (View sample, stock numbers are linked to corresponding stock ordering pages).
- All the gene models correspond to TAIR9 release of Arabidopsis genome.

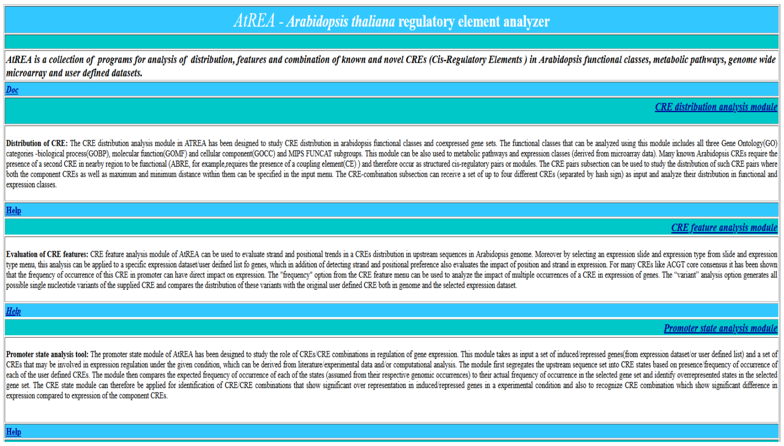
How to Cite ***AGRIS***

AGRIS has been viewed
144044
times

7. *Arabidopsis thaliana* regulatory element analyzer (AtREA) http://www.bioinformatics.org/grn/atrea.html

Arabidopsis thaliana Regulatory Element Analyzer (ATREA) is a bioinformatics tool that is used to predict cis-regulatory elements in within the Arabidopsis thaliana genome. These elements play a significant role in regulating the gene expression within the genome. They are located near the genes they control, in the promoter regions, enhancer regions, or other regions of the genome.

ATREA utilizes a combination of machine learning techniques, such as the artificial neural networks and support vector machines, to analyse the DNA sequences and identify potential cis-regulatory elements. The tool is trained on a large dataset of known cis-regulatory elements and uses this information to predict new elements in the genome.



AtREA - Arabidopsis thaliana **regulatory element analyzer**

AtREA is a collection of programs for analysis of distribution, features and combination of known and novel CREs (Cis-Regulatory Elements) in Arabidopsis functional classes, metabolic pathways, genome wide microarray and user defined datasets.

Doc

CRE distribution analysis module

Distribution of CRE: The CRE distribution analysis module in ATREA has been designed to study CRE distribution in arabidopsis functional classes and coexpressed gene sets. The functional classes that can be analyzed using this module includes all three Gene Ontology(GO) categories -biological process(GOBP), molecular function(GOMF) and cellular component(GOCC) and MIPS FUNCAT subgroups. This module can be also used to metabolic pathways and expression classes (derived from microarray data). Many known Arabidopsis CREs require the presence of a second CRE in nearby region to be functional (ABRE, for example,requires the presence of a coupling element(CE)) and therefore occur as structured cis-regulatory pairs or modules. The CRE pairs subsection can be used to study the distribution of such CRE pairs where both the component CREs as well as maximum and minimum distance within them can be specified in the input menu. The CRE-combination subsection can receive a set of up to four different CREs (separated by hash sign) as input and analyze their distribution in functional and expression classes.

Help

CRE feature analysis module

Evaluation of CRE features: CRE feature analysis module of AtREA can be used to evaluate strand and positional trends in a CREs distribution in upstream sequences in Arabidopsis genome. Moreover by selecting an expression slide and expression type from slide and expression type menu, this analysis can be applied to a specific expression dataset/user deifned list fo genes, which in addition of detecting strand and positional preference also evaluates the impact of position and strand in expression. For many CREs like ACGT core consensus it has been shown that the frequency of occurrence of this CRE in promoter can have direct impact on expression. The "frequency" option from the CRE feature menu can be used to analyze the impact of multiple occurrences of a CRE in expression of genes. The “variant” analysis option generates all possible single nucleotide variants of the supplied CRE and compares the distribution of these variants with the original user defined CRE both in genome and the selected expression dataset.

Help

Promoter state analysis module

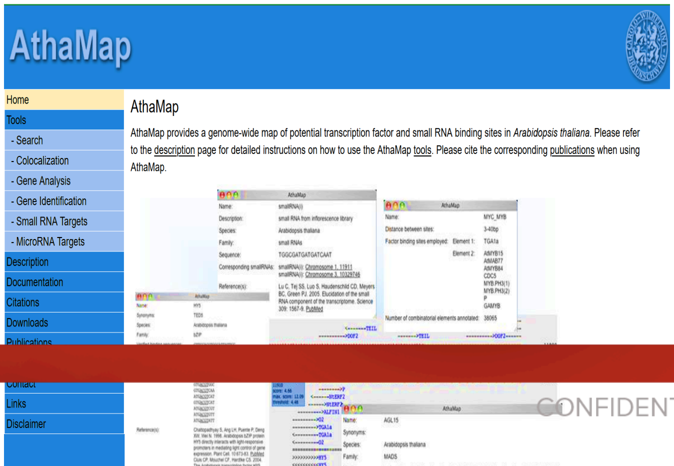
Promoter state analysis tool: The promoter state module of AtREA has been designed to study the role of CREs/CRE combinations in regulation of gene expression. This module takes as input a set of induced/repressed genes(from expression dataset/or user defined list) and a set of CREs that may be involved in expression regulation under the given condition, which can be derived from literature/experimental data and/or computational analysis. The module first segregates the upstream sequence set into CRE states based on presence/frequency of occurrence of each of the user defined CREs. The module then compares the expected frequency of occurrence of each of the states (assumed from their respective genomic occurrences) to their actual frequency of occurrence in the selected gene set and identify overrepresented states in the selected gene set. The CRE state module can therefore be applied for identification of CRE/CRE combinations that show significant over representation in induced/repressed genes in a experimental condition and also to recognize CRE combination which show significant difference in expression compared to expression of the component CREs.

Help

ATREA offers an easy-to-use interface for users to submit their DNA sequences for analysis and view the results in different formats, such as graphical displays of the predicted cis-regulatory elements and their positions in the genome. Overall, ATREA is a valuable tool for researchers studying gene regulation in Arabidopsis and other plants, as it can help identify new cis-regulatory elements and understand the mechanisms that control gene expression in these organisms.

8. AthaMap (http://www.athamap.de/)

AthaMap is a comprehensive genetic map of the model plant Arabidopsis thaliana created based on data from a variety of mapping techniques, including restriction fragment length polymorphism (RFLP), amplified fragment length



AthaMap
Home
Tools
- Search
- Colocalization
- Gene Analysis
- Gene Identification
- Small RNA Targets
- MicroRNA Targets
Description
Documentation
Citations
Downloads
Publications
Contact
Links
Disclaimer
AthaMap
Name: smallRNA(i)
Description: small RNA from inflorescence library
Species: Arabidopsis thaliana
Family: small RNAs
Sequence: TGGCGATGATGATCAAT
Corresponding smallRNAs: smallRNA(i): Chromosome 1, 11911 smallRNA(i): Chromosome 3, 10329746
Reference(s): Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ. 2005. Elucidation of the small RNA component of the transcriptome. Science 309: 1567-9. PubMed
AthaMap
Name: MYC_MYB
Distance between sites: 3-40bp
Factor binding sites employed: Element 1: TGA1a
Element 2: AtMYB15 AtMAB77 AtMYB84 CDC5 MYB.PH3(1) MYB.PH3(2) P GAMYB
Number of combinatorial elements annotated: 38065
AthaMap
Name: HY5
Synonyms: TED5
Species: Arabidopsis thaliana
Family: bZIP
AthaMap
Name: AGL15
Synonyms:
Species: Arabidopsis thaliana
Family: MADS

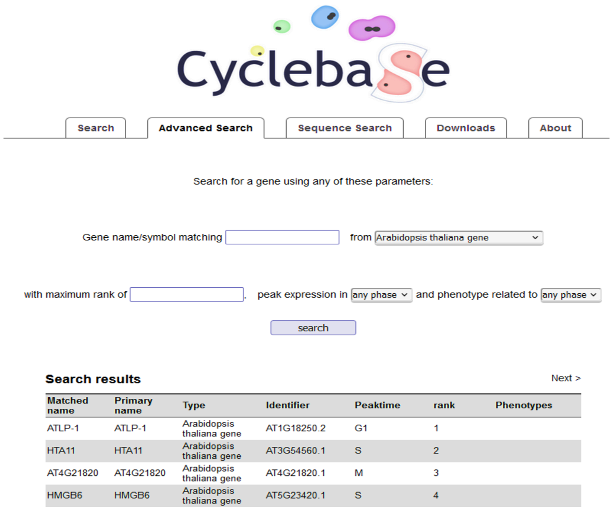
AthaMap

AthaMap provides a genome-wide map of potential transcription factor and small RNA binding sites in *Arabidopsis thaliana*. Please refer to the description page for detailed instructions on how to use the AthaMap tools. Please cite the corresponding publications when using AthaMap.

polymorphism (AFLP), and simple sequence repeat (SSR) markers. The map consists of over 2,500 markers that span the entire Arabidopsis genome, with an average marker density of one marker per 5 kb. AthaMap was used to identify the locations of genes associated with various traits, such as flowering time, seed development, and disease resistance, and helped to identify the locations of transposable elements, which make up a large portion of the Arabidopsis genome. The AthaMap provides researchers an important tool in the field of plant genetics and genomics and has been an important tool for the study of plant development and evolution.

9. Cyclebase (https://cyclebase.org/Advanced%20Search?type=3702)

Cyclebase is a database for plant circadian clock research with focus on the dynamic regulation of clock-controlled genes. It provides a comprehensive collection of time-series data on gene expression, protein abundance, and protein-protein interactions, as well as information on the regulation of these processes by various signalling pathways. Not only that, it includes information on clock-controlled genes, their promoters, and their cis- and trans-acting regulatory elements.This allows researchers to gain a detailed understanding of how the clock works at the molecular level, and how it is regulated by environmental cues such as light and temperature.



Search for a gene using any of these parameters:

Gene name/symbol matching □ from Arabidopsis thaliana gene

with maximum rank of □, peak expression in any phase and phenotype related to any phase

search

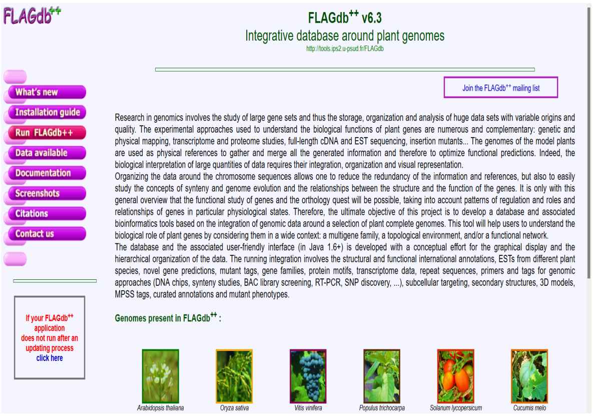
Search results Next >

Matched name	Primary name	Type	Identifier	Peaktime	rank	Phenotypes
ATLP-1	ATLP-1	Arabidopsis thaliana gene	AT1G18250.2	G1	1	
HTA11	HTA11	Arabidopsis thaliana gene	AT3G54560.1	S	2	
AT4G21820	AT4G21820	Arabidopsis thaliana gene	AT4G21820.1	M	3	
HMGB6	HMGB6	Arabidopsis thaliana gene	AT5G23420.1	S	4	

In addition to its data resources, Cyclebase also provides a variety of tools for data analysis and visualization which allows users to search and browse the database, perform gene set enrichment analysis, network analysis, and other types of data mining.

10.FLAGdb++ v6.3 (http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtml)

FLAGdb++ v6.3 is a database for the gene ontology, gene interactions, and protein-protein interactions. It also provides information on gene expression patterns under different environmental conditions, including different developmental stages with respect to treatments with different hormones or abiotic factors. It provides a comprehensive collection of information on the function, structure, and regulation of Arabidopsis genes, as well as tools for gene set enrichment analysis, network analysis, and other types of data mining along a user-friendly interface for data visualization.



FLAGdb^{++} v6.3

Integrative database around plant genomes

http://tools.ips2.u-psud.fr/FLAGdb

Research in genomics involves the study of large gene sets and thus the storage, organization and analysis of huge data sets with variable origins and quality. The experimental approaches used to understand the biological functions of plant genes are numerous and complementary: genetic and physical mapping, transcriptome and proteome studies, full-length cDNA and EST sequencing, insertion mutants... The genomes of the model plants are used as physical references to gather and merge all the generated information and therefore to optimize functional predictions. Indeed, the biological interpretation of large quantities of data requires their integration, organization and visual representation.

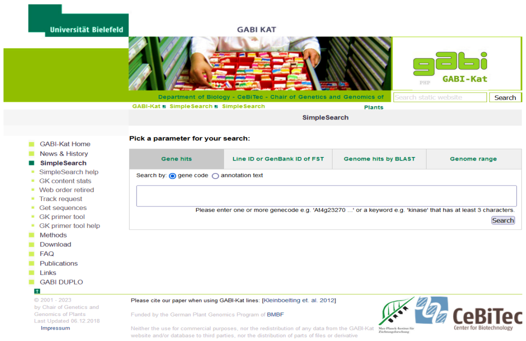
Organizing the data around the chromosome sequences allows one to reduce the redundancy of the information and references, but also to easily study the concepts of synteny and genome evolution and the relationships between the structure and the function of the genes. It is only with this general overview that the functional study of genes and the orthology quest will be possible, taking into account patterns of regulation and roles and relationships of genes in particular physiological states. Therefore, the ultimate objective of this project is to develop a database and associated bioinformatics tools based on the integration of genomic data around a selection of plant complete genomes. This tool will help users to understand the biological role of plant genes by considering them in a wide context: a multigene family, a topological environment, and/or a functional network.

The database and the associated user-friendly interface (in Java 1.6+) is developed with a conceptual effort for the graphical display and the hierarchical organization of the data. The running integration involves the structural and functional international annotations, ESTs from different plant species, novel gene predictions, mutant tags, gene families, protein motifs, transcriptome data, repeat sequences, primers and tags for genomic approaches (DNA chips, synteny studies, BAC library screening, RT-PCR, SNP discovery, ...), subcellular targeting, secondary structures, 3D models, MPSS tags, curated annotations and mutant phenotypes.

Genomes present in FLAGdb^{++} :

11. GABI-Kat (https://www.gabi-kat.de/)

GABI-kat is a functional genomics project that aims to identify and study the functions of all protein kinases in the plant Arabidopsis thaliana. It employs a combination of genetic, biochemical, and bioinformatic approaches to systematically identify and characterize all kinases and their substrates, as well as their roles in various physiological and developmental processes. For researchers it is expected to provide valuable insights into the



SimpleSearch

Pick a parameter for your search:

Search by: ● gene code ○ annotation text

Please enter one or more genecode e.g. 'At4g23270 ...' or a keyword e.g. 'kinase' that has at least 3 characters.

© 2001 - 2023 by Chair of Genetics and Genomics of Plants Last Updated 06.12.2018

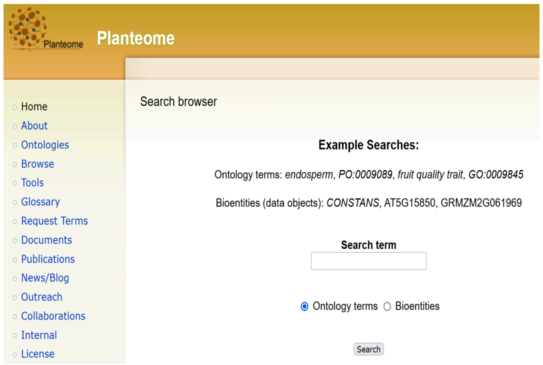
Please cite our paper when using GABI-Kat lines: [Kleinboelting et. al. 2012]

Funded by the German Plant Genomics Program of BMBF

Neither the use for commercial purposes, nor the redistribution of any data from the GABI-Kat website and/or database to third parties, nor the distribution of parts of files or derivative molecular mechanisms underlying plant growth and adaptation to changing environments.

12.Planteome (https://planteome.org/)

The Plant Ontology (PO) is a database of plant structure and growth stages, developed through collaborative effort between several institutions and is funded by the National Science Foundation. They also provide following information:



Search browser

Example Searches:

Ontology terms: *endosperm*, *PO:0009089*, *fruit quality trait*, *GO:0009845*

Bioentities (data objects): *CONSTANS*, AT5G15850, GRMZM2G061969

Search term

● Ontology terms ○ Bioentities

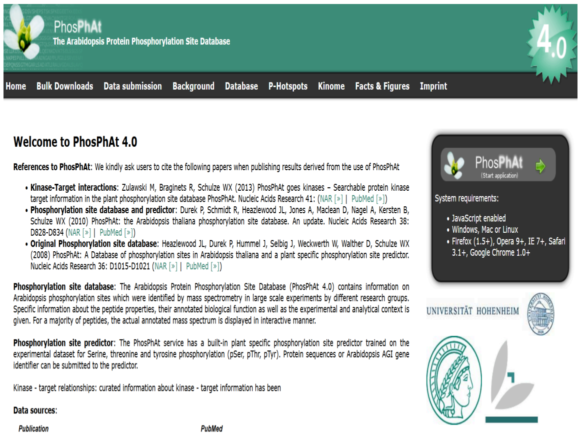
The Plant Phenotype Ontology (PPO) which describes the observable characteristics of plants.

The Plant Trait Ontology (TO) which describes the measurable characteristics of plants.

The Plant Environment Ontology (ENVO) which describes the environmental factors that affect plant growth and development.

13.PhosPhat3.0 (https://phosphat.uni-hohenheim.de/)

PhosPhAt 3.0 is a comprehensive database that contains information on phosphorylation sites identified in the plant Arabidopsis thaliana using mass spectrometry. The database includes detailed information on the properties of the peptides, their annotated biological function, and the experimental and analytical context in which they were



Welcome to PhosPhAt 4.0

References to PhosPhAt: We kindly ask users to cite the following papers when publishing results derived from the use of PhosPhAt

- **Kinase-Target interactions**: Zulawski M, Braginets R, Schulze WX (2013) PhosPhAt goes kinases – Searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. Nucleic Acids Research 41: (NAR [»] | PubMed [»])
- **Phosphorylation site database and predictor**: Durek P, Schmidt R, Heazlewood JL, Jones A, Maclean D, Nagel A, Kersten B, Schulze WX (2010) PhosPhAt: the Arabidopsis thaliana phosphorylation site database. An update. Nucleic Acids Research 38: D828-D834 (NAR [»] | PubMed [»])
- **Original Phosphorylation site database**: Heazlewood JL, Durek P, Hummel J, Selbig J, Weckwerth W, Walther D, Schulze WX (2008) PhosPhAt: A Database of phosphorylation sites in Arabidopsis thaliana and a plant specific phosphorylation site predictor. Nucleic Acids Research 36: D1015-D1021 (NAR [»] | PubMed [»])

Phosphorylation site database: The Arabidopsis Protein Phosphorylation Site Database (PhosPhAt 4.0) contains information on Arabidopsis phosphorylation sites which were identified by mass spectrometry in large scale experiments by different research groups. Specific information about the peptide properties, their annotated biological function as well as the experimental and analytical context is given. For a majority of peptides, the actual annotated mass spectrum is displayed in interactive manner.

Phosphorylation site predictor: The PhosPhAt service has a built-in plant specific phosphorylation site predictor trained on the experimental dataset for Serine, threonine and tyrosine phosphorylation (pSer, pThr, pTyr). Protein sequences or Arabidopsis AGI gene identifier can be submitted to the predictor.

Kinase - target relationships: curated information about kinase - target information has been

Data sources:

Publication *PubMed*

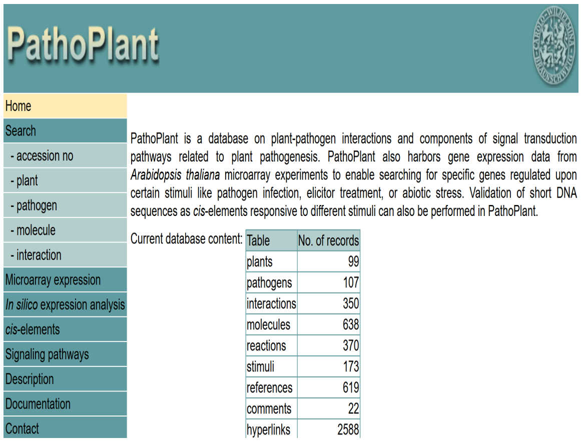
System requirements:

- JavaScript enabled
- Windows, Mac or Linux
- Firefox (1.5+), Opera 9+, IE 7+, Safari 3.1+, Google Chrome 1.0+

identified. A majority of the peptides also have interactive mass spectra available for viewing. The database also provides interactive visualization of annotated fragmentation spectra and the ability to export spectra and peptide sequences for use in other applications. Additionally, the database has dynamic links to other web resources, providing access to external protein-related data. For phosphorylation sites with information about dynamic behavior in response to external stimuli, the database displays simple time-resolved diagrams. The database also includes predictions for pT and pY sites and updated pS predictions. Users can also access the prediction algorithm to predict phosphorylation on any user-uploaded protein sequence. The database also maps Pfam domain structures onto the protein sequence display next to experimental and predicted phosphorylation sites and functional annotation of proteins using MAPMAN ontology.

14.PathoPlant (http://www.pathoplant.de/)

PathoPlant is a database provides a comprehensive overview of the molecular and genetic interactions between plants and pathogens, including bacteria, fungi, viruses, and nematodes. The database includes information on different types of plant-pathogen interactions, such as host-pathogen recognition, signalling, and defence response. It also includes information on pathogen virulence factors, plant resistance genes, and the molecular mechanisms underlying these interactions.



PathoPlant is a database on plant-pathogen interactions and components of signal transduction pathways related to plant pathogenesis. PathoPlant also harbors gene expression data from *Arabidopsis thaliana* microarray experiments to enable searching for specific genes regulated upon certain stimuli like pathogen infection, elicitor treatment, or abiotic stress. Validation of short DNA sequences as *cis*-elements responsive to different stimuli can also be performed in PathoPlant.

Current database content:

Table	No. of records
plants	99
pathogens	107
interactions	350
molecules	638
reactions	370
stimuli	173
references	619
comments	22
hyperlinks	2588

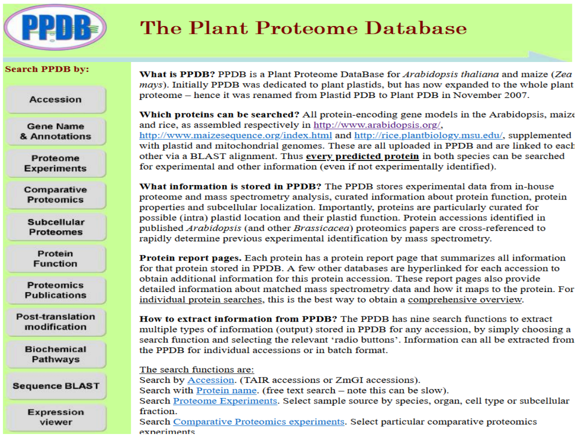
The PathoPlant database is curated by a team of experts and is updated regularly to ensure the accuracy and completeness of the information provided. The database is accessible through a user-friendly interface that allows users to search and retrieve data based on various criteria, such as pathogen or plant species, gene or protein names, or functional categories.

In addition to the database, the PathoPlant also provides several other resources such as:

- A collection of tools for data analysis and visualization, such as sequence alignment, phylogenetic analysis, and gene expression profiling.
- A set of bioinformatics pipelines for the identification and annotation of pathogen virulence factors and plant resistance genes.
- A forum for discussion and collaboration among researchers in the field of plant-microbe interactions.

15.Plant proteome Database(PPDB) (http://ppdb.tc.cornell.edu/)

The Plant Proteome Database (PPDB) is a publicly available resource that provides comprehensive information on plant proteins and their functional annotation. It is funded and run between Klaas J. van Wijk Lab of Cornell University and the Computational Biology Service Unit of Cornell Life Sciences Core Laboratories Center. The database contains information on protein sequences, structures, functional domains, post-translational modifications, interactions, and expression patterns for a variety of plant species.



What is PPDB? PPDB is a Plant Proteome DataBase for *Arabidopsis thaliana* and maize (*Zea mays*). Initially PPDB was dedicated to plant plastids, but has now expanded to the whole plant proteome – hence it was renamed from Plastid PDB to Plant PDB in November 2007.

Which proteins can be searched? All protein-encoding gene models in the Arabidopsis, maize and rice, as assembled respectively in http://www.arabidopsis.org/, http://www.maizesequence.org/index.html and http://rice.plantbiology.msu.edu/, supplemented with plastid and mitochondrial genomes. These are all uploaded in PPDB and are linked to each other via a BLAST alignment. Thus **every predicted protein** in both species can be searched for experimental and other information (even if not experimentally identified).

What information is stored in PPDB? The PPDB stores experimental data from in-house proteome and mass spectrometry analysis, curated information about protein function, protein properties and subcellular localization. Importantly, proteins are particularly curated for possible (intra) plastid location and their plastid function. Protein accessions identified in published *Arabidopsis* (and other *Brassicacea*) proteomics papers are cross-referenced to rapidly determine previous experimental identification by mass spectrometry.

Protein report pages. Each protein has a protein report page that summarizes all information for that protein stored in PPDB. A few other databases are hyperlinked for each accession to obtain additional information for this protein accession. These report pages also provide detailed information about matched mass spectrometry data and how it maps to the protein. For individual protein searches, this is the best way to obtain a comprehensive overview.

How to extract information from PPDB? The PPDB has nine search functions to extract multiple types of information (output) stored in PPDB for any accession, by simply choosing a search function and selecting the relevant 'radio buttons'. Information can all be extracted from the PPDB for individual accessions or in batch format.

The search functions are:
Search by Accession. (TAIR accessions or ZmGI accessions).
Search with Protein name. (free text search – note this can be slow).
Search Proteome Experiments. Select sample source by species, organ, cell type or subcellular fraction.
Search Comparative Proteomics experiments. Select particular comparative proteomics experiments

PPDB is built on the principle of integrating and cross-referencing data from various sources, including experimental data from high-throughput proteomic studies, computational predictions, and manual curation. The main goal of PPDB is to provide a centralized, curated repository for data on predicted and experimentally determined proteins in Arabidopsis thaliana and maize, including their annotated functions, molecular and biophysical properties, and experimental and predicted post-translational modifications. The database is particularly useful for mass spectrometry-based identifications, allowing users to evaluate the significance of experimental identifications and information on post-translational modifications.

Users can access the content of PPDB through its web interface (http://ppdb.tc.cornell.edu/), which offers multiple search options such as gene identification number, functional annotation, and various protein properties. The database also includes active links to other databases such as TAIR and TIGR for easy access to additional information.

The PPDB is designed to be user-friendly and provides user options to search for proteins based on various criteria, such as protein name, gene name, or accession number along with different visualization tools to view protein sequences and structures in various formats, such as multiple sequence alignments, phylogenetic trees, and three-dimensional structures.

Some of the resources are mentioned below:

- A set of bioinformatics tools for data analysis, such as sequence alignment, phylogenetic analysis, and gene expression profiling
- A collection of protein-related publications and resources for further reading
- A forum for discussion and collaboration among researchers in the field of plant proteomics

CONCLUSION

Online databases are an essential resource for researchers working with Arabidopsis, as they provide a wealth of information on the genetic makeup, functional annotation, and molecular properties of the plant. Some examples of well-established and widely used databases for Arabidopsis include the Arabidopsis Information Resource (TAIR), the Arabidopsis thaliana Genome Portal (AtGP), and the Arabidopsis Subcellular Localization Database (ASCLD).

TAIR provides comprehensive information on the genetic makeup of Arabidopsis, including the complete genome sequence, gene models, functional annotation, and expression data. AtGP is another valuable resource for studying the genome of Arabidopsis, offering a variety of tools and resources for functional genomics research, such as gene expression data, genetic and physical maps, and genome browser. ASCLD, on the other hand, is a specialized database that focuses on the subcellular localization of proteins in Arabidopsis, providing information on the experimental and predicted localization of proteins across different cellular compartments.

Other databases such as PhosPhAt 3.0, Plantome, and PPDB also offer specialized information on phosphorylation sites, functional annotation, and proteome data respectively. These databases are essential resources for researchers studying the molecular mechanisms of Arabidopsis, providing detailed information on protein function, post-translational modifications, and molecular interactions.

In conclusion, online databases are a vital resource for the Arabidopsis research community, providing a wealth of information on the genetic makeup, functional annotation, and molecular properties of the plant. These resources are essential for understanding the complex biology of Arabidopsis and can aid in the discovery of new insights and applications.