Woody Plant
DATABASE

Woody Plant User Manual

Version v1.1 | 2025

Website URL: https://woodyplant.com/

Contact@woodyplant.com

Table of Contents

Introduction

Woody plants, typically encompassing trees, shrubs, and woody lianas, are generally perennial in nature. The presence of wood as a structural tissue facilitates the annual growth of woody plants from above-ground stems, resulting in some of them becoming the largest and tallest terrestrial plants. Here, we create the **Woody Plant Multi-Omics Database (WP-MOD)**, a digital platform crafted to unlock the genetic secrets of woody plants. At the latest version of WP-MOD, it includes the multi-omics information of 373 woody plants with whole-genome references, transcriptomes, methylome, sRNAomics, degradome, etc. The mission of WP-MOD is to collect and provide as much data and related information of woody plants as possible.

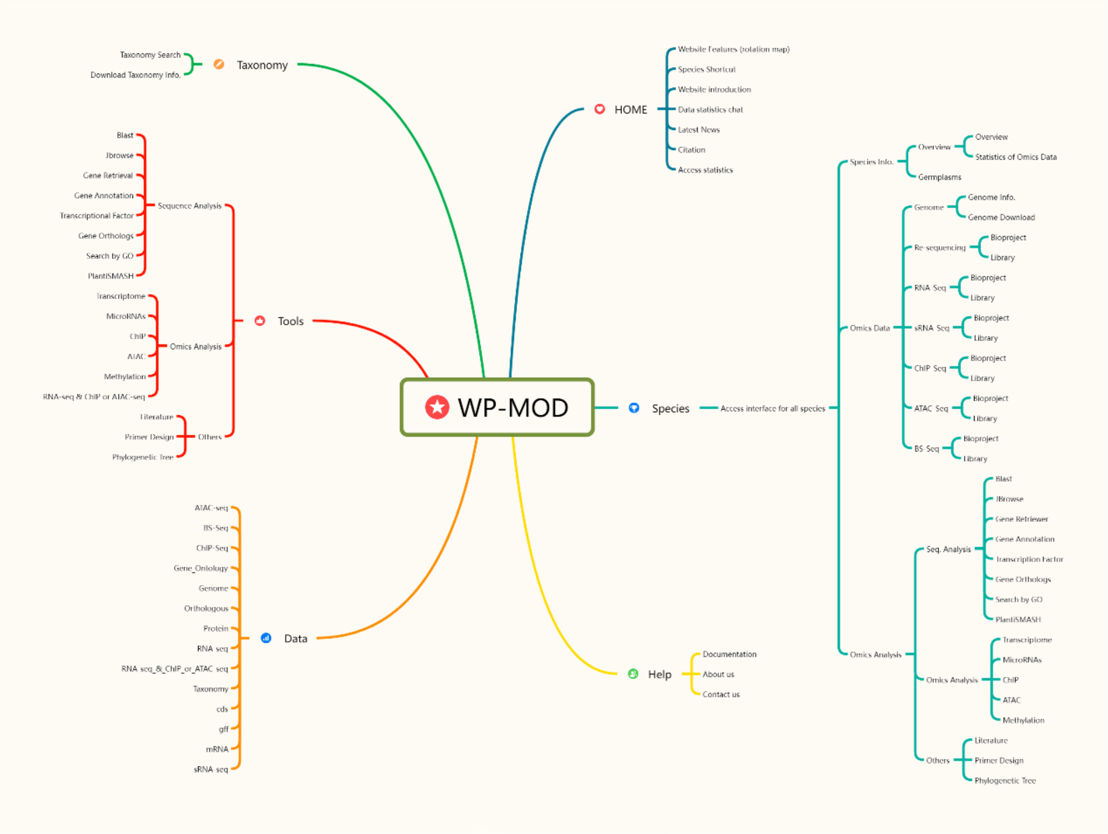
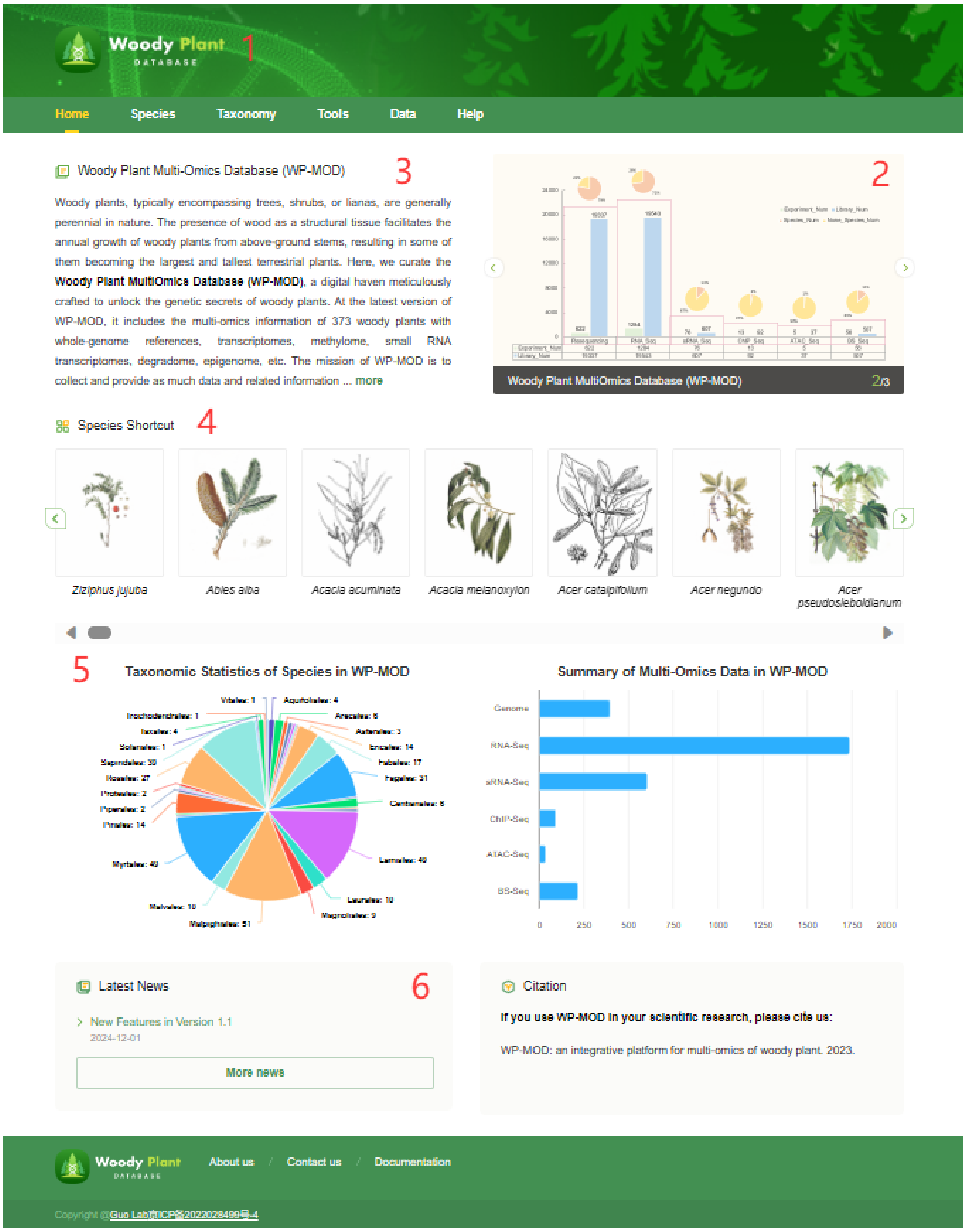


Figure 1. Framework of WP-MOD.

1. Home

The link to the Woody Plant Database home page is https://woodyplant.com/.

For access to data, users can visit the Home, Species, Taxonomy, Tools, Data, and Help sections in the navigation bar.



Woody Plant Multi-Omics Database (WP-MOD)

Woody plants, typically encompassing trees, shrubs, or lianas, are generally perennial in nature. The presence of wood as a structural tissue facilitates the annual growth of woody plants from above-ground stems, resulting in some of them becoming the largest and tallest terrestrial plants. Here, we curate the **Woody Plant MultiOmics Database (WP-MOD)**, a digital haven meticulously crafted to unlock the genetic secrets of woody plants. At the latest version of WP-MOD, it includes the multi-omics information of 373 woody plants with whole-genome references, transcriptomes, methylome, small RNA transcriptomes, degradome, epigenome, etc. The mission of WP-MOD is to collect and provide as much data and related information ... more

If you use WP-MOD in your scientific research, please cite us:

WP-MOD: an integrative platform for multi-omics of woody plant. 2023.

Figure 2. Overview of Homepage.

1) WP-MOD Logo

Users can return to the home page from any WP-MOD page by clicking the logo.

2) Navigation bar

WP-MOD provides a navigation bar at the top of all of its pages to allow users to easily move between sections of the database.

3) Introduction

Brief summary of data in the database. Goals and expectations of WP-MOD.

4) Species Shortcut

By clicking on the image for the species of interest, go to the Omics page where you can access all the data and tools available of this species.

5) Data Overview

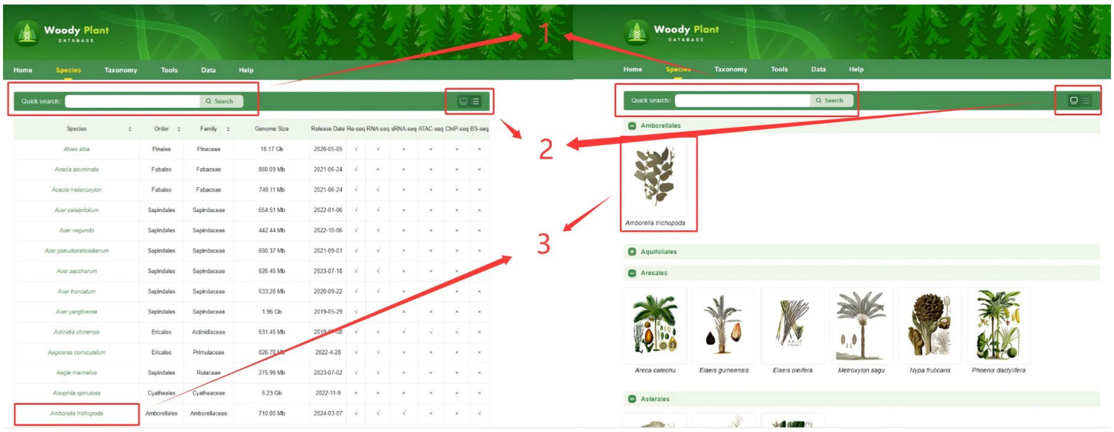
These charts offer an intuitive way to explore species diversity and data distribution. The Taxonomic Statistics pie chart shows the distribution of plant species across taxonomic orders, while the Multi-Omics Data Summary bar chart shows the available omics datasets, including Genome, RNA-Seq, and sRNA-Seq, providing a quick overview of species diversity and data coverage.

6) Latest News

Updated data and new versions of WP-MOD.

2. Species

The Species page provides access to 373 plant species in WP-MOD, allowing users to efficiently browse and search for species. Features include:



Species	Order	Family	Genome Size	Release Date	Re-seq	RNA-seq	sRNA-seq	ATAC-seq	ChIP-seq	BS-seq
Abies alba	Pinales	Pinaceae	18.17 Gb	2020-05-05	√	√	×	×	×	×
Acacia acuminata	Fabales	Fabaceae	880.09 Mb	2021-06-24	√	×	×	×	×	×
Acacia melanoxylon	Fabales	Fabaceae	749.11 Mb	2021-06-24	√	√	×	×	×	×
Acer catalpifolium	Sapindales	Sapindaceae	654.51 Mb	2022-01-06	√	√	×	×	×	×
Acer negundo	Sapindales	Sapindaceae	442.44 Mb	2022-10-06	√	√	×	×	×	×
Acer pseudosieboldianum	Sapindales	Sapindaceae	690.37 Mb	2021-09-01	√	√	×	×	×	×
Acer saccharum	Sapindales	Sapindaceae	626.45 Mb	2023-07-18	√	√	×	×	×	[illegible]
Acer truncatum	Sapindales	Sapindaceae	633.28 Mb	2020-09-22	√	√	×	[illegible]	×	×
Acer yangbiense	Sapindales	Sapindaceae	1.96 Gb	2019-05-29	√	[illegible]	×	×	×	×
Actinidia chinensis	Ericales	Actinidiaceae	631.45 Mb	[illegible]	√	√	√	√	√	×
Aegiceras corniculatum	Ericales	Primulaceae	826.70 Mb	2022-4-28	√	√	×	×	×	×
Aegle marmelos	Sapindales	Rutaceae	275.99 Mb	2023-07-02	√	√	×	×	×	×
Alsophila spinulosa	Cyatheales	Cyatheaceae	6.23 Gb	2022-11-9	×	×	×	×	×	×
Amborella trichopoda	Amborellales	Amborellaceae	710.00 Mb	2024-03-07	√	√	√	×	×	√

Figure 3. Species Page Overview.

1) Quick Search

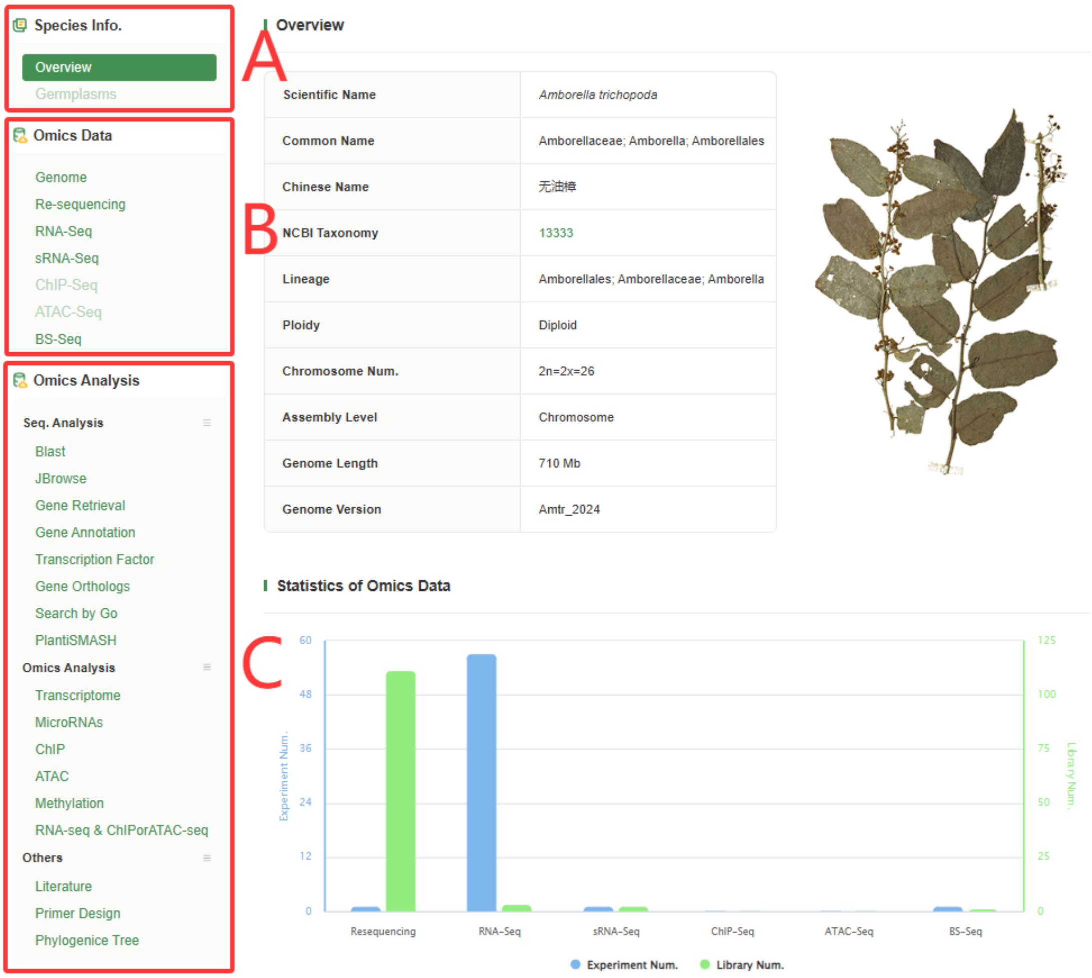
Instantly find species by entering a name in the search box.

2) View Modes

Switch between table view and figure view.

3) Species Details

Click on a species image or name to access its dedicated page with comprehensive information and available datasets (eg: *Amborella trichopoda*). The species detail page is divided into three main sections:



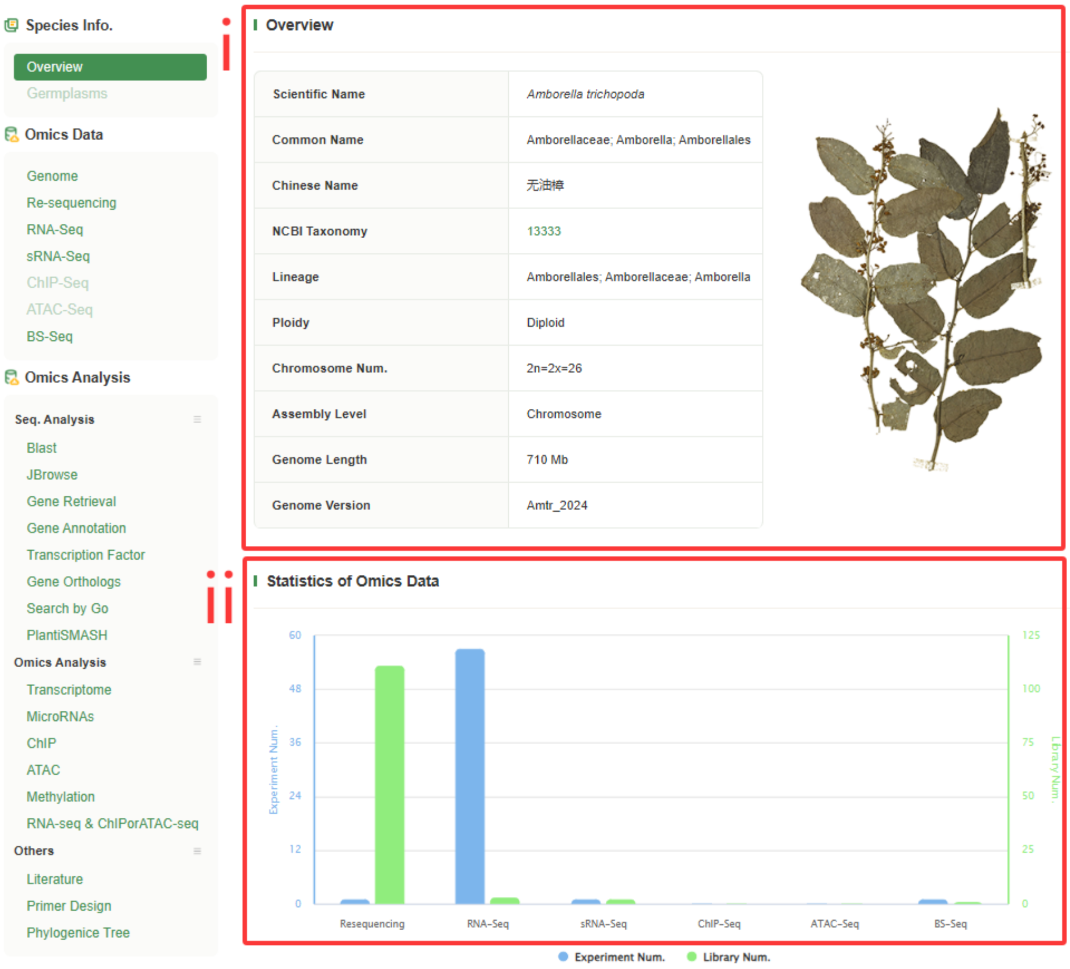
Scientific Name	*Amborella trichopoda*
Common Name	Amborellaceae; Amborella; Amborellales
Chinese Name	无油樟
NCBI Taxonomy	13333
Lineage	Amborellales; Amborellaceae; Amborella
Ploidy	Diploid
Chromosome Num.	2n=2x=26
Assembly Level	Chromosome
Genome Length	710 Mb
Genome Version	Amtr_2024

Figure 4. Species Detail Page Overview.

A. Species Info.

This section provides basic information about this species:

a) **Overview** (eg: *Amborella trichopoda*)



Scientific Name	*Amborella trichopoda*
Common Name	Amborellaceae; Amborella; Amborellales
Chinese Name	无油樟
NCBI Taxonomy	13333
Lineage	Amborellales; Amborellaceae; Amborella
Ploidy	Diploid
Chromosome Num.	2n=2x=26
Assembly Level	Chromosome
Genome Length	710 Mb
Genome Version	Amtr_2024

Figure 5. Species Overview Page.

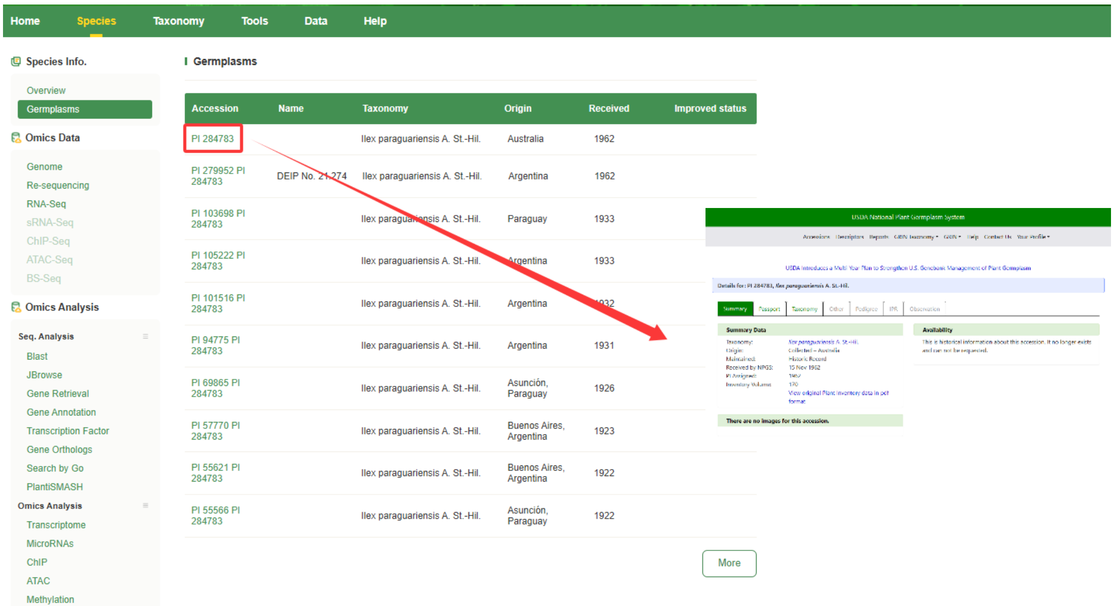
i. Overview

The Overview section provides key taxonomic and genomic details about the species, including its scientific name, common name, taxonomy ID, lineage, ploidy, and chromosome number. It also includes genome assembly information such as assembly level, genome size, and version, helping users understand the sequencing and annotation status. A reference species image is displayed for visual identification.

ii. Statistics of Omics Data

The Statistics of Omics Data section presents a graphical summary of available sequencing datasets, including resequencing, RNA-Seq, sRNA-Seq, ChIP-Seq, ATAC-Seq, and BS-Seq. The experiment count (blue bars) indicates the number of sequencing projects conducted, while the library count (green bars) represents the available sequencing libraries, providing a quick overview of omics data coverage for the species.

b) **Germplasms** (eg: *Ilex paraguariensis*)



Accession	Name	Taxonomy	Origin	Received	Improved status
PI 284783		Ilex paraguariensis A. St.-Hil.	Australia	1962	
PI 279952 PI 284783	DEIP No. 21.274	Ilex paraguariensis A. St.-Hil.	Argentina	1962	
PI 103698 PI 284783		Ilex paraguariensis A. St.-Hil.	Paraguay	1933	
PI 105222 PI 284783		Ilex paraguariensis A. St.-Hil.	Argentina	1933	
PI 101516 PI 284783		Ilex paraguariensis A. St.-Hil.	Argentina	1932	
PI 94775 PI 284783		Ilex paraguariensis A. St.-Hil.	Argentina	1931	
PI 69865 PI 284783		Ilex paraguariensis A. St.-Hil.	Asunción, Paraguay	1926	
PI 57770 PI 284783		Ilex paraguariensis A. St.-Hil.	Buenos Aires, Argentina	1923	
PI 55621 PI 284783		Ilex paraguariensis A. St.-Hil.	Buenos Aires, Argentina	1922	
PI 55566 PI 284783		Ilex paraguariensis A. St.-Hil.	Asunción, Paraguay	1922	

USDA Introduces a Multi Year Plan to Strengthen U.S. Genebank Management of Plant Germplasm

Details for: PI 284783, *Ilex paraguariensis* A. St.-Hil.

Summary Data

Taxonomy:	*Ilex paraguariensis* A. St.-Hil.
Origin:	Collected – Australia
Maintained:	Historic Record
Received by NPGS:	15 Nov 1962
PI Assigned:	1962
Inventory Volume	170
	View original Plant Inventory data in pdf format

There are no images for this accession.

Availability

This is historical information about this accession. It no longer exists and can not be requested.

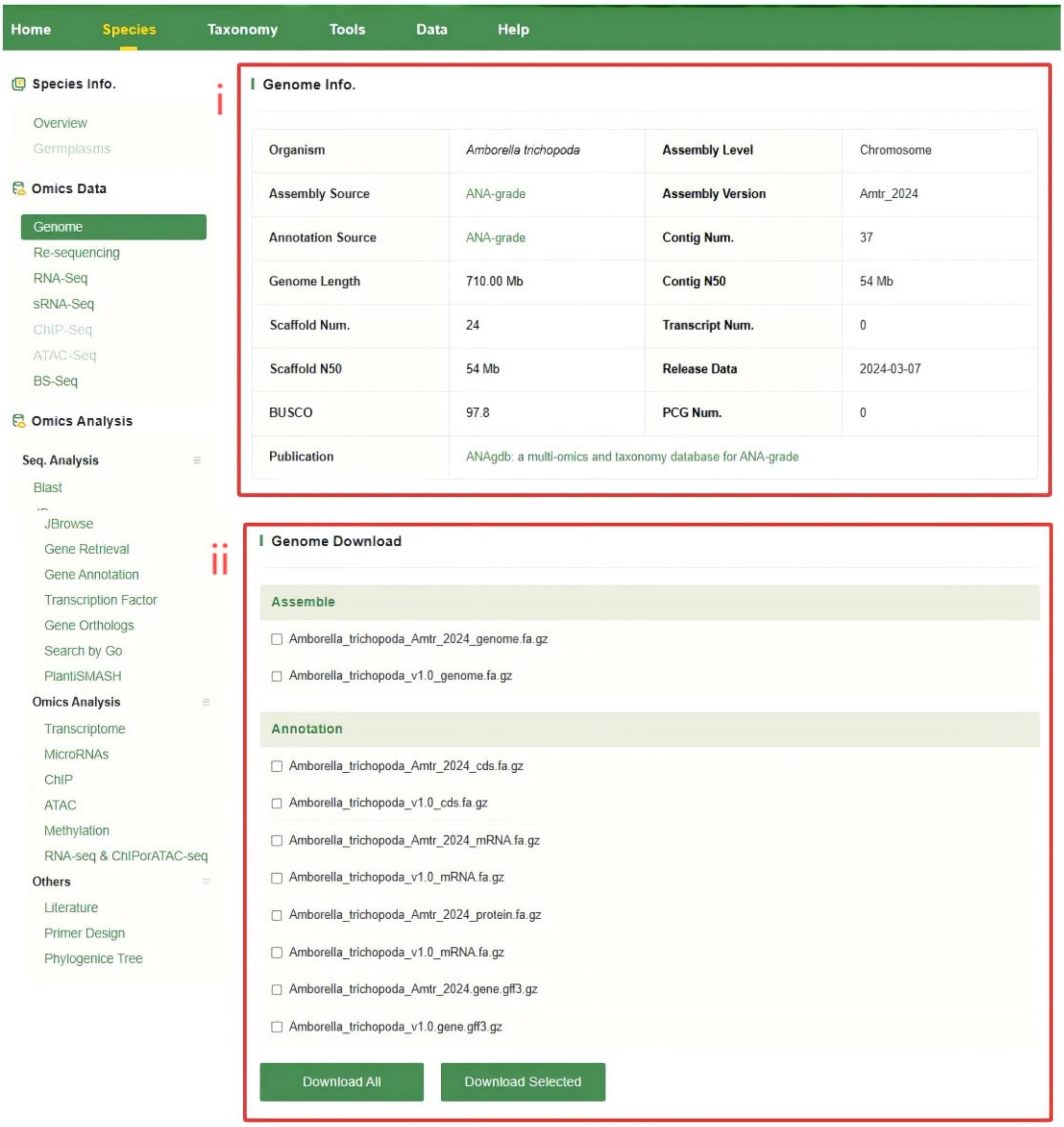
Figure 6. Germplasms Page Overview.

The Germplasms page provides a list of germplasm accessions sourced from the USDA National Plant Germplasm System (GRIN-Global). Each entry includes details such as accession number, taxonomy, origin, received date, and improvement status, allowing users to explore the historical and geographic background of the species' genetic resources. Clicking on an accession number redirects users to the corresponding GRIN-Global page, where they can access comprehensive records.

B. **Omics Data** (eg: *Amborella trichopoda*)

The Omics Data section provides access to genomic and transcriptomic datasets for the species. It includes Genome data and other omics datasets such as Re-sequencing, RNA-Seq, sRNA-Seq, ChIP-Seq, ATAC-Seq, and BS-Seq.

a) **Genome**



Genome Info.

Organism	*Amborella trichopoda*	Assembly Level	Chromosome
Assembly Source	ANA-grade	Assembly Version	Amtr_2024
Annotation Source	ANA-grade	Contig Num.	37
Genome Length	710.00 Mb	Contig N50	54 Mb
Scaffold Num.	24	Transcript Num.	0
Scaffold N50	54 Mb	Release Data	2024-03-07
BUSCO	97.8	PCG Num.	0
Publication	ANAgdb: a multi-omics and taxonomy database for ANA-grade		

Genome Download

Assemble

☐ Amborella_trichopoda_Amtr_2024_genome.fa.gz
☐ Amborella_trichopoda_v1.0_genome.fa.gz

Annotation

☐ Amborella_trichopoda_Amtr_2024_cds.fa.gz
☐ Amborella_trichopoda_v1.0_cds.fa.gz
☐ Amborella_trichopoda_Amtr_2024_mRNA.fa.gz
☐ Amborella_trichopoda_v1.0_mRNA.fa.gz
☐ Amborella_trichopoda_Amtr_2024_protein.fa.gz
☐ Amborella_trichopoda_v1.0_mRNA.fa.gz
☐ Amborella_trichopoda_Amtr_2024.gene.gff3.gz
☐ Amborella_trichopoda_v1.0.gene.gff3.gz

Figure 7. Genome Page Overview.

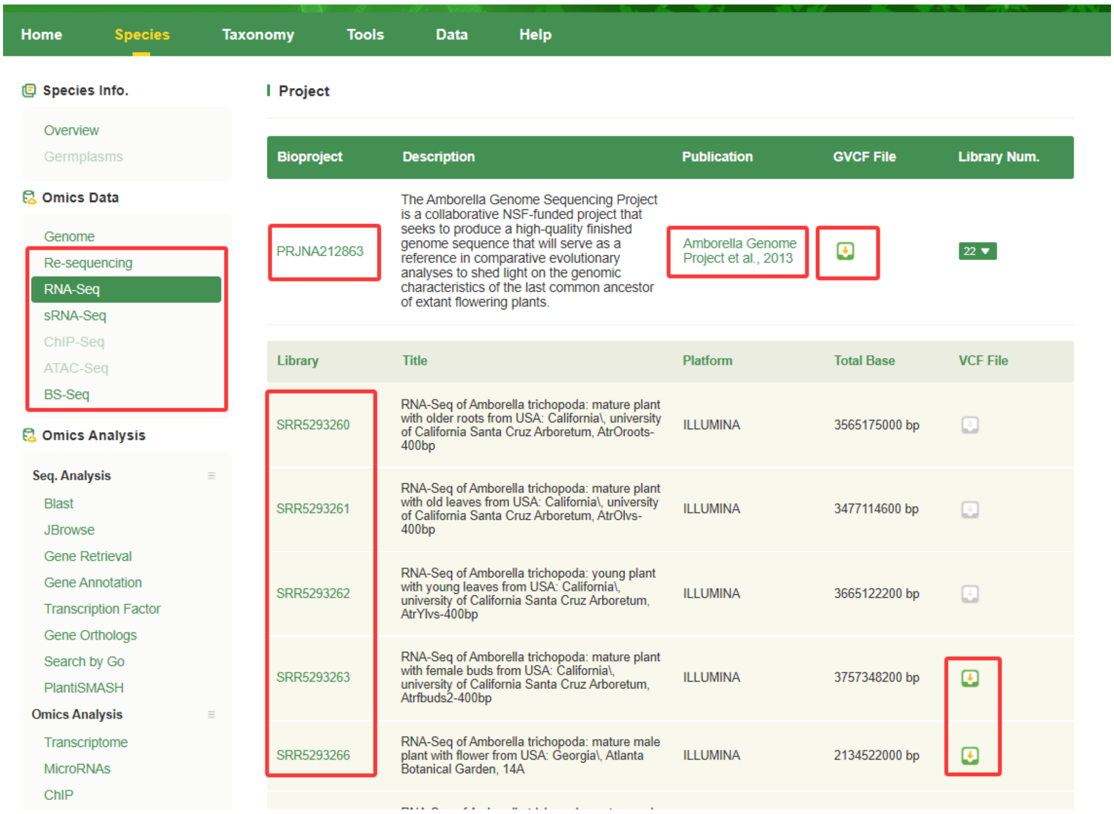
i. Genome Info.

Displays essential genome assembly details, including assembly source, genome size, scaffold numbers, contig statistics, BUSCO score, and related publications.

ii. Download

Provides genome sequence and annotation files, including assembled genome data, CDS, mRNA, protein sequences, and GFF annotation files. Users can download all files or select specific ones.

b) **Other Omics Data** (eg: RNA-Seq)



Bioproject	Description	Publication	GVCF File	Library Num.
PRJNA212863	The Amborella Genome Sequencing Project is a collaborative NSF-funded project that seeks to produce a high-quality finished genome sequence that will serve as a reference in comparative evolutionary analyses to shed light on the genomic characteristics of the last common ancestor of extant flowering plants.	Amborella Genome Project et al., 2013	⬇	22 ▼

Library	Title	Platform	Total Base	VCF File
SRR5293260	RNA-Seq of Amborella trichopoda: mature plant with older roots from USA: California\, university of California Santa Cruz Arboretum, AtrOroots-400bp	ILLUMINA	3565175000 bp	⇩
SRR5293261	RNA-Seq of Amborella trichopoda: mature plant with old leaves from USA: California\, university of California Santa Cruz Arboretum, AtrOlvs-400bp	ILLUMINA	3477114600 bp	⇩
SRR5293262	RNA-Seq of Amborella trichopoda: young plant with young leaves from USA: California\, university of California Santa Cruz Arboretum, AtrYlvs-400bp	ILLUMINA	3665122200 bp	⇩
SRR5293263	RNA-Seq of Amborella trichopoda: mature plant with female buds from USA: California\, university of California Santa Cruz Arboretum, Atrfbuds2-400bp	ILLUMINA	3757348200 bp	⬇
SRR5293266	RNA-Seq of Amborella trichopoda: mature male plant with flower from USA: Georgia\, Atlanta Botanical Garden, 14A	ILLUMINA	2134522000 bp	⬇

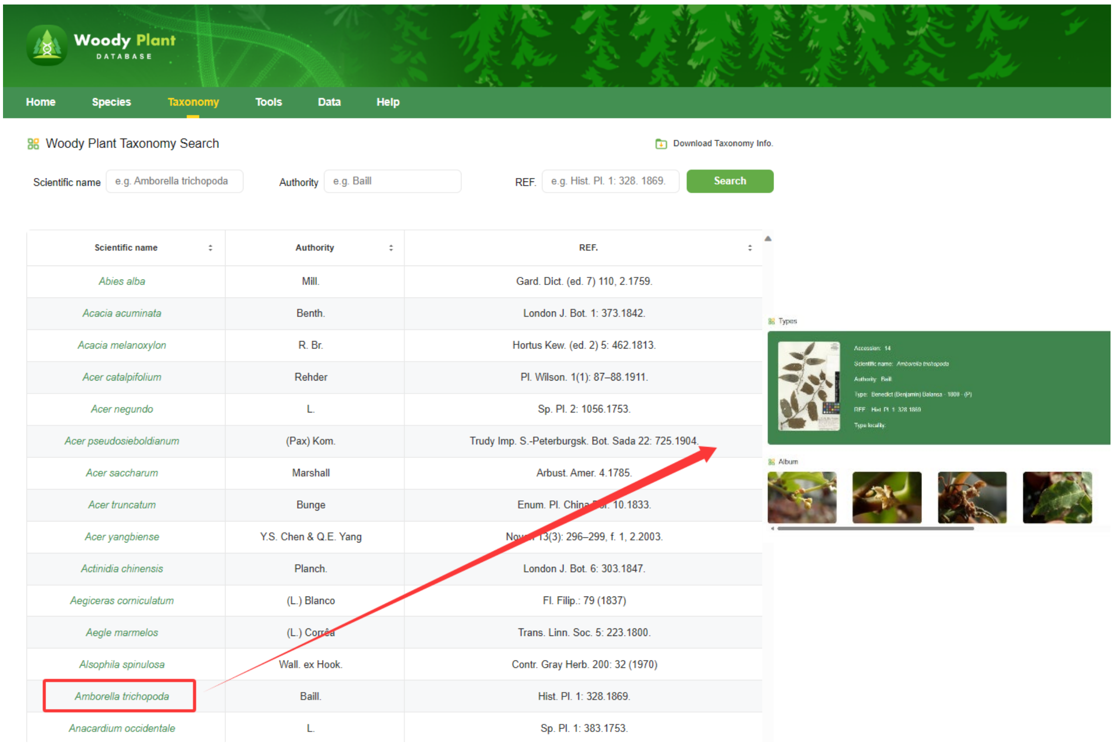
Figure 8. RNA-seq Page Overview.

Clicking the Bioproject ID or Library Number redirects users to the corresponding NCBI database page for more detailed information, while clicking the publication link takes users to the original research article. If a green download button appears, it indicates that the RNA-Seq dataset has been reprocessed and is available for download.

C. Omics Analysis

The Omics Analysis section provides quick access to various bioinformatics tools, including sequence analysis, omics data processing, and additional resources like literature search and primer design. These tools are identical to those available in the Tools interface, and clicking on any option will redirect users to the corresponding tool page for further analysis. For detailed descriptions and usage instructions, users can refer to the Tools page.

3. Taxonomy



Woody Plant Taxonomy Search

Scientific name: e.g. Amborella trichopoda
Authority: e.g. Baill
REF.: e.g. Hist. Pl. 1: 328. 1869.

Scientific name	Authority	REF.
Abies alba	Mill.	Gard. Dict. (ed. 7) 110, 2.1759.
Acacia acuminata	Benth.	London J. Bot. 1: 373.1842.
Acacia melanoxylon	R. Br.	Hortus Kew. (ed. 2) 5: 462.1813.
Acer catalpifolium	Rehder	Pl. Wilson. 1(1): 87–88.1911.
Acer negundo	L.	Sp. Pl. 2: 1056.1753.
Acer pseudosieboldianum	(Pax) Kom.	Trudy Imp. S.-Peterburgsk. Bot. Sada 22: 725.1904.
Acer saccharum	Marshall	Arbust. Amer. 4.1785.
Acer truncatum	Bunge	Enum. Pl. China Bor. 10.1833.
Acer yangbiense	Y.S. Chen & Q.E. Yang	Novon 13(3): 296–299, f. 1, 2.2003.
Actinidia chinensis	Planch.	London J. Bot. 6: 303.1847.
Aegiceras corniculatum	(L.) Blanco	Fl. Filip.: 79 (1837)
Aegle marmelos	(L.) Corrêa	Trans. Linn. Soc. 5: 223.1800.
Alsophila spinulosa	Wall. ex Hook.	Contr. Gray Herb. 200: 32 (1970)
Amborella trichopoda	Baill.	Hist. Pl. 1: 328.1869.
Anacardium occidentale	L.	Sp. Pl. 1: 383.1753.

Accession: 14
Scientific name: *Amborella trichopoda*
Authority: Baill
REF.: Hist. Pl. 1: 328.1869
Type locality:

Figure 9. Taxonomy Page Overview.

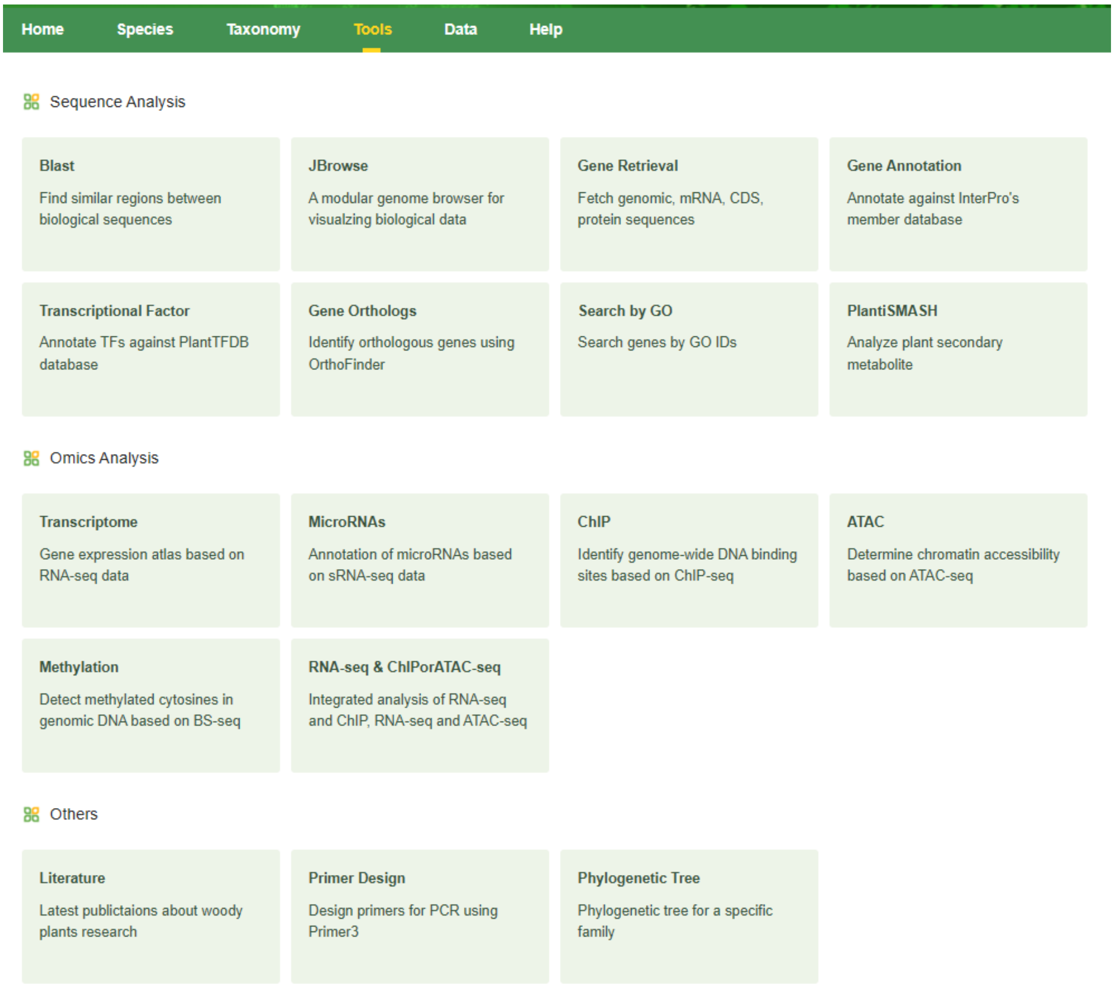
1) Taxonomy Page Overview

The Taxonomy page includes 373 species, allowing users to search by scientific name, authority, or reference to find specific taxa. Clicking on a species name redirects users to a detailed page that provides taxonomic classification, authority, reference, and type information.

2) Species Taxonomy Details Page (eg: *Amborella trichopoda*)

The species taxonomy details page provides essential taxonomic information, including the scientific name, authority, reference, type locality, and specimen image. It also features an image album showcasing real photos of the species. Additional details such as herbarium placement, leaf habit, and links to external databases like POWO and Tropicos are integrated to offer users a more comprehensive taxonomic resource.

4. Tools



Sequence Analysis

Blast
Find similar regions between biological sequences

JBrowse
A modular genome browser for visualzing biological data

Gene Retrieval
Fetch genomic, mRNA, CDS, protein sequences

Gene Annotation
Annotate against InterPro's member database

Transcriptional Factor
Annotate TFs against PlantTFDB database

Gene Orthologs
Identify orthologous genes using OrthoFinder

Search by GO
Search genes by GO IDs

PlantiSMASH
Analyze plant secondary metabolite

Omics Analysis

Transcriptome
Gene expression atlas based on RNA-seq data

MicroRNAs
Annotation of microRNAs based on sRNA-seq data

ChIP
Identify genome-wide DNA binding sites based on ChIP-seq

ATAC
Determine chromatin accessibility based on ATAC-seq

Methylation
Detect methylated cytosines in genomic DNA based on BS-seq

RNA-seq & ChIPorATAC-seq
Integrated analysis of RNA-seq and ChIP, RNA-seq and ATAC-seq

Others

Literature
Latest publictaions about woody plants research

Primer Design
Design primers for PCR using Primer3

Phylogenetic Tree
Phylogenetic tree for a specific family

Figure 10. Tools Page Overview.

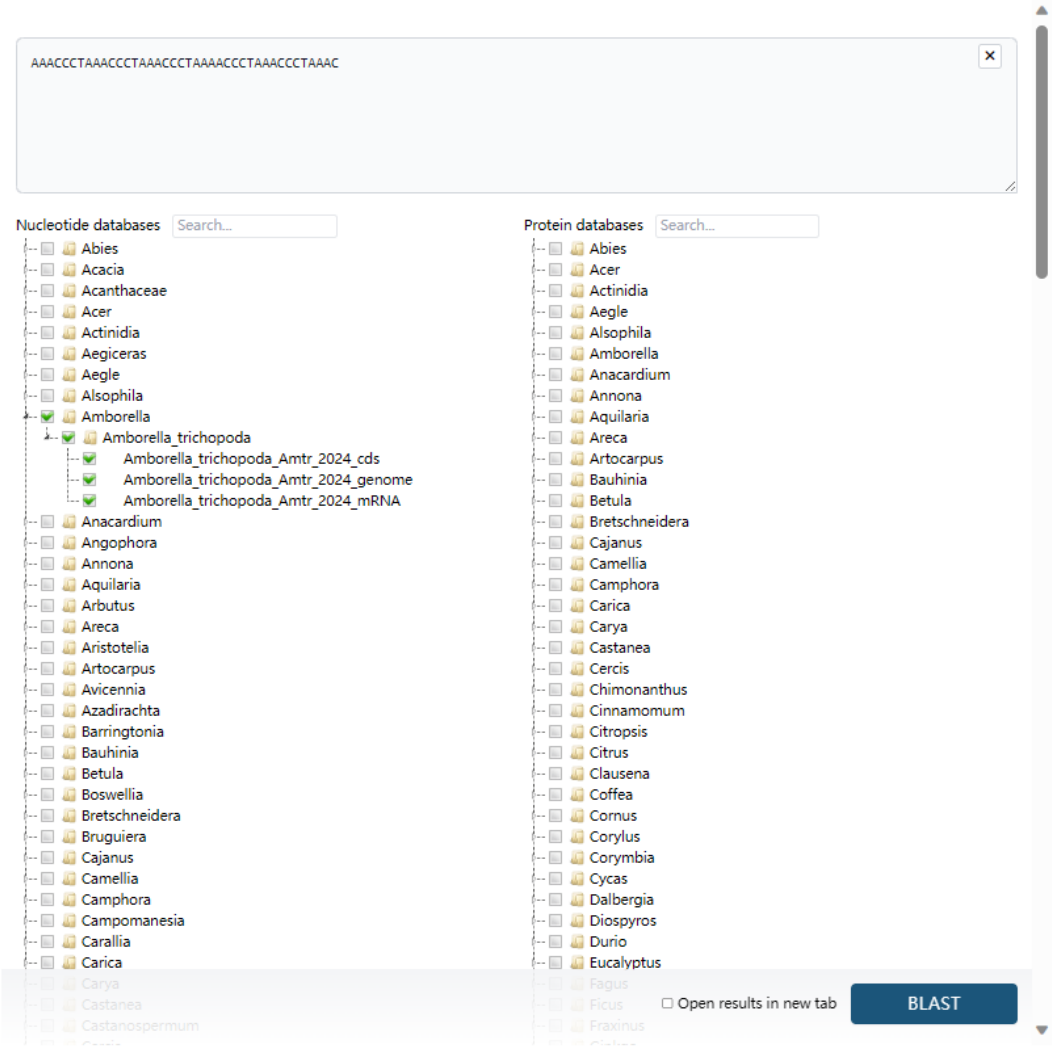
The Tools page provides a collection of bioinformatics tools for analyzing genomic, transcriptomic, and other omics data. These tools support sequence analysis, functional annotation, gene expression studies, and evolutionary research, helping users explore various biological datasets efficiently.

1) Sequence Analysis

This section includes tools for sequence alignment, genome browsing, gene retrieval, and functional annotation. It allows users to compare sequences using BLAST, visualize genomic data with JBrowse, and retrieve genetic information such as mRNA and CDS. Additional tools support transcription factor annotation, gene ortholog identification, and secondary metabolite analysis.

A. **Blast**

The BLAST page allows users to perform sequence similarity searches by inputting a nucleotide or protein sequence. Users can select reference databases from a wide range of plant species, including genomic, mRNA, and CDS data. After selecting the appropriate database, users can run the BLAST search to find homologous sequences, with an option to open results in a new tab.



```
AAACCCTAAACCCTAAACCCTAAAACCCTAAACCCTAAAC
```

Figure 11. Blast page Overview.

B. **JBrowse** (eg: *Acer negundo*, CM046694.1)

JBrowse is an interactive genome browser that allows users to visualize, analyze, and navigate genomic data efficiently. It supports multiple genome views, enabling users to examine sequence alignments, gene annotations, and structural variations. The tool provides flexibility in selecting assemblies, searching genomic regions, and displaying various data tracks for detailed analysis.

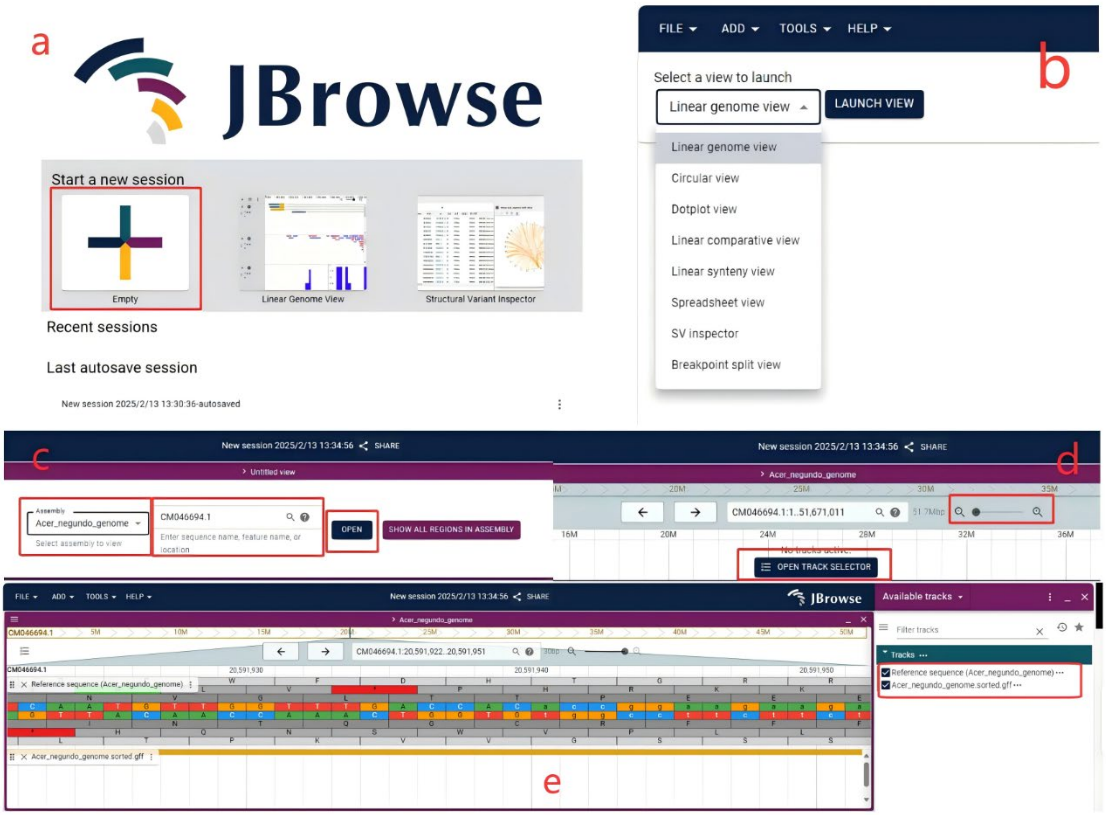


Figure 12. JBrowse page Overview.

a) **Start a New Session** (Fig. a)

Upon launching JBrowse, users are presented with options to start a new session. Clicking on "Empty" allows users to begin a fresh session where they can load and explore genomic data.

b) **Select a View Type** (Fig. b)

Users can choose from multiple viewing modes, such as: Linear Genome View (default), Circular View, Dotplot View, Comparative and Synteny Views, Spreadsheet View for tabular data.

c) **Load a Genome Assembly** (Fig. c)

Users must first select an assembly from the available datasets. They can enter a specific sequence name, feature name, or genomic location to navigate directly to a particular region. Clicking "Open" will load the specified genomic data.

d) **Navigate and Adjust the Genome View** (Fig. d)

Once a genome assembly is loaded, users can: Zoom in/out using the magnification controls. Move along the sequence using navigation arrows. And click "Open Track Selector" to add or remove data tracks for visualization.

e) **View and Manage Tracks** (Fig. e)

Tracks display reference sequences, gene annotations, and other genomic features. Users can: Select tracks from the "Available Tracks" panel. Toggle visibility for individual tracks. And explore sequence alignments, annotations, and additional genomic insights.

C. **Gene Retrieval** (eg: *Acer catalpifolium*, GWHGASIS002255)

The Gene Retrieval tool allows users to search for genes by selecting a species and entering a gene identifier. It provides a comprehensive view of gene information, including sequence retrieval and structural visualization. This tool supports genomic research by enabling users to extract genomic, transcript, and protein sequences while also analyzing gene structure in a graphical format.

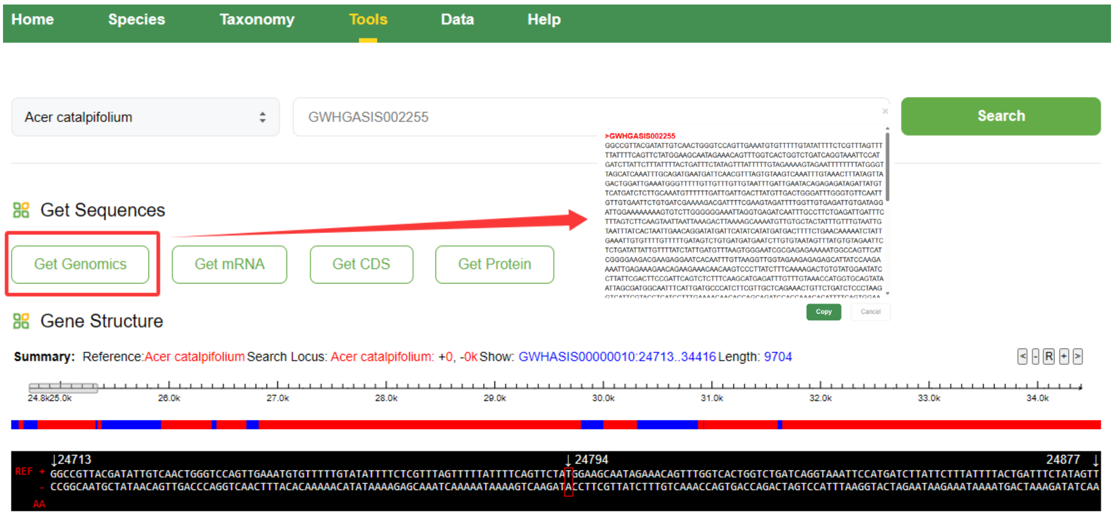


Figure 13. Gene Retrieval page Overview.

a) Get Sequences

The Get Sequences section allows users to retrieve different types of sequences associated with a gene. By clicking "Get Genomics", "Get mRNA", "Get CDS", or "Get Protein", users can access and download the corresponding sequences in FASTA format. This feature is essential for downstream applications such as gene function analysis, molecular cloning, and comparative genomics.

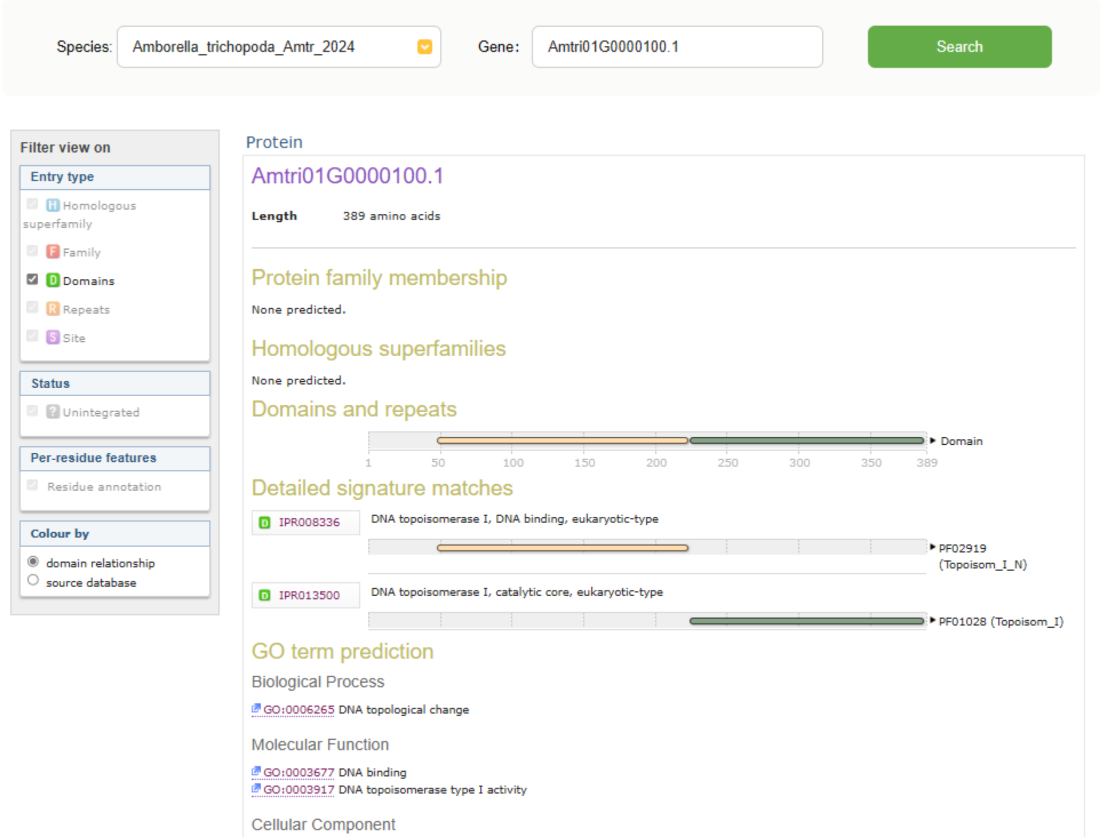
b) Gene Structure

The Gene Structure section provides a visual representation of the gene’s exon-intron organization, along with sequence annotations. Users can explore specific loci and sequence alignments, allowing for an in-depth understanding of gene architecture. The tool enables easy navigation through genomic regions and facilitates functional genomic studies.

D. Gene Annotation (eg: *Amborella trichopoda*, Amtri01G0000100.1)

The Gene Annotation page displays functional annotations generated using InterProScan27 (v5.24), providing insights into the protein family, domains, repeats, and Gene Ontology (GO) terms associated with a given gene. This analysis helps users understand the potential functions and biological roles of the gene's protein product.

Users can explore detailed domain signatures, homologous superfamilies, and GO term predictions, including biological processes and molecular functions linked to the annotated protein. The interactive visualization allows filtering by entry type, status, per-residue features, and domain relationships, making it easier to analyze functional elements within the protein sequence.



Species: Amborella_trichopoda_Amtr_2024

Gene: Amtri01G0000100.1

Protein

Amtri01G0000100.1

Length 389 amino acids

Protein family membership

None predicted.

Homologous superfamilies

None predicted.

Domains and repeats

Detailed signature matches

D IPR008336 DNA topoisomerase I, DNA binding, eukaryotic-type

D IPR013500 DNA topoisomerase I, catalytic core, eukaryotic-type

GO term prediction

Biological Process

GO:0006265 DNA topological change

Molecular Function

GO:0003677 DNA binding
GO:0003917 DNA topoisomerase type I activity

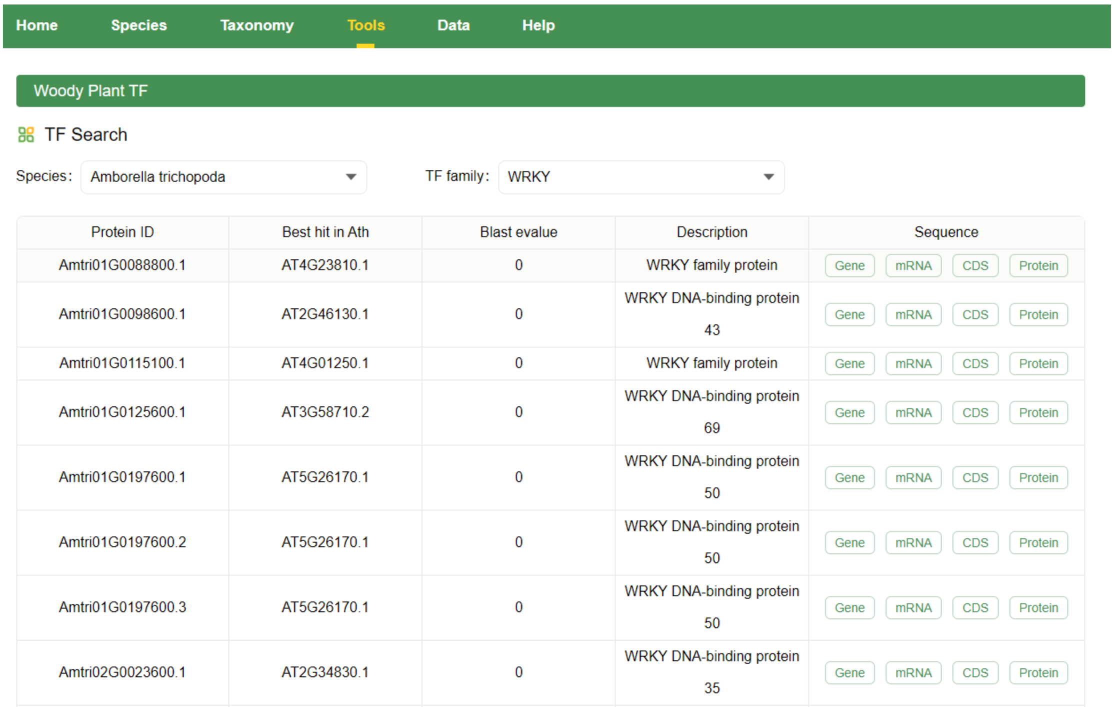
Cellular Component

Figure 14. Gene Annotation Page Overview.

E. **Transcriptional Factor** (eg: *Amborella trichopoda*, WRKY)

The Transcription Factor (TF) Search tool allows users to identify and analyze transcription factors in woody plants. It is based on PlantTFDB and utilizes ESTScan 3.0 for transcription factor prediction. Users can search for specific TF families and retrieve matching TFs for a selected species.

The results include Protein ID, best hit in Arabidopsis (Ath), BLAST e-value, description, and options to retrieve gene, mRNA, CDS, and protein sequences. This tool helps researchers explore regulatory proteins involved in gene expression, facilitating functional genomics and comparative studies.



Species: Amborella trichopoda

TF family: WRKY

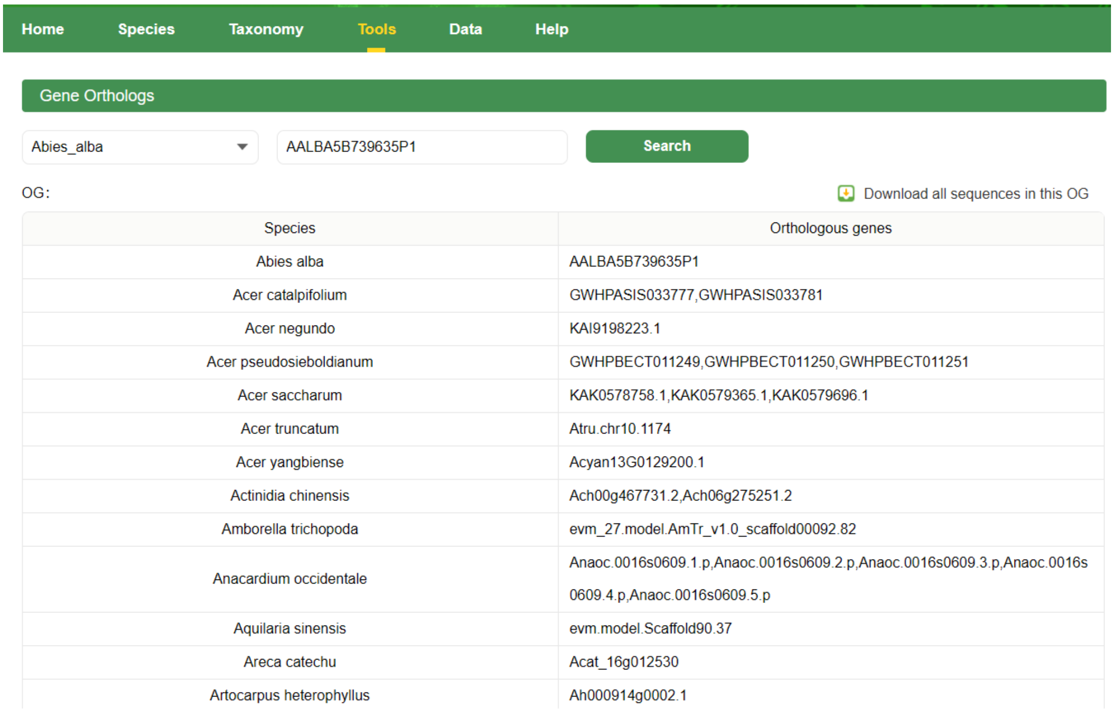
Protein ID	Best hit in Ath	Blast evalue	Description	Sequence
Amtri01G0088800.1	AT4G23810.1	0	WRKY family protein	Gene mRNA CDS Protein
Amtri01G0098600.1	AT2G46130.1	0	WRKY DNA-binding protein 43	Gene mRNA CDS Protein
Amtri01G0115100.1	AT4G01250.1	0	WRKY family protein	Gene mRNA CDS Protein
Amtri01G0125600.1	AT3G58710.2	0	WRKY DNA-binding protein 69	Gene mRNA CDS Protein
Amtri01G0197600.1	AT5G26170.1	0	WRKY DNA-binding protein 50	Gene mRNA CDS Protein
Amtri01G0197600.2	AT5G26170.1	0	WRKY DNA-binding protein 50	Gene mRNA CDS Protein
Amtri01G0197600.3	AT5G26170.1	0	WRKY DNA-binding protein 50	Gene mRNA CDS Protein
Amtri02G0023600.1	AT2G34830.1	0	WRKY DNA-binding protein 35	Gene mRNA CDS Protein

Figure 15. Transcriptional Factor Page Overview.

F. **Gene Orthologs** (eg: *Abies alba*, AALBA5B739635P1)

The Gene Orthologs tool allows users to search for orthologous genes across 96 woody plant species. By selecting a species and entering a gene identifier, users can retrieve a list of genes that share evolutionary relationships across different plant genomes.

The results display species names and their corresponding orthologous genes, enabling comparative genomic analysis. Additionally, users can download all sequences in the orthologous group (OG) for further studies, supporting research in gene function evolution and species adaptation.



Abies_alba

AALBA5B739635P1

OG:

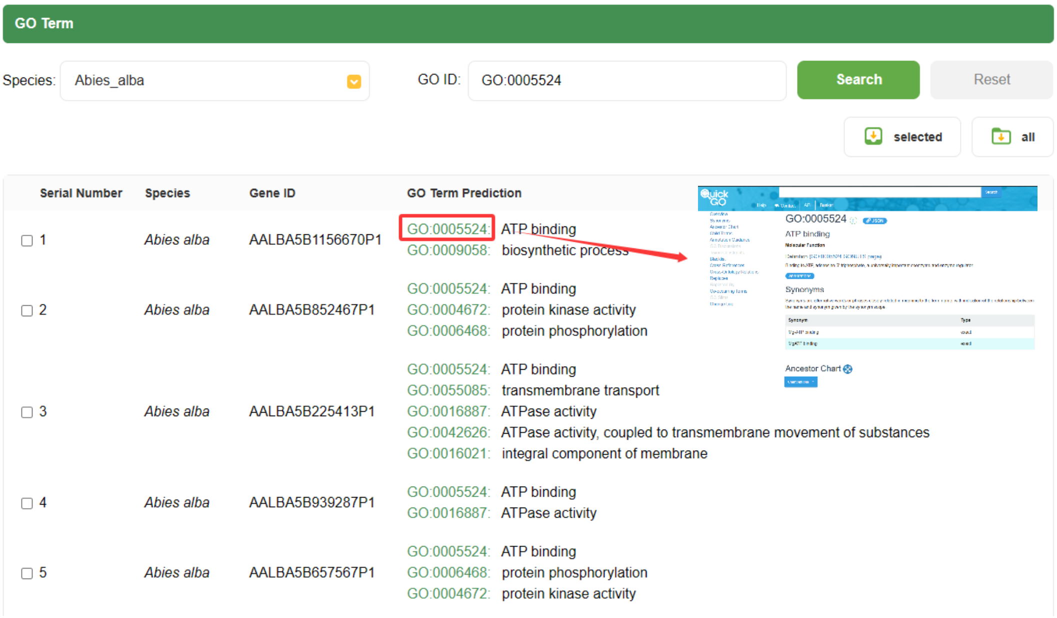
Species	Orthologous genes
Abies alba	AALBA5B739635P1
Acer catalpifolium	GWHPASIS033777,GWHPASIS033781
Acer negundo	KAI9198223.1
Acer pseudosieboldianum	GWHPBECT011249,GWHPBECT011250,GWHPBECT011251
Acer saccharum	KAK0578758.1,KAK0579365.1,KAK0579696.1
Acer truncatum	Atru.chr10.1174
Acer yangbiense	Acyan13G0129200.1
Actinidia chinensis	Ach00g467731.2,Ach06g275251.2
Amborella trichopoda	evm_27.model.AmTr_v1.0_scaffold00092.82
Anacardium occidentale	Anaoc.0016s0609.1.p,Anaoc.0016s0609.2.p,Anaoc.0016s0609.3.p,Anaoc.0016s0609.4.p,Anaoc.0016s0609.5.p
Aquilaria sinensis	evm.model.Scaffold90.37
Areca catechu	Acat_16g012530
Artocarpus heterophyllus	Ah000914g0002.1

Figure 16. Gene Orthologs Page Overview.

G. Search by GO (eg: *Abies alba*, GO:0005524)

The Search by GO tool enables users to search for genes associated with specific Gene Ontology (GO) terms within a selected species. By entering a GO ID, users can retrieve a list of genes annotated with that term, including information such as species, gene ID, and functional prediction. This helps researchers explore genes related to specific biological processes, molecular functions, or cellular components.

Clicking on a GO ID redirects users to an external QuickGO page, where they can access detailed information about the GO term, including its definition, synonyms, ancestor relationships, and associated annotations. This integration provides deeper insights into gene functions and their biological roles.



Species: Abies_alba

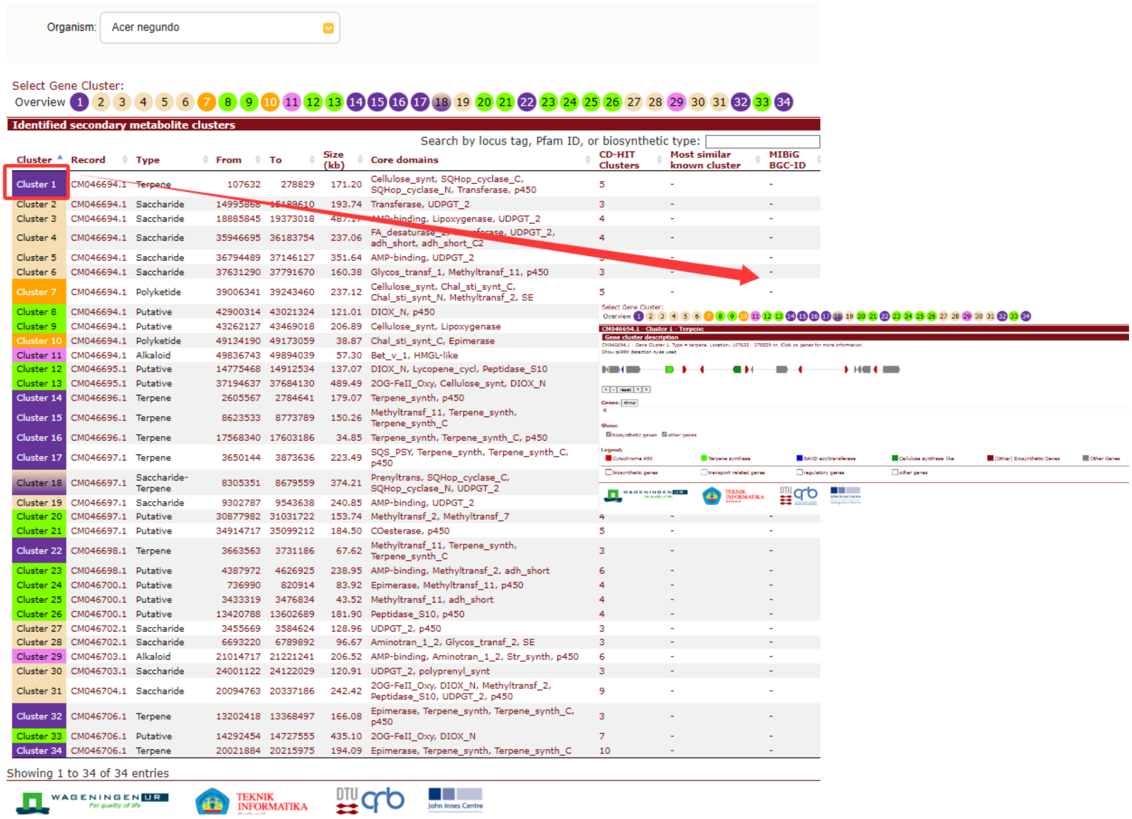
GO ID: GO:0005524

	Serial Number	Species	Gene ID	GO Term Prediction
☐	1	*Abies alba*	AALBA5B1156670P1	GO:0005524: ATP binding GO:0009058: biosynthetic process
☐	2	*Abies alba*	AALBA5B852467P1	GO:0005524: ATP binding GO:0004672: protein kinase activity GO:0006468: protein phosphorylation
☐	3	*Abies alba*	AALBA5B225413P1	GO:0005524: ATP binding GO:0055085: transmembrane transport GO:0016887: ATPase activity GO:0042626: ATPase activity, coupled to transmembrane movement of substances GO:0016021: integral component of membrane
☐	4	*Abies alba*	AALBA5B939287P1	GO:0005524: ATP binding GO:0016887: ATPase activity
☐	5	*Abies alba*	AALBA5B657567P1	GO:0005524: ATP binding GO:0006468: protein phosphorylation GO:0004672: protein kinase activity

Figure 17. Search by GO Page Overview.

H. **PlantiSMASH** (eg: *Acer negundo*, Cluster 1)

The PlantiSMASH tool is designed for identifying and analyzing secondary metabolite biosynthetic gene clusters (BGCs) in plant genomes. It helps researchers explore genes involved in the biosynthesis of specialized metabolites such as terpenes, alkaloids, polyketides, and saccharides. By predicting and categorizing these clusters, PlantiSMASH facilitates the study of plant metabolic pathways and their potential applications in biotechnology and drug discovery.



Cluster	Record	Type	From	To	Size (kb)	Core domains	CD-HIT Clusters	Most similar known cluster	MIBiG BGC-ID
Cluster 1	CM046694.1	Terpene	107632	278829	171.20	Cellulose_synt, SQHop_cyclase_C, SQHop_cyclase_N, Transferase, p450	5	-	-
Cluster 2	CM046694.1	Saccharide	14995868	15189610	193.74	Transferase, UDPGT_2	3	-	-
Cluster 3	CM046694.1	Saccharide	18885845	19373018	487.17	AMP-binding, Lipoxygenase, UDPGT_2	4	-	-
Cluster 4	CM046694.1	Saccharide	35946695	36183754	237.06	FA_desaturase_2, Transferase, UDPGT_2, adh_short, adh_short_C2	4	-	-
Cluster 5	CM046694.1	Saccharide	36794489	37146127	351.64	AMP-binding, UDPGT_2	3	-	-
Cluster 6	CM046694.1	Saccharide	37631290	37791670	160.38	Glycos_transf_1, Methyltransf_11, p450	3	-	-
Cluster 7	CM046694.1	Polyketide	39006341	39243460	237.12	Cellulose_synt, Chal_sti_synt_C, Chal_sti_synt_N, Methyltransf_2, SE	5	-	-
Cluster 8	CM046694.1	Putative	42900314	43021324	121.01	DIOX_N, p450			
Cluster 9	CM046694.1	Putative	43262127	43469018	206.89	Cellulose_synt, Lipoxygenase			
Cluster 10	CM046694.1	Polyketide	49134190	49173059	38.87	Chal_sti_synt_C, Epimerase			
Cluster 11	CM046694.1	Alkaloid	49836743	49894039	57.30	Bet_v_1, HMGL-like			
Cluster 12	CM046695.1	Putative	14775468	14912534	137.07	DIOX_N, Lycopene_cycl, Peptidase_S10			
Cluster 13	CM046695.1	Putative	37194637	37684130	489.49	2OG-FeII_Oxy, Cellulose_synt, DIOX_N			
Cluster 14	CM046696.1	Terpene	2605567	2784641	179.07	Terpene_synth, p450			
Cluster 15	CM046696.1	Terpene	8623533	8773789	150.26	Methyltransf_11, Terpene_synth, Terpene_synth_C			
Cluster 16	CM046696.1	Terpene	17568340	17603186	34.85	Terpene_synth, Terpene_synth_C, p450			
Cluster 17	CM046697.1	Terpene	3650144	3873636	223.49	SQS_PSY, Terpene_synth, Terpene_synth_C, p450			
Cluster 18	CM046697.1	Saccharide-Terpene	8305351	8679559	374.21	Prenyltrans, SQHop_cyclase_C, SQHop_cyclase_N, UDPGT_2			
Cluster 19	CM046697.1	Saccharide	9302787	9543638	240.85	AMP-binding, UDPGT_2			
Cluster 20	CM046697.1	Putative	30877982	31031722	153.74	Methyltransf_2, Methyltransf_7	4	-	-
Cluster 21	CM046697.1	Putative	34914717	35099212	184.50	COesterase, p450	5	-	-
Cluster 22	CM046698.1	Terpene	3663563	3731186	67.62	Methyltransf_11, Terpene_synth, Terpene_synth_C	3	-	-
Cluster 23	CM046698.1	Putative	4387972	4626925	238.95	AMP-binding, Methyltransf_2, adh_short	6	-	-
Cluster 24	CM046700.1	Putative	736990	820914	83.92	Epimerase, Methyltransf_11, p450	4	-	-
Cluster 25	CM046700.1	Putative	3433319	3476834	43.52	Methyltransf_11, adh_short	4	-	-
Cluster 26	CM046700.1	Putative	13420788	13602689	181.90	Peptidase_S10, p450	4	-	-
Cluster 27	CM046702.1	Saccharide	3455669	3584624	128.96	UDPGT_2, p450	3	-	-
Cluster 28	CM046702.1	Saccharide	6693220	6789892	96.67	Aminotran_1_2, Glycos_transf_2, SE	3	-	-
Cluster 29	CM046703.1	Alkaloid	21014717	21221241	206.52	AMP-binding, Aminotran_1_2, Str_synth, p450	6	-	-
Cluster 30	CM046703.1	Saccharide	24001122	24122029	120.91	UDPGT_2, polyprenyl_synt	3	-	-
Cluster 31	CM046704.1	Saccharide	20094763	20337186	242.42	2OG-FeII_Oxy, DIOX_N, Methyltransf_2, Peptidase_S10, UDPGT_2, p450	9	-	-
Cluster 32	CM046706.1	Terpene	13202418	13368497	166.08	Epimerase, Terpene_synth, Terpene_synth_C, p450	3	-	-
Cluster 33	CM046706.1	Putative	14292454	14727555	435.10	2OG-FeII_Oxy, DIOX_N	7	-	-
Cluster 34	CM046706.1	Terpene	20021884	20215975	194.09	Epimerase, Terpene_synth, Terpene_synth_C	10	-	-

Showing 1 to 34 of 34 entries

Figure 18. PlantiSMASH Page Overview.

a) Cluster Overview

The Cluster Overview page provides a summary of all identified biosynthetic gene clusters in a selected species. Each cluster is classified by type and includes details such as genomic coordinates, cluster size, core domains, and biosynthetic similarity. Users can search for specific loci or biosynthetic types and select a cluster ID to view its detailed gene structure and organization.

b) Detailed Gene Cluster View

Clicking on a cluster opens the Detailed Gene Cluster View, where users can examine the arrangement of genes within the cluster through a graphical representation. Genes are color-coded based on their function, such as biosynthetic, regulatory, and transport-related genes, allowing for easy interpretation of metabolic pathways. Users can also click on individual genes to access functional annotations, aiding in the study of secondary metabolite production in plants.

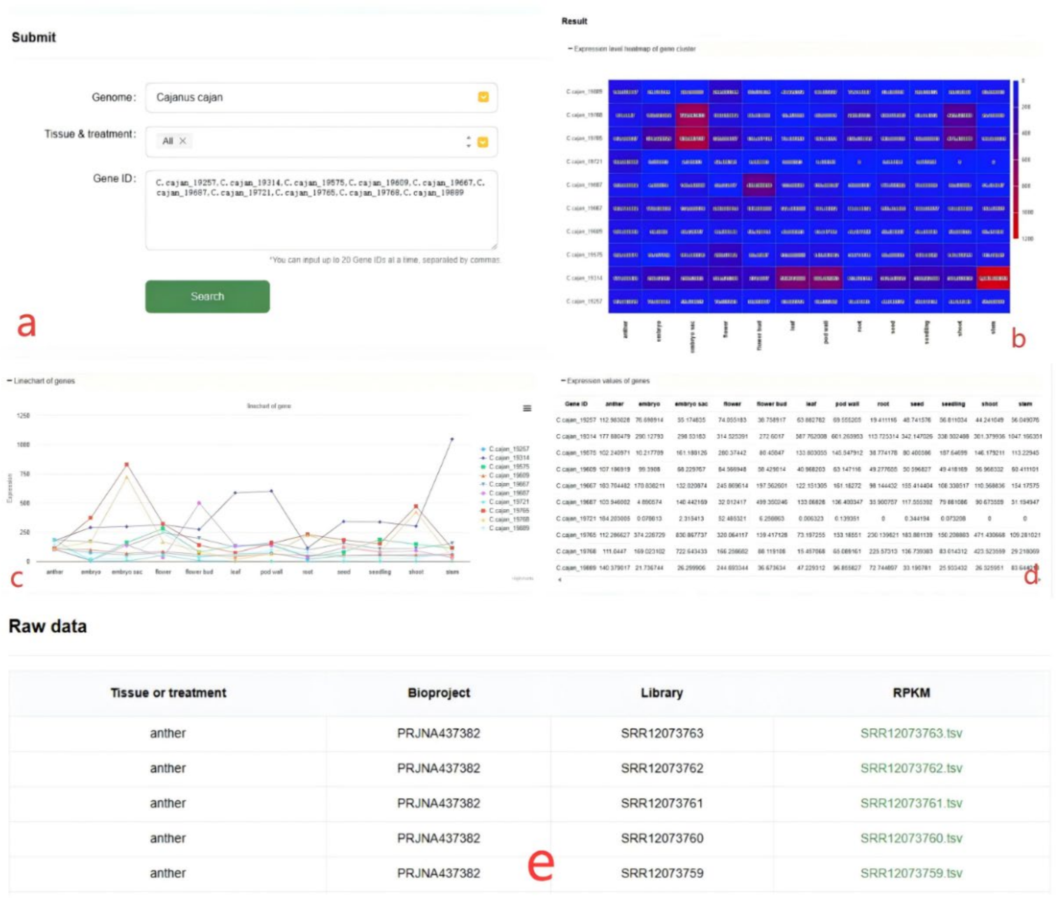
2) Omics Analysis

This section provides tools for analyzing transcriptome, epigenome, and chromatin accessibility data. Users can explore RNA-Seq expression profiles, microRNA annotation, and DNA-binding sites from ChIP-Seq data. Additionally, integrated analysis options allow users to combine RNA-Seq with ChIP-Seq or ATAC-Seq for a deeper understanding of regulatory mechanisms.

A. Transcriptome

The Transcriptomes tool provides a comprehensive analysis of gene expression across various tissues and treatments. Users can input gene IDs to retrieve expression data, visualize patterns through heatmaps and line charts, and access raw transcriptomic data. This tool helps researchers explore gene activity in different biological contexts and conditions.

Cajanus cajan: C.cajan_19257, C.cajan_19314, C.cajan_19575, C.cajan_19609, C.cajan_19667, C.cajan_19687, C.cajan_19721, C.cajan_19765, C.cajan_19768, C.cajan_19889



Genome: Cajanus cajan

Tissue & treatment: All

Gene ID: C.cajan_19257,C.cajan_19314,C.cajan_19575,C.cajan_19609,C.cajan_19667,C.cajan_19687,C.cajan_19721,C.cajan_19765,C.cajan_19768,C.cajan_19889

*You can input up to 20 Gene IDs at a time, separated by commas.

Tissue or treatment	Bioproject	Library	RPKM
anther	PRJNA437382	SRR12073763	SRR12073763.tsv
anther	PRJNA437382	SRR12073762	SRR12073762.tsv
anther	PRJNA437382	SRR12073761	SRR12073761.tsv
anther	PRJNA437382	SRR12073760	SRR12073760.tsv
anther	PRJNA437382	SRR12073759	SRR12073759.tsv

Figure 19. Transcriptomes Page Overview.

a) **Submit Query** (Fig. a)

Users can select a genome, specify tissues or treatments, and input up to 20 gene IDs separated by commas. Clicking "Search" retrieves the expression data for the selected genes, allowing further visualization and analysis.

b) **Expression Heatmap** (Fig. b)

The heatmap visually represents the expression levels of queried genes across different tissues. Higher expression values are displayed in red, while lower values appear in blue. This provides a quick overview of expression patterns in various biological conditions.

c) **Line Chart of Gene Expression** (Fig. c)

This interactive line chart plots expression trends of genes across different tissues. Each line represents a gene, allowing users to compare expression dynamics and identify tissue-specific patterns.

d) **Expression Values Table** (Fig. d)

The table provides a numerical representation of gene expression levels across tissues. Users can examine specific values to understand the quantitative differences in gene expression.

e) **Raw Data Download** (Fig. e)

This section contains Bioproject IDs, libraries, and RPKM expression values for transcriptomic datasets. Users can download raw expression data files (.tsv format) for further analysis and integration into bioinformatics workflows.

f) Software and Parameters

i. Trim Galore (Adapter Trimming)

Software: Trim Galore

Parameters: '-q 20 --stringency 3 --length 20 --gzip'

Purpose: Removes sequencing adapters and low-quality bases from raw RNA-seq reads.

ii. HISAT2 (Read Mapping)

Software: HISAT2

Parameters: '-p 10 -x genome_index -1 -2 '

Purpose: Maps cleaned RNA-seq reads to the reference genome.

iii. StringTie (Transcript Assembly and Quantification)

Software: StringTie

Parameters: '-p 10 -G genome.gtf -e'

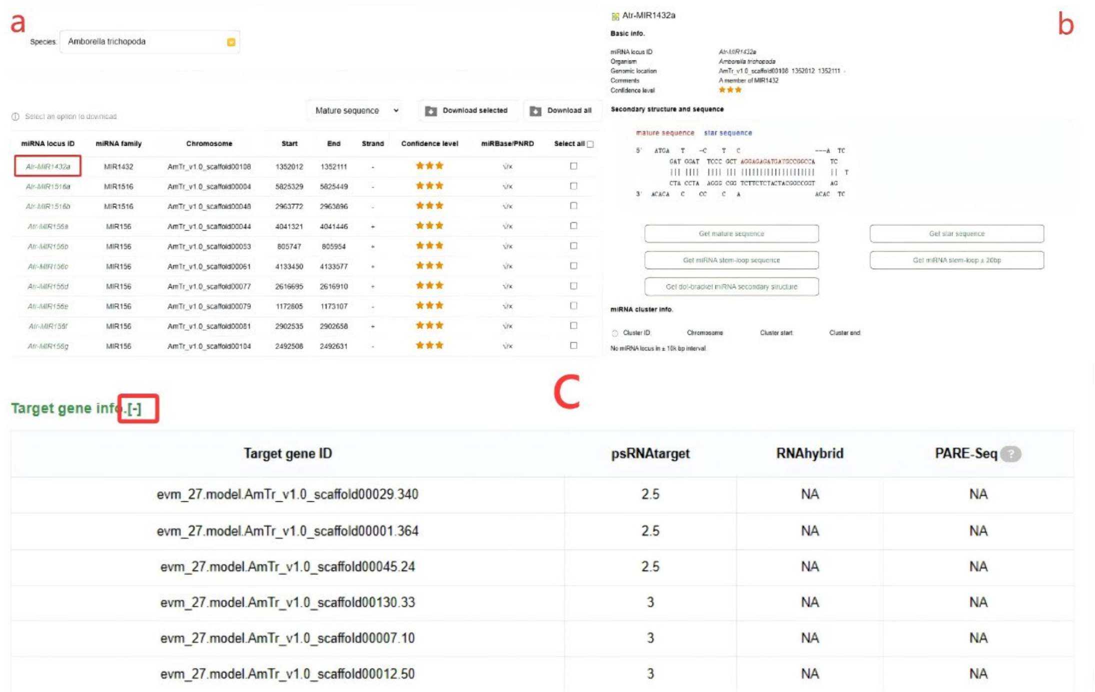
Purpose: Assembles transcripts and estimates expression levels (FPKM, TPM) from mapped reads. Additionally, the prepDE.py script is used with StringTie outputs to extract count matrices for downstream differential expression analysis (e.g., DESeq2).

This pipeline ensures accurate RNA-seq data processing, including adapter trimming, read alignment, transcript assembly, and expression quantification, while facilitating downstream analysis with extracted count matrices.

B. **MicroRNAs** (eg: *Amborella trichopoda*, Atr-MIR1432a)

The MicroRNAs page provides comprehensive information on predicted and annotated miRNAs across different species. Users can search for miRNA loci, explore their genomic positions, and analyze their sequences and secondary structures.

Additionally, the tool includes predicted target genes, helping to understand potential regulatory interactions.



miRNA locus ID	miRNA family	Chromosome	Start	End	Strand	Confidence level	miRBase/PNRD	Select all ☐
Atr-MIR1432a	MIR1432	AmTr_v1.0_scaffold00108	1352012	1352111	-	★★★	√/x	☐
Atr-MIR1516a	MIR1516	AmTr_v1.0_scaffold00004	5825329	5825449	-	★★★	√/x	☐
Atr-MIR1516b	MIR1516	AmTr_v1.0_scaffold00048	2963772	2963896	-	★★★	√/x	☐
Atr-MIR156a	MIR156	AmTr_v1.0_scaffold00044	4041321	4041446	+	★★★	√/x	☐
Atr-MIR156b	MIR156	AmTr_v1.0_scaffold00053	805747	805954	+	★★★	√/x	☐
Atr-MIR156c	MIR156	AmTr_v1.0_scaffold00061	4133450	4133577	+	★★★	√/x	☐
Atr-MIR156d	MIR156	AmTr_v1.0_scaffold00077	2616695	2616910	+	★★★	√/x	☐
Atr-MIR156e	MIR156	AmTr_v1.0_scaffold00079	1172805	1173107	-	★★★	√/x	☐
Atr-MIR156f	MIR156	AmTr_v1.0_scaffold00081	2902535	2902658	+	★★★	√/x	☐
Atr-MIR156g	MIR156	AmTr_v1.0_scaffold00104	2492508	2492631	-	★★★	√/x	☐

Atr-MIR1432a

Basic info.

miRNA locus ID	*Atr-MIR1432a*
Organism	*Amborella trichopoda*
Genomic location	AmTr_v1.0_scaffold00108 1352012 1352111 -
Comments	A member of MIR1432
Confidence level	★★★

Secondary structure and sequence

```
5'   ATGA    T    -C    T   C                     ---A  TC
        GAT GGAT  TCCC GCT AGGAGAGATGATGCCGGCCA     TC
        ||| ||||  |||| ||| ||||||||||||||||||||     ||  T
        CTA CCTA  AGGG CGG TCTTCTCTACTACGGCCGGT     AG
3'  ACACA    C    CC    C   A                     ACAC  TC
```

miRNA cluster info.

○ Cluster ID	Chromosome	Cluster start	Cluster end

No miRNA locus in ± 10k bp interval.

Target gene ID	psRNAtarget	RNAhybrid	PARE-Seq ?
evm_27.model.AmTr_v1.0_scaffold00029.340	2.5	NA	NA
evm_27.model.AmTr_v1.0_scaffold00001.364	2.5	NA	NA
evm_27.model.AmTr_v1.0_scaffold00045.24	2.5	NA	NA
evm_27.model.AmTr_v1.0_scaffold00130.33	3	NA	NA
evm_27.model.AmTr_v1.0_scaffold00007.10	3	NA	NA
evm_27.model.AmTr_v1.0_scaffold00012.50	3	NA	NA

Figure 20. MicroRNAs Page Overview.

a) miRNA Locus Search and Selection

Users can search for miRNAs by selecting a species and browsing the available miRNA loci. The results include detailed annotations such as the miRNA family, chromosome location, start and end positions, strand orientation, and confidence level. Users can also download selected sequences in various formats.

b) miRNA Structural and Sequence Analysis

Clicking on a specific miRNA locus opens a detailed page displaying its sequence, secondary structure, and genomic location. This section allows users to download the mature sequence, star sequence, stem-loop sequence, and dot-bracket notation of the miRNA's predicted secondary structure.

c) Predicted miRNA Target Genes

The target gene prediction section provides a list of putative target genes regulated by the selected miRNA. It includes target gene IDs and scores from multiple prediction algorithms (e.g., psRNAtarget, RNAhybrid, PARE-Seq) to assess binding strength and regulatory potential.

d) Software and Parameters

i. Trim Galore (Adapter Trimming)

Software: Trim Galore

Parameters: '--small_rna --length 18 --max_length 26 --dont_gzip --suppress_warn'

Purpose: Removes adapters and filters raw reads to retain only sRNAs with a length between 18 and 26 nucleotides.

ii. miRDeep-P2 (miRNA Identification)

Software: miRDeep-P2

Parameters: '-g genome.fa -x bowtie_index -L 15 -R 5 -p 1 -q'

Purpose: Identifies novel and known miRNAs by aligning small RNA reads to the reference genome using Bowtie and predicting miRNA precursors with RNAfold.

iii. miRAnno (miRNA Annotation)

Software: miRAnno

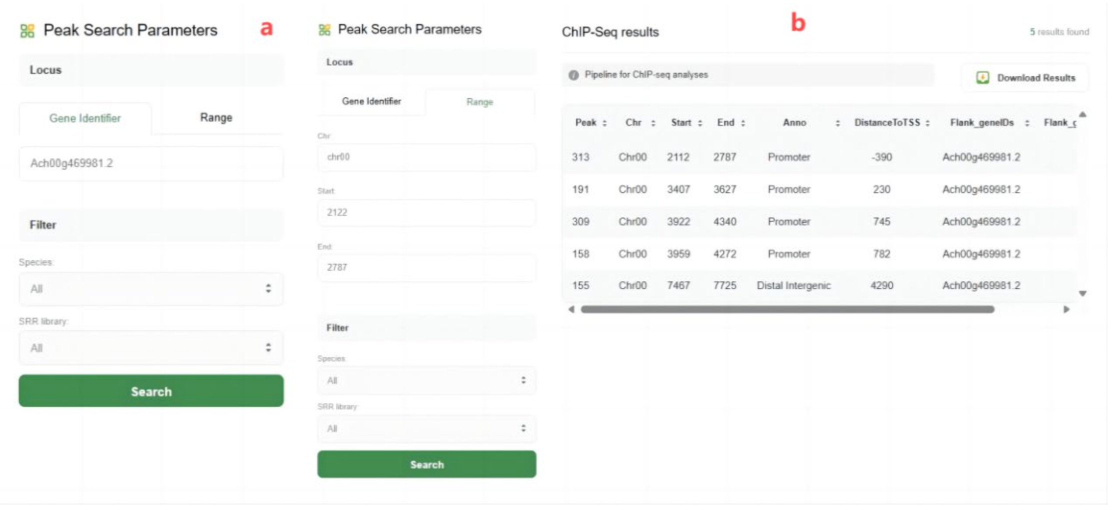
Parameters: '-s -g genome.fa'

Purpose: Annotates miRNAs based on the PmiREN database, linking predicted miRNAs to their known functions and species-specific roles.

This pipeline ensures high-quality miRNA discovery and annotation by integrating adapter trimming, miRNA identification, and functional annotation into a streamlined workflow.

C. **ChIP** (eg: Ach00g469981.2, Chr00, 2122, 2787)

The ChIP-Seq page enables users to analyze chromatin immunoprecipitation sequencing (ChIP-Seq) data by identifying protein-DNA interactions. It provides functionalities to search for peaks associated with specific genes or genomic regions, filter results based on species and sequencing libraries, and obtain detailed peak annotations such as promoter regions, intergenic regions, and distances to transcription start sites (TSS).



Gene Identifier: Ach00g469981.2

Species: All

SRR library: All

Chr: chr00

Start: 2122

End: 2787

Species: All

SRR library: All

Peak	Chr	Start	End	Anno	DistanceToTSS	Flank_geneIDs	Flank_
313	Chr00	2112	2787	Promoter	-390	Ach00g469981.2	
191	Chr00	3407	3627	Promoter	230	Ach00g469981.2	
309	Chr00	3922	4340	Promoter	745	Ach00g469981.2	
158	Chr00	3959	4272	Promoter	782	Ach00g469981.2	
155	Chr00	7467	7725	Distal Intergenic	4290	Ach00g469981.2	

Figure 21. ChIP Page Overview.

a) Peak Search Parameters

Users can search for peaks by either gene identifier or a specific genomic range (chromosome, start, and end positions). Filters are available to refine results based on species and SRR sequencing libraries, allowing users to customize their search for relevant ChIP-Seq data.

b) ChIP-Seq Results

The results table displays detected ChIP-Seq peaks along with their genomic coordinates (chromosome, start, and end positions). Additionally, it provides peak annotations, indicating whether a peak is located in a promoter or distal intergenic region, and the distance to the transcription start site (TSS). Each peak is linked to the flanking gene(s), helping researchers explore potential regulatory interactions.

Users can download the ChIP-Seq results for further analysis by clicking the

"Download Results" button. This feature allows for offline examination and integration with external bioinformatics tools for deeper insights into transcription factor binding and epigenetic modifications.

c) Software and Parameters

i. Trim Galore (Adapter Trimming)

Software: Trim Galore

Parameters: '-q 20 --stringency 3 --length 20 --gzip'

Purpose: Removes adapters and low-quality bases from ChIP-Seq raw reads, preparing data for downstream alignment.

ii. Bowtie2 (Read Mapping)

Software: Bowtie2

Parameters: '-p 10 -x genome_index -1 -2 '

Purpose: Maps cleaned ChIP-Seq reads to the reference genome, producing sorted BAM files for peak calling.

iii. MACS2 (Peak Calling)

Software: MACS2

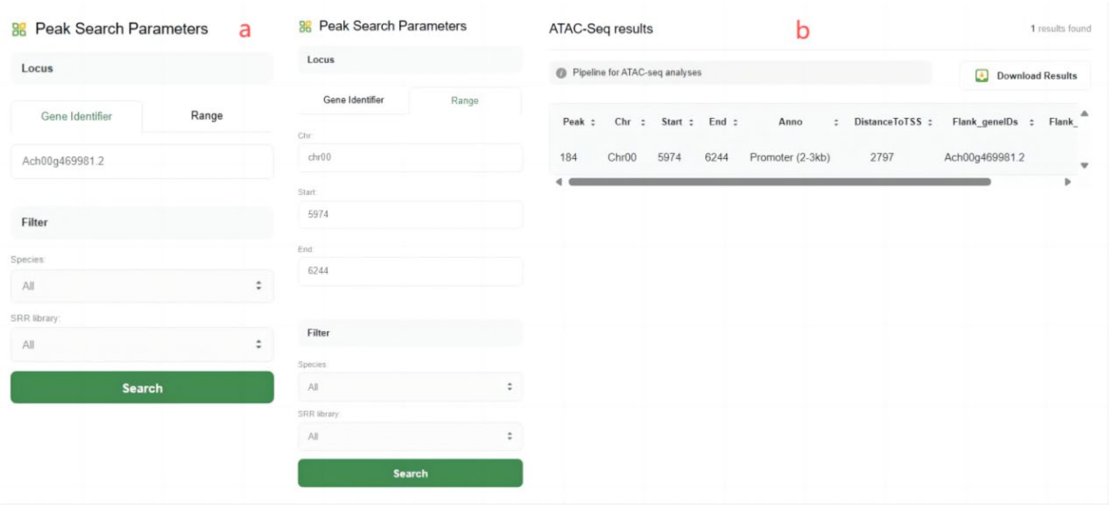
Parameters: '--nomodel -f AUTO --keep-dup 1 -q 0.05 -B -g '

Purpose: Identifies significant enrichment regions (peaks) in the genome that represent potential DNA-protein interaction sites.

This pipeline ensures high-quality ChIP-Seq analysis by integrating adapter trimming, read alignment, and peak calling to identify DNA-protein interaction sites.

D. **ATAC** (eg: Ach00g469981.2, Chr00, 5974, 6244)

The ATAC-Seq page allows users to analyze chromatin accessibility data, providing insights into open chromatin regions that may regulate gene expression. This tool helps researchers identify regulatory elements such as enhancers and promoters, offering options to search peaks linked to specific genes or genomic loci. The results include peak annotations, transcription start site (TSS) distances, and flanking genes, with an option to export data for further analysis.



Gene Identifier: Ach00g469981.2

Species: All

SRR library: All

Chr: chr00

Start: 5974

End: 6244

Species: All

SRR library: All

Peak	Chr	Start	End	Anno	DistanceToTSS	Flank_geneIDs	Flank_
184	Chr00	5974	6244	Promoter (2-3kb)	2797	Ach00g469981.2	

Figure 22. ATAC Page Overview.

a) Peak Search Parameters

Users can search for chromatin accessibility peaks using either a gene identifier or a specific genomic range by defining chromosome coordinates. Filters are available to refine searches by species and sequencing library, enabling targeted data retrieval for specific experimental conditions.

b) ATAC-Seq Results

The results table displays detected ATAC-Seq peaks, including chromosomal location, start and end positions, and annotations indicating whether a peak is located within a promoter, enhancer, or other functional region. Additionally, the distance to the nearest transcription start site (TSS) is provided, helping users assess regulatory potential.

Users can download ATAC-Seq results in tabular format for further computational analysis. This feature facilitates downstream investigations into chromatin accessibility patterns, transcriptional regulation, and functional genomics research.

c) Software and Parameters

i. Trim Galore (Adapter Trimming)

Software: Trim Galore

Parameters: '-q 20 --stringency 3 --length 20 --gzip'

Purpose: Removes adapters and low-quality bases from ATAC-Seq raw reads, ensuring clean data for downstream alignment.

ii. Bowtie2 (Read Mapping)

Software: Bowtie2

Parameters: '-p 10 -x genome_index -1 -2 '

Purpose: Maps cleaned ATAC-Seq reads to the reference genome, producing sorted BAM files for peak calling.

iii. MACS2 (Peak Calling)

Software: MACS2

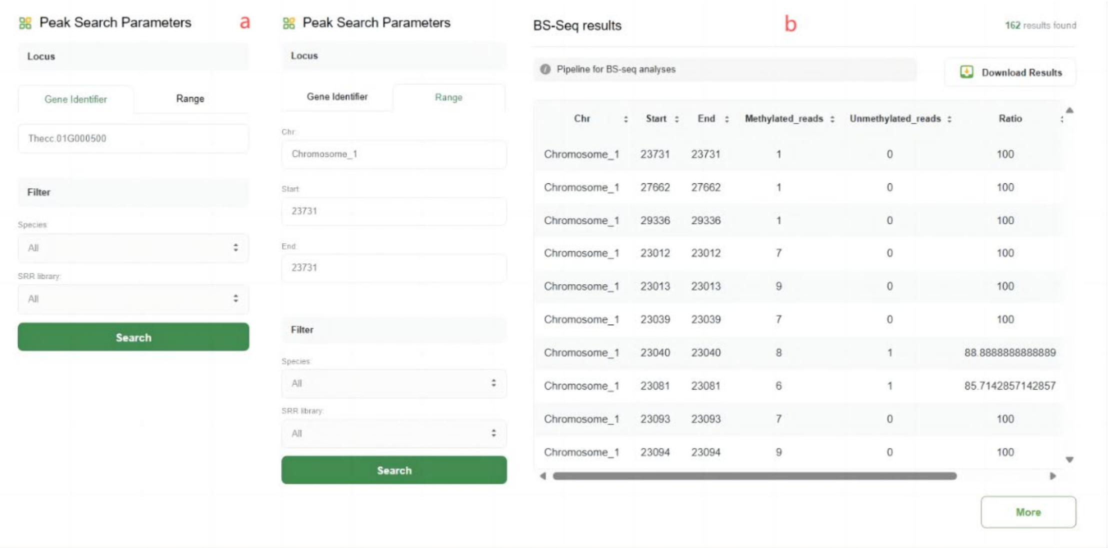
Parameters: '--nomodel -f AUTO --keep-dup 1 -q 0.05 -B -g '

Purpose: Identifies significant enrichment regions (peaks) in the genome, representing open chromatin sites and accessible regulatory elements.

This pipeline ensures reliable ATAC-Seq data analysis by integrating adapter trimming, read alignment, and peak calling to study chromatin accessibility and open chromatin regions.

E. Methylation (eg: Thecc.01G000300, Chr12, 14592, 19924)

The BS-Seq (Bisulfite Sequencing) page provides users with DNA methylation data, allowing them to explore methylation patterns across different genomic regions. This tool helps researchers understand epigenetic modifications by distinguishing between methylated and unmethylated reads, providing insights into gene regulation, development, and environmental responses. Users can search for methylation data by gene ID or specific genomic loci, and the results include detailed methylation ratios along with filtering and export options.



Thecc.01G000500

Species: All

SRR library: All

Chr: Chromosome_1

Start: 23731

End: 23731

Species: All

SRR library: All

162 results found

Pipeline for BS-seq analyses

Chr	Start	End	Methylated_reads	Unmethylated_reads	Ratio
Chromosome_1	23731	23731	1	0	100
Chromosome_1	27662	27662	1	0	100
Chromosome_1	29336	29336	1	0	100
Chromosome_1	23012	23012	7	0	100
Chromosome_1	23013	23013	9	0	100
Chromosome_1	23039	23039	7	0	100
Chromosome_1	23040	23040	8	1	88.8888888888889
Chromosome_1	23081	23081	6	1	85.7142857142857
Chromosome_1	23093	23093	7	0	100
Chromosome_1	23094	23094	9	0	100

Figure 23. Methylation Page Overview.

a) Peak Search Parameters

Users can search for methylation peaks using either a gene identifier or a specific chromosomal range by specifying coordinates. Filters are available to narrow searches by species and sequencing libraries, enabling focused retrieval of methylation data relevant to specific biological conditions.

b) BS-Seq Results

The results table displays key methylation statistics, including chromosomal positions, number of methylated and unmethylated reads, and methylation ratios. This data helps users assess the methylation status of specific genomic sites, aiding in the study of gene expression regulation and epigenetic modifications.

Users can export the methylation data in a tabular format for downstream computational analyses. This allows for further investigation of DNA methylation dynamics, identification of differentially methylated regions (DMRs), and integration with transcriptomic and genomic datasets.

c) Software and Parameters

i. Trim Galore (Adapter Trimming)

Software: Trim Galore

Parameters: `-q 20 --stringency 3 --length 20 --gzip`

Purpose: Removes adapters and filters low-quality bases from BS-Seq raw reads to prepare clean data for downstream analysis.

ii. Bismark (Alignment and Methylation Analysis)

Software: Bismark

Parameters:

Step 1: 'bismark_genome_preparation' to index the reference genome.

Step 2: '--bowtie2 -N 0 -L 20 -p ' for read alignment.

Step 3: 'deduplicate_bismark' to remove PCR duplicates.

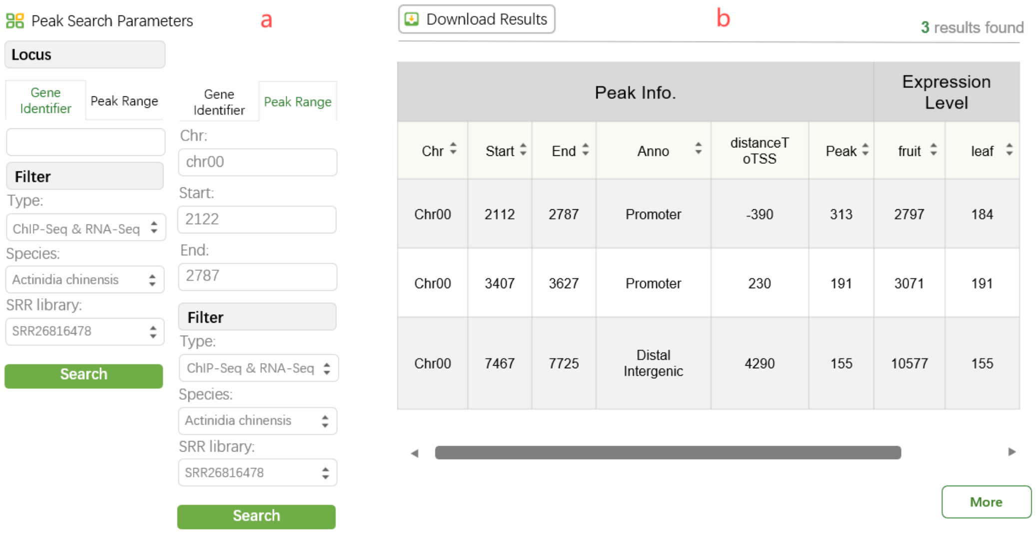
Step 4: 'bismark_methylation_extractor --bedGraph --counts --comprehensive' to extract methylation calls.

Purpose: Aligns reads to the bisulfite-converted reference genome, removes duplicates, and extracts comprehensive methylation data for further analysis.

This pipeline ensures accurate DNA methylation analysis, integrating adapter trimming, alignment, deduplication, and methylation call extraction to provide high-quality methylation profiles.

F. **RNA-seq & ChIP or ATAC-seq** (eg: ChIP-seq & RNA-seq, Actinidia chinensis, SRR26816478, chr00, 2122, 2787)

The RNA-Seq & ChIP or ATAC-Seq page integrates transcriptomic (RNA-Seq) and epigenomic (ChIP-Seq or ATAC-Seq) data, allowing users to examine gene expression in conjunction with regulatory element interactions. This functionality helps researchers analyze how chromatin accessibility or transcription factor binding correlates with gene expression, providing insights into gene regulation mechanisms. Users can search for peaks related to specific genes, filter by species and sequencing libraries, and view expression levels across different tissues.



Type: ChIP-Seq & RNA-Seq
Species: Actinidia chinensis
SRR library: SRR26816478

Chr: chr00
Start: 2122
End: 2787
Type: ChIP-Seq & RNA-Seq
Species: Actinidia chinensis
SRR library: SRR26816478

3 results found

Peak Info.						Expression Level	
Chr	Start	End	Anno	distanceToTSS	Peak	fruit	leaf
Chr00	2112	2787	Promoter	-390	313	2797	184
Chr00	3407	3627	Promoter	230	191	3071	191
Chr00	7467	7725	Distal Intergenic	4290	155	10577	155

Figure 24. RNA-seq & ChIP or ATAC-seq Page Overview.

a) Peak Search Parameters

Users can input either a gene identifier or a specific chromosomal range to identify peaks of interest. Filters are available to refine searches by data type (ChIP-Seq & RNA-Seq), species, and sequencing library, enabling precise retrieval of relevant regulatory interactions.

b) RNA-seq & ChIP or ATAC-seq Results

The results display detected regulatory peaks, including chromosomal positions, annotation type (e.g., promoter or intergenic region), distance to transcription start sites (TSS), and peak intensity values. This information helps identify potential regulatory elements influencing gene expression.

For each identified peak, gene expression levels in different tissues (e.g., fruit and leaf) are presented. This enables correlation between regulatory peak intensity and gene expression, assisting in understanding transcriptional regulation under different biological conditions.

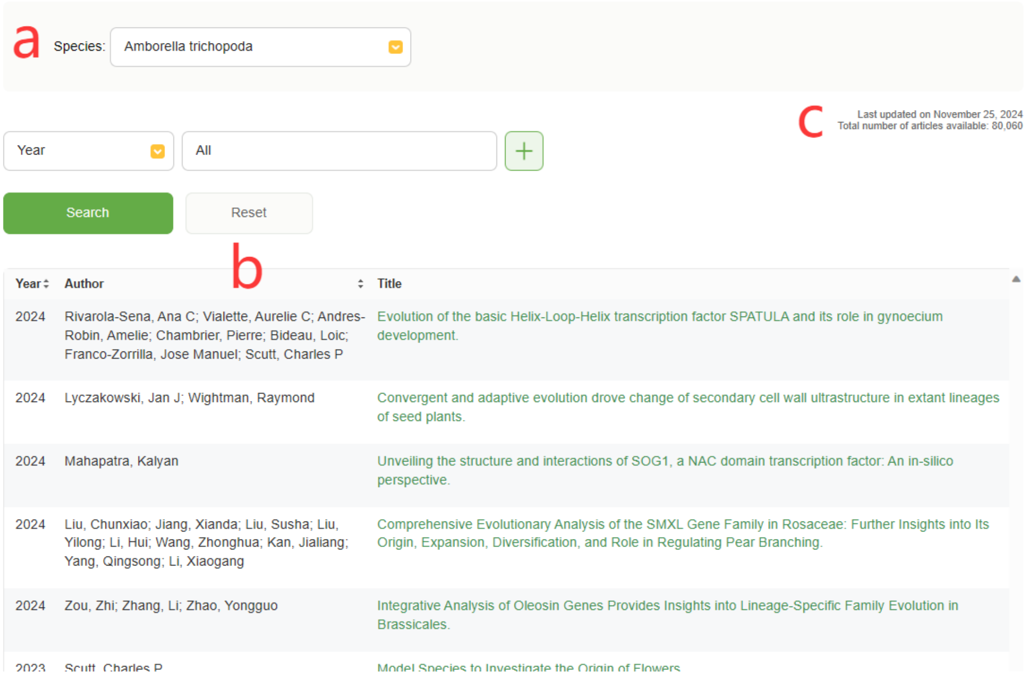
Users can download the results for further analysis, facilitating comparative studies between transcription factor binding, chromatin accessibility, and gene expression across different species and conditions.

3) Others

This section includes resources for literature search, primer design, and phylogenetic tree construction. Users can access the latest research publications, design PCR primers, and generate phylogenetic trees for specific plant families.

A. Literature (eg: *Amborella trichopoda*)

The Literature page enables users to search for research articles related to a specific species. Users can filter literature based on various criteria, such as year, author, title, journal, keyword, and abstract. The platform provides access to a vast collection of published studies, ensuring that users can stay updated with the latest scientific research.



Species: Amborella trichopoda

Last updated on November 25, 2024
Total number of articles available: 80,060

Year: All

Year	Author	Title
2024	Rivarola-Sena, Ana C; Vialette, Aurelie C; Andres-Robin, Amelie; Chambrier, Pierre; Bideau, Loic; Franco-Zorrilla, Jose Manuel; Scutt, Charles P	Evolution of the basic Helix-Loop-Helix transcription factor SPATULA and its role in gynoecium development.
2024	Lyczakowski, Jan J; Wightman, Raymond	Convergent and adaptive evolution drove change of secondary cell wall ultrastructure in extant lineages of seed plants.
2024	Mahapatra, Kalyan	Unveiling the structure and interactions of SOG1, a NAC domain transcription factor: An in-silico perspective.
2024	Liu, Chunxiao; Jiang, Xianda; Liu, Susha; Liu, Yilong; Li, Hui; Wang, Zhonghua; Kan, Jialiang; Yang, Qingsong; Li, Xiaogang	Comprehensive Evolutionary Analysis of the SMXL Gene Family in Rosaceae: Further Insights into Its Origin, Expansion, Diversification, and Role in Regulating Pear Branching.
2024	Zou, Zhi; Zhang, Li; Zhao, Yongguo	Integrative Analysis of Oleosin Genes Provides Insights into Lineage-Specific Family Evolution in Brassicales.
2023	Scutt, Charles P	Model Species to Investigate the Origin of Flowers

Figure 24. Literature Page Overview.

a) Search and Filter

Users can select a species and refine their search by specifying additional filters such as publication year, author name, or keywords. The interface allows multiple search criteria to be applied simultaneously, enabling precise literature retrieval.

b) Literature Results Display

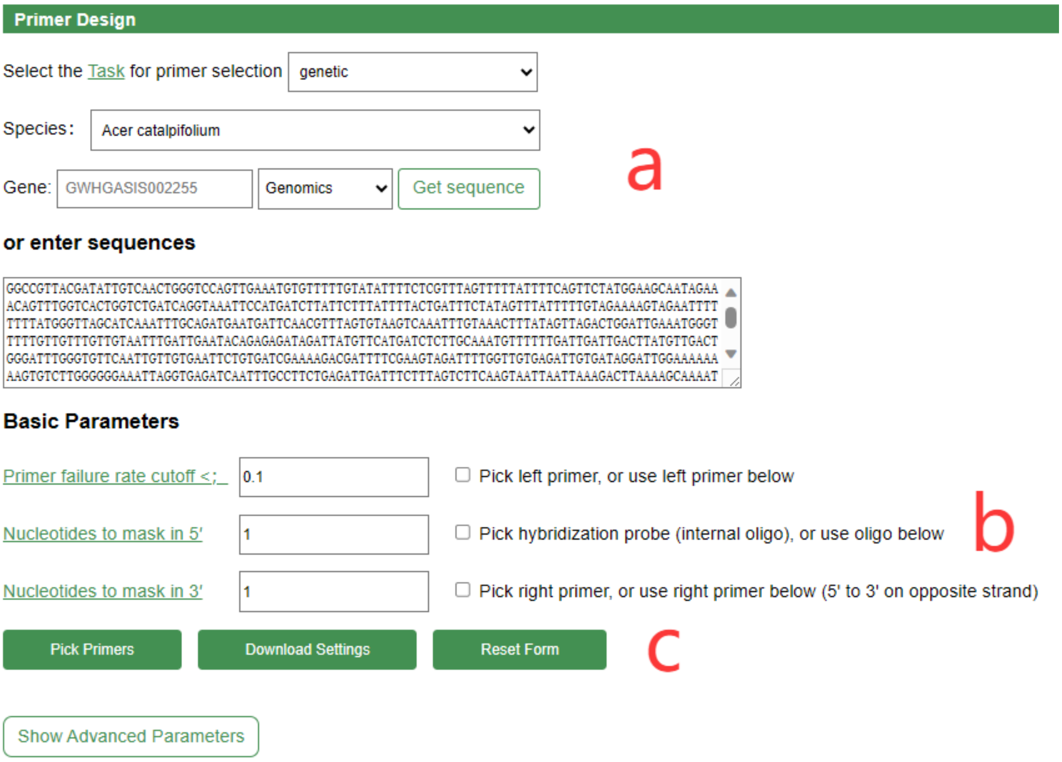
The search results are displayed in a tabular format, showing key details such as the publication year, authors, and article titles. Clicking on a title redirects users to the original research article for further reading. The database is regularly updated to ensure access to the most recent publications.

c) Data Overview and Management

At the top right corner, users can view the last update date and the total number of available articles. A reset option is also available to clear filters and start a new search, ensuring an efficient and seamless user experience.

B. **Primer Design** (eg: *Acer catalpifolium*, GWHGASIS002255)

The Primer Design tool allows users to design primers for genetic analysis efficiently. By selecting specific genes or entering sequences manually, users can customize primer parameters to suit their experimental needs. The tool integrates various options for sequence retrieval, masking nucleotides, and optimizing primer selection, ensuring high-quality primer design for PCR and other molecular applications.



Select the Task for primer selection: genetic

Species: Acer catalpifolium

Gene: GWHGASIS002255 Genomics

or enter sequences

```
GGCCGTTACGATATTGTCAACTGGGTCCAGTTGAAATGTGTTTTTGTATATTTTCTCGTTTAGTTTTTATTTTCAGTTCTATGGAAGCAATAGAA
ACAGTTTGGTCACTGGTCTGATCAGGTAAATTCCATGATCTTATTCTTTATTTTACTGATTTCTATAGTTTATTTTTGTAGAAAAGTAGAATTTT
TTTTATGGGTTAGCATCAAATTTGCAGATGAATGATTCAACGTTTAGTGTAAGTCAAATTTGTAAACTTTATAGTTAGACTGGATTGAAATGGGT
TTTTGTTGTTTGTTGTAATTTGATTGAATACAGAGAGATAGATTATGTTCATGATCTCTTGCAAATGTTTTTTGATTGATTGACTTATGTTGACT
GGGATTTGGGTGTTCAATTGTTGTGAATTCTGTGATCGAAAAGACGATTTTCGAAGTAGATTTTGGTTGTGAGATTGTGATAGGATTGGAAAAAA
AAGTGTCTTGGGGGGAAATTAGGTGAGATCAATTTGCCTTCTGAGATTGATTTCTTTAGTCTTCAAGTAATTAATTAAAGACTTAAAAGCAAAAT
```

Basic Parameters

Primer failure rate cutoff <;_ 0.1

Nucleotides to mask in 5' 1

Nucleotides to mask in 3' 1

☐ Pick left primer, or use left primer below

☐ Pick hybridization probe (internal oligo), or use oligo below

☐ Pick right primer, or use right primer below (5' to 3' on opposite strand)

Figure 25. Primer Design Page Overview.

a) Gene Selection and Sequence Input

Users can either select a species and gene ID to retrieve genomic, mRNA, CDS, or protein sequences or manually enter a custom sequence. This flexibility allows researchers to design primers for a wide range of target sequences.

b) Basic Parameter Configuration

The interface provides key settings such as primer failure rate cutoff and nucleotide masking at the 5’ and 3’ ends. Additional options include selecting left or right primers and hybridization probes, ensuring tailored primer design for different experimental setups.

c) Primer Design and Download Options

Once the parameters are set, users can generate primers using the Pick Primers button.

The tool also allows users to download settings for future use or reset the form to start a new design process. The Show Advanced Parameters button provides access to additional customization settings for expert users.

C. **Phylogenetic Tree** (eg: Aquifoliaceae)

The Phylogenetic Tree tool allows users to visualize evolutionary relationships among different species or genes using a hierarchical tree structure. This interactive tool supports various visualization modes, filtering options, and branch selection, providing researchers with a powerful platform to analyze phylogenetic relationships.

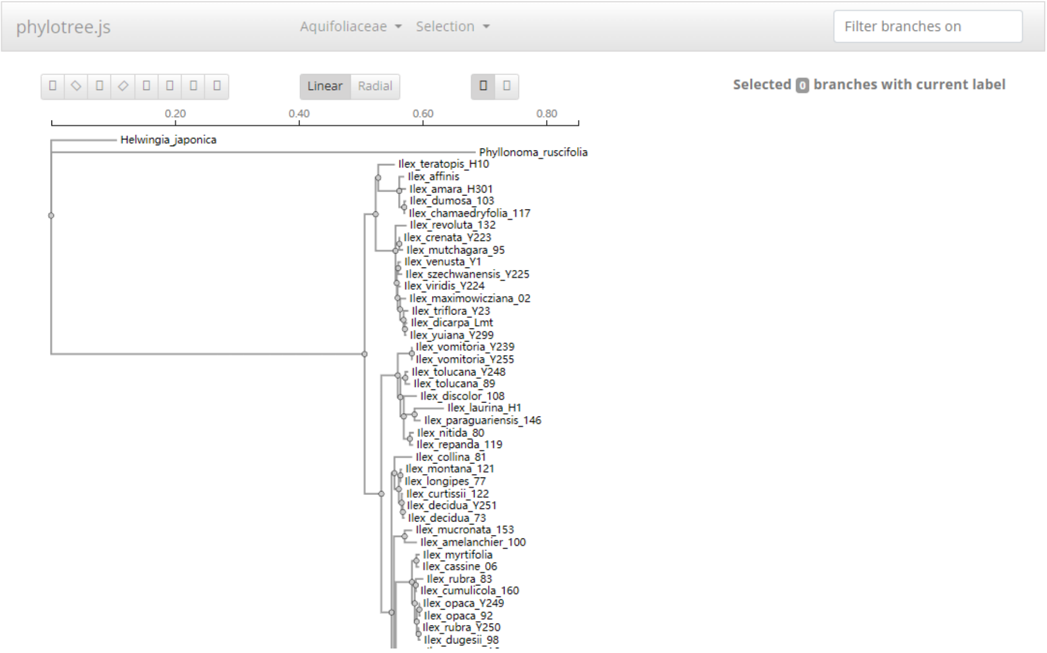


Figure 26. Phylogenetic Tree Page Overview.

a) **Tree Visualization and Navigation**

Users can choose between linear and radial tree representations to best fit their analytical needs. The interface provides zooming and panning functionalities, making it easy to explore complex phylogenetic structures.

b) **Species and Branch Selection**

The dropdown menus allow users to filter and select specific taxa or genes of interest.

Users can highlight branches, compare evolutionary distances, and focus on specific clades to extract meaningful insights.

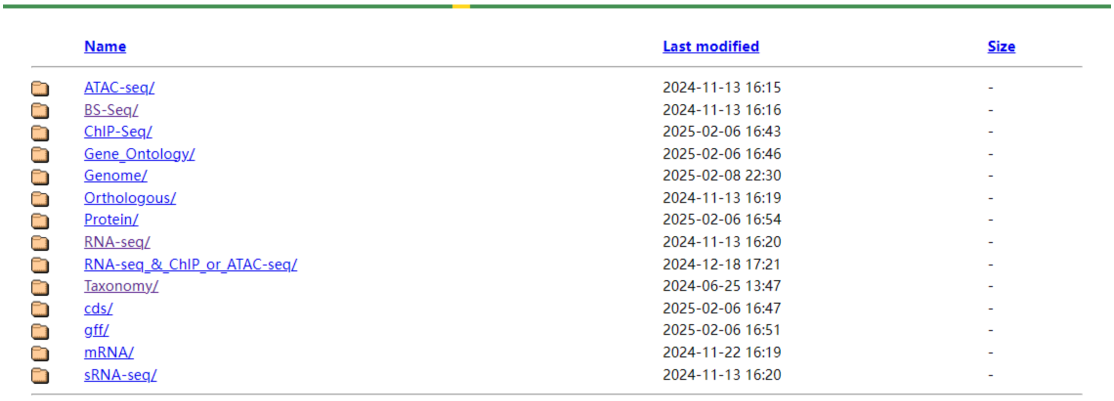
c) Filtering and Data Exploration

By using the Filter branches on option, users can dynamically refine the displayed tree to highlight relevant evolutionary relationships. This feature enables researchers to focus on specific groups, compare divergence points, and analyze the hierarchical clustering of species or genes.

5. Data

The DATA page serves as a central repository for all raw and processed data generated by the database, systematically categorized by species or file type. Users can access various datasets, including ATAC-seq, ChIP-seq, BS-Seq, RNA-seq, Gene Ontology annotations, orthologous gene data, genome sequences, and structural annotations. This organization ensures efficient retrieval of relevant genomic, transcriptomic, and functional data.

Each file entry includes metadata, such as the last modified date, ensuring users have access to the latest datasets. Researchers can navigate through these directories to download essential files for further analysis, supporting a wide range of genomics and bioinformatics studies.



Name	Last modified	Size
ATAC-seq/	2024-11-13 16:15	-
BS-Seq/	2024-11-13 16:16	-
ChIP-Seq/	2025-02-06 16:43	-
Gene_Ontology/	2025-02-06 16:46	-
Genome/	2025-02-08 22:30	-
Orthologous/	2024-11-13 16:19	-
Protein/	2025-02-06 16:54	-
RNA-seq/	2024-11-13 16:20	-
RNA-seq_&_ChIP_or_ATAC-seq/	2024-12-18 17:21	-
Taxonomy/	2024-06-25 13:47	-
cds/	2025-02-06 16:47	-
gff/	2025-02-06 16:51	-
mRNA/	2024-11-22 16:19	-
sRNA-seq/	2024-11-13 16:20	-

Figure 27. Data Page Overview.

6. Help

The Help page provides users with essential resources, including documentation, an about section, and contact information. Users can report issues, provide feedback, or seek assistance via email (Contact@woodyplant.com) or by visiting the listed physical address in Nanjing, China.

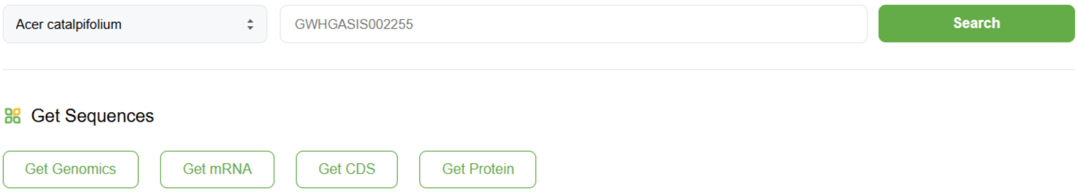
7. Q & A

1) How to find information about a specific gene for a particular species?

To retrieve detailed information about a specific gene in a woody plant species, users can follow these steps:

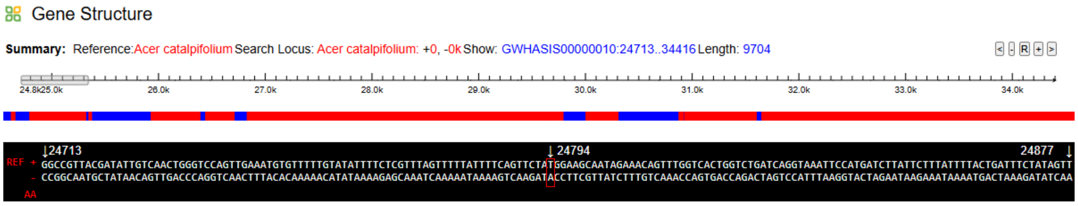
A. Step 1: Use the Gene Retrieval Tool

- Tools > Gene Retrieval (https://woodyplant.com/generetrieval).
- Select the desired species from the dropdown list.
- Enter the gene ID and click Search.
- This will return key sequence data, including:
    - Genomic sequence
    - mRNA sequence
    - CDS (coding sequence)
    - Protein sequence


Acer catalpifolium
GWHGASIS002255

B. Step 2: Explore Gene Structure

- After retrieving the gene, users can analyze its structural features.
- The Gene Structure section displays exon-intron organization and base-pair positioning, aiding in understanding gene architecture.

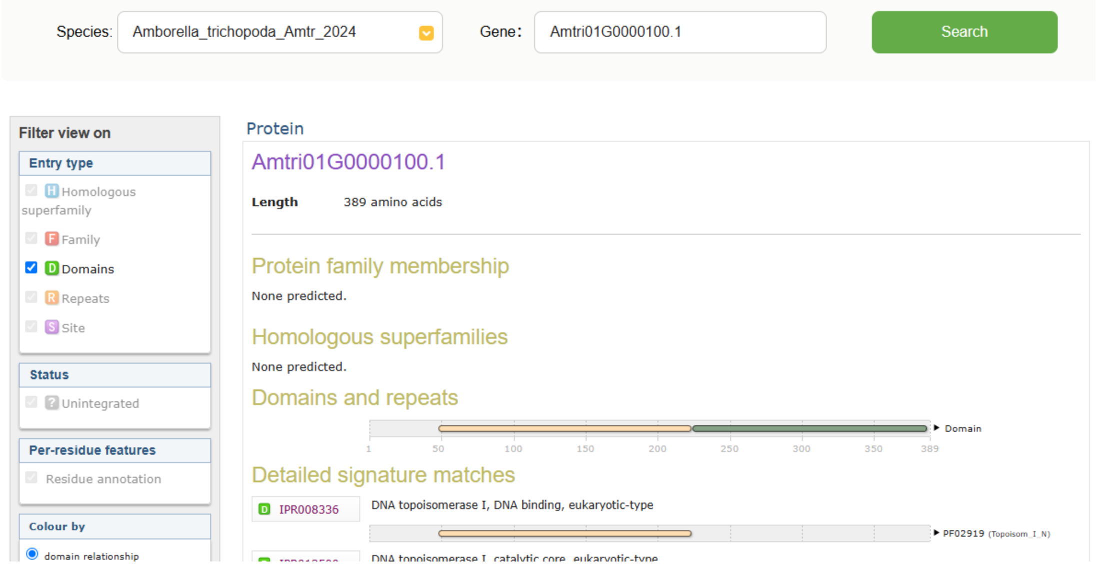

Gene Structure

Summary: Reference:Acer catalpifolium Search Locus: Acer catalpifolium: +0, -0k Show: GWHASIS00000010:24713..34416 Length: 9704

```
 ↓24713                                                                         ↓ 24794                                                                                   24877 ↓
REF + GGCCGTTACGATATTGTCAACTGGGTCCAGTTGAAATGTGTTTTTGTATATTTTCTCGTTTAGTTTTTATTTTCAGTTCTATGGAAGCAATAGAAACAGTTTGGTCACTGGTCTGATCAGGTAAATTCCATGATCTTATTCTTTATTTTACTGATTTCTATAGTT
    - CCGGCAATGCTATAACAGTTGACCCAGGTCAACTTTACACAAAAACATATAAAAGAGCAAATCAAAAATAAAAGTCAAGATACCTTCGTTATCTTTGTCAAACCAGTGACCAGACTAGTCCATTTAAGGTACTAGAATAAGAAATAAAATGACTAAAGATATCAA
 AA
```

C. Step 3: Investigate Gene Function via Annotation

- Tools > Gene Annotation (https://woodyplant.com/toolannotation).
- Input the gene ID to obtain InterProScan-based functional predictions.
- This will provide insights into:
    - Protein domains and families
    - GO term predictions (biological processes, molecular functions, and cellular components)
    - Functional annotations from homologous proteins


Species: Amborella_trichopoda_Amtr_2024

Gene: Amtri01G0000100.1

Protein

Amtri01G0000100.1

Length 389 amino acids

Protein family membership

None predicted.

Homologous superfamilies

None predicted.

Domains and repeats

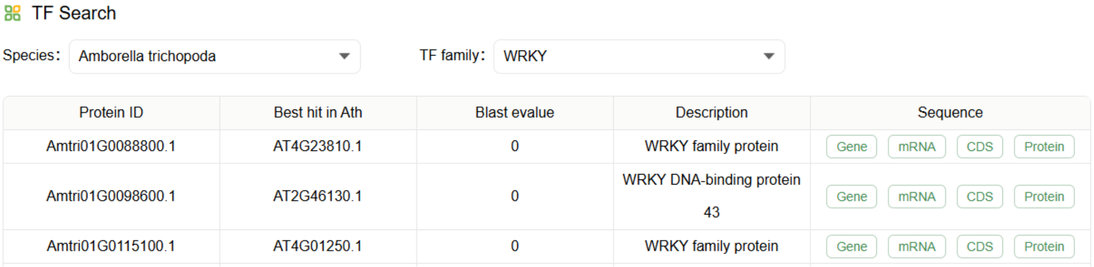
Detailed signature matches

D IPR008336 DNA topoisomerase I, DNA binding, eukaryotic-type

DNA topoisomerase I, catalytic core, eukaryotic-type

D. Step 4: Identify Related Transcription Factors

- Tools > Transcriptional Factor (https://woodyplant.com/tooltransfactor).
- This will display associated TF families, regulatory potential, and corresponding orthologs.


TF Search

Species: Amborella trichopoda

TF family: WRKY

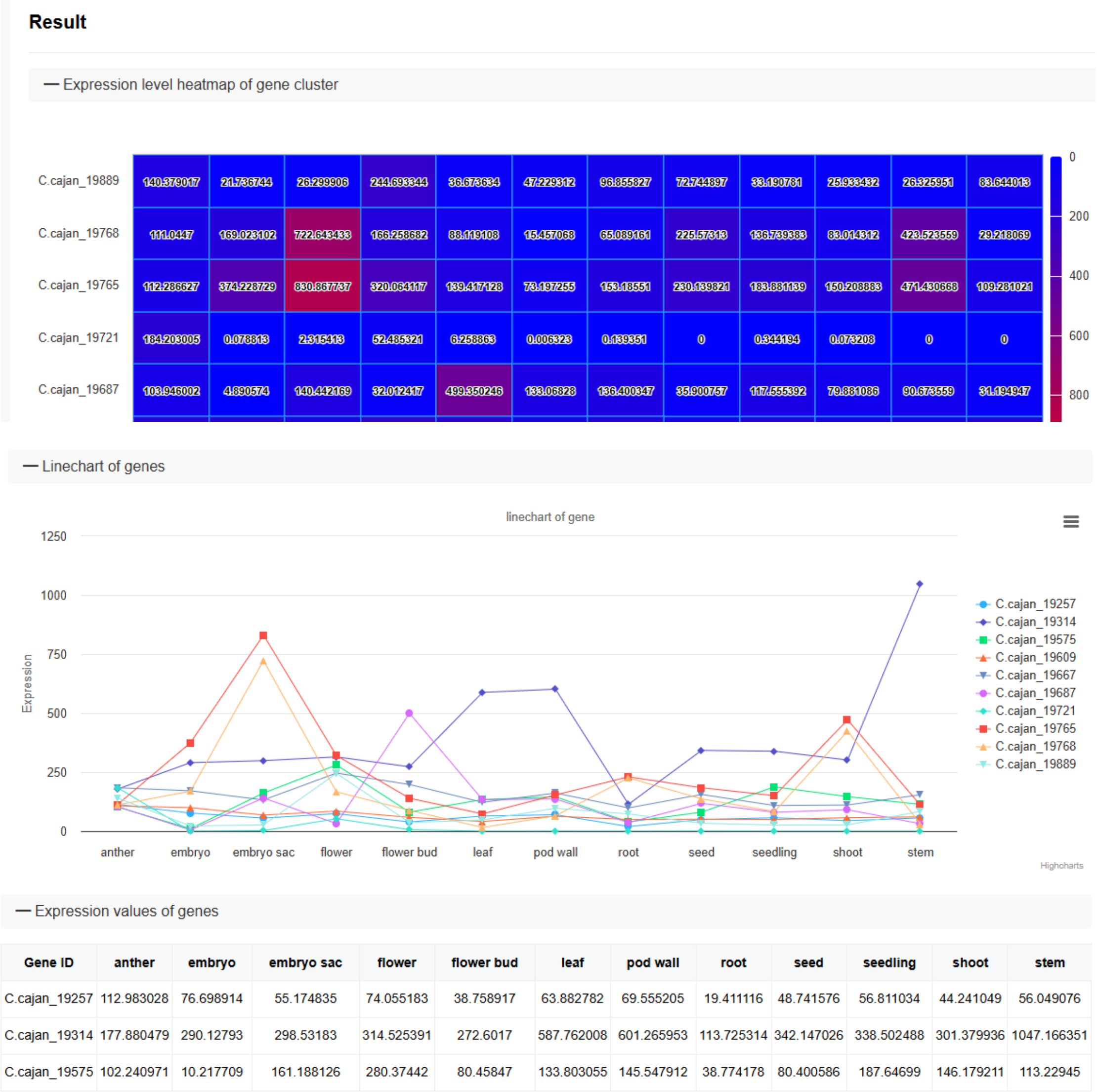
Protein ID	Best hit in Ath	Blast evalue	Description	Sequence
Amtri01G0088800.1	AT4G23810.1	0	WRKY family protein	Gene mRNA CDS Protein
Amtri01G0098600.1	AT2G46130.1	0	WRKY DNA-binding protein 43	Gene mRNA CDS Protein
Amtri01G0115100.1	AT4G01250.1	0	WRKY family protein	Gene mRNA CDS Protein

E. Step 5: Explore Expression Patterns

- Tools > Transcriptome (https://woodyplant.com/tooltreatment)
- Select the species and input the gene ID to generate:
    - Expression heatmaps for different plant tissues.
    - Line charts showing dynamic expression trends.
    - Expression value tables for statistical analysis.

F. Additional Research Ideas

- Compare the expression levels of the gene across different tissues (e.g., leaves, roots, flowers) to infer potential functional roles.
- Analyze differential expression under various treatments (e.g., drought, salinity) to study its stress response.
- Use ChIP-Seq and ATAC-Seq data to check whether this gene is a target of transcriptional regulation.



Result

— Expression level heatmap of gene cluster

C.cajan_19889	140.379017	21.736744	26.299906	244.693344	36.673634	47.229312	96.855827	72.744897	33.190781	25.933432	26.325951	83.644013
C.cajan_19768	111.0447	169.023102	722.643433	166.258682	88.119108	15.457068	65.089161	225.57313	136.739383	83.014312	423.523559	29.218069
C.cajan_19765	112.286627	374.228729	830.867737	320.064117	139.417128	73.197255	153.18551	230.139821	183.881139	150.208883	471.430668	109.281021
C.cajan_19721	184.203005	0.078813	2.315413	52.485321	6.258863	0.006323	0.139351	0	0.344194	0.073208	0	0
C.cajan_19687	103.946002	4.890574	140.442169	32.012417	499.350246	133.06828	136.400347	35.900757	117.555392	79.881086	90.673559	31.194947

— Linechart of genes

— Expression values of genes

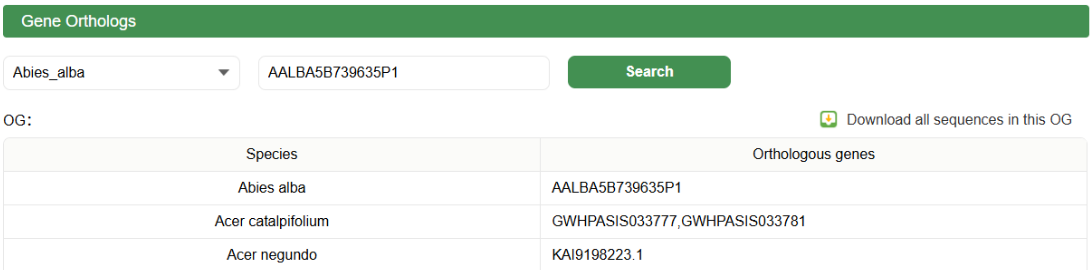
Gene ID	anther	embryo	embryo sac	flower	flower bud	leaf	pod wall	root	seed	seedling	shoot	stem
C.cajan_19257	112.983028	76.698914	55.174835	74.055183	38.758917	63.882782	69.555205	19.411116	48.741576	56.811034	44.241049	56.049076
C.cajan_19314	177.880479	290.12793	298.53183	314.525391	272.6017	587.762008	601.265953	113.725314	342.147026	338.502488	301.379936	1047.166351
C.cajan_19575	102.240971	10.217709	161.188126	280.37442	80.45847	133.803055	145.547912	38.774178	80.400586	187.64699	146.179211	113.22945

2) How to access data on a gene family across all woody plant species?

To explore the presence and evolution of a specific gene family across multiple woody plant species, users can employ the following strategies:

A. Step 1: Search for Orthologs Across Species

- Tools > Gene Orthologs (https://woodyplant.com/geneorthologs).
- Enter a reference gene ID or select a species.
- This will retrieve orthologous genes across 96 woody plant species, allowing for:
  - Phylogenetic comparisons of gene conservation and divergence.
  - Cross-species functional annotation transfer.



Gene Orthologs

Abies_alba

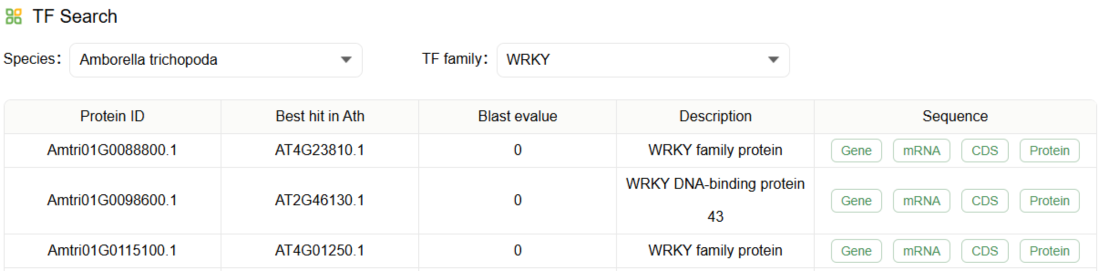
AALBA5B739635P1

OG:

Species	Orthologous genes
Abies alba	AALBA5B739635P1
Acer catalpifolium	GWHPASIS033777,GWHPASIS033781
Acer negundo	KAI9198223.1

B. Step 2: Explore Transcription Factor Families

- Tools > Transcriptional Factor.
- Select the species and input the TF family name (e.g., WRKY, MYB).
- This will return all TFs belonging to the selected family, along with:
  - Best hits from Arabidopsis, aiding functional inference.
  - BLAST e-values, providing confidence scores for homology.



TF Search

Species: Amborella trichopoda

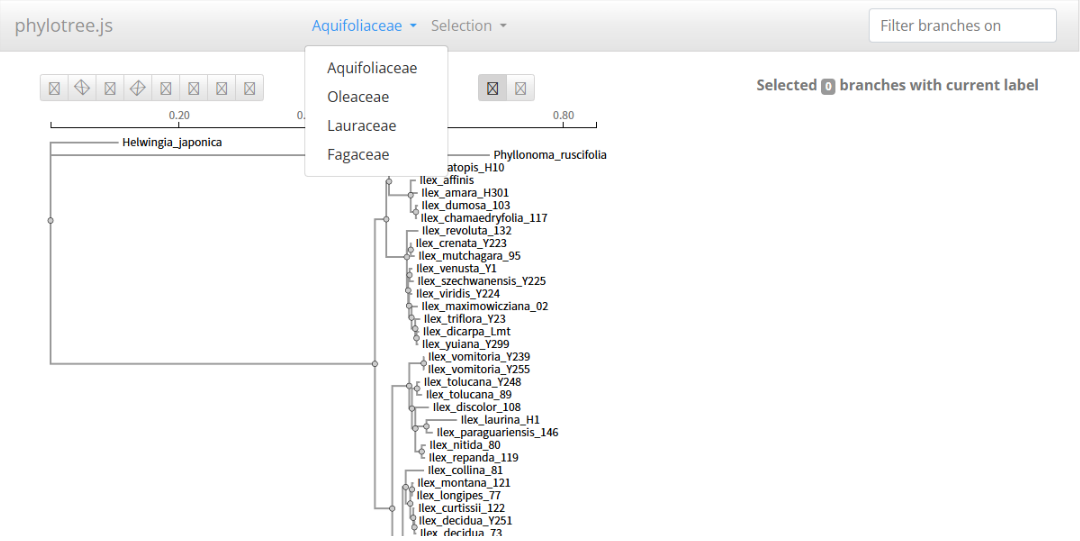
TF family: WRKY

Protein ID	Best hit in Ath	Blast evalue	Description	Sequence
Amtri01G0088800.1	AT4G23810.1	0	WRKY family protein	Gene mRNA CDS Protein
Amtri01G0098600.1	AT2G46130.1	0	WRKY DNA-binding protein 43	Gene mRNA CDS Protein
Amtri01G0115100.1	AT4G01250.1	0	WRKY family protein	Gene mRNA CDS Protein

C. Step 3: Conduct Phylogenetic Analysis

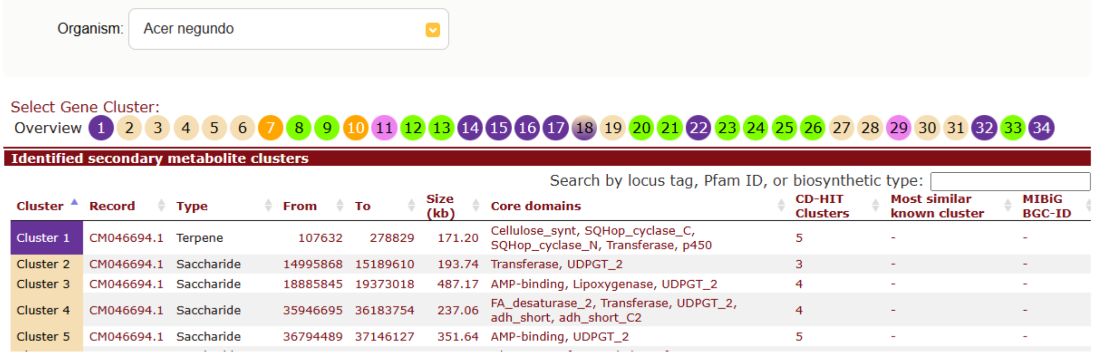
- Tools > Phylogenetic Tree (https://woodyplant.com/phylotreejs).

- Upload gene sequences from multiple species and generate:
    - Linear or radial phylogenetic trees.
    - Branch selection and filtering tools to highlight subgroups.



D. Step 4: Investigate Secondary Metabolite Biosynthesis

- Tools > PlantiSMASH (https://woodyplant.com/toolpathway).
- Select a species and analyze gene clusters related to biosynthetic pathways.
- This can reveal conserved pathways across species, providing insights into metabolic evolution.



Organism: Acer negundo

Identified secondary metabolite clusters

Search by locus tag, Pfam ID, or biosynthetic type:

Cluster	Record	Type	From	To	Size (kb)	Core domains	CD-HIT Clusters	Most similar known cluster	MIBiG BGC-ID
Cluster 1	CM046694.1	Terpene	107632	278829	171.20	Cellulose_synt, SQHop_cyclase_C, SQHop_cyclase_N, Transferase, p450	5	-	-
Cluster 2	CM046694.1	Saccharide	14995868	15189610	193.74	Transferase, UDPGT_2	3	-	-
Cluster 3	CM046694.1	Saccharide	18885845	19373018	487.17	AMP-binding, Lipoxygenase, UDPGT_2	4	-	-
Cluster 4	CM046694.1	Saccharide	35946695	36183754	237.06	FA_desaturase_2, Transferase, UDPGT_2, adh_short, adh_short_C2	4	-	-
Cluster 5	CM046694.1	Saccharide	36794489	37146127	351.64	AMP-binding, UDPGT_2	5	-	-

E. Additional Research Ideas:

- Study the expansion and contraction of gene families across evolutionary lineages.
- Compare expression patterns in different species to identify species-specific regulatory roles.
- Investigate whether specific orthologs are under positive selection by analyzing sequence divergence.

3) How to quickly download data from WP-MOD for customized analysis?

For large-scale bioinformatics analyses, users may need to download all results in WP-MOD for their following study. The DATA page provides access to all source files.

- Visit the DATA page (https://woodyplant.com/ftp?path=data) to obtain the following data in the respective directory.
    - ATAC-seq/
    - BS-Seq/
    - ChIP-Seq/
    - Gene_Ontology/
    - Genome/
    - Orthologous/
    - Protein/
    - RNA-seq/
    - RNA-seq_&_ChIP_or_ATAC-seq/
    - Taxonomy/
    - cds/
    - gff/
    - mRNA/
    - sRNA-seq/