BMC Genomics

DATABASE

Open Access

ASAPv2: an improved platform for exploring gene function in *Angelica sinensis*

Wei Zhu[1†], Silan Wu[1†], Xingxing Zhao[1], Jiaotong Yang[1*] and Qiaoqiao Xiao[1*]

Abstract

Angelica sinensis (Danggui) is a long-valued herb in traditional Chinese medicine for treating gynecological and other health issues. Modern research has identified numerous bioactive compounds in *A. sinensis* roots, but the genetic and biosynthetic pathways underlying these constituents remain partially understood. The recent availability of high-quality *A. sinensis* genomes and extensive omic data has opened new opportunities for gene function analysis. Integrating these resources can significantly aid in uncovering gene functions in *A. sinensis's* secondary metabolic pathways. We created ASAPv2, an improved platform for analyzing *A. sinensis* gene functions. It combines two reference genomes, 97 transcriptome datasets from diverse tissues and treatment conditions, and metabolomic profiles from existing studies. All genes in the database are annotated with functional information by sequence similarity to public databases and domain analyses. Key gene families have been identified including 5,594 transcription factors, 2,665 protein kinases, 1,746 Carbohydrate-Active enZymes, 1,727 transporters, and 2,770 ubiquitin-related enzymes. Furthermore, ASAPv2 identified 1,055,573 positive co-expression relationships and 86,486 protein–protein interaction networks predicted by orthology to *Arabidopsis*. The platform, built on a LAMP architecture, offers a web interface with browsing, search, and visualization tools. Users can query genes, perform BLAST searches, view genomic regions in JBrowse, examine gene expression heatmaps, explore network relationships, and retrieve sequences. In addition, by mining transcriptome data through the data platform, we identified a candidate gene, SOC1, involved in flowering regulation. Multiple lines of functional evidence obtained via the platform support this finding, indicating that the platform can play an auxiliary role in acquiring and mining gene function information. We hope ASAPv2 (accessible at www.gzybioinformatics.cn/ASAPv2) will enhance *A. sinensis* research and enable new discoveries.

Keywords *Angelica sinensis*, Multi-omics integration, Gene function analysis, Secondary metabolism, Bioinformatics platform

[†]Wei Zhu and Silan Wu contributed equally to this work.

*Correspondence:
Jiaotong Yang
y_jiaotong@163.com
Qiaoqiao Xiao
xqqiao2021@163.com
[1]Guizhou University of Traditional Chinese Medicine, Guiyang, Guizhou 550025, China

Background

Angelica sinensis (*A. sinensis*) (Oliv.) Diels, also called Danggui or "female ginseng", is a key traditional Chinese medicinal herb used for almost 2000 years [1, 2]. Its dried roots are valued for enriching blood and regulating the menstrual cycle [1]. Recent studies have revealed the hepatoprotective, neuroprotective, anti-inflammatory, and anti-fibrotic effects of *A. sinensis*, linked to its diverse bioactive components. *A. sinensis* polysaccharide (ASP) forms nanoparticles that target the liver, improving drug distribution and protecting against acute alcoholic liver injury by reducing oxidative stress [3]. ASP also inhibits epithelial–mesenchymal transition and pulmonary fibrosis [4]. However, the biosynthetic pathways and specific genes for these compounds are not yet fully understood.

Recent genomic and sequencing advancements have boosted *A. sinensis* research. In 2022, the first chromosome-level genome assembly for *A. sinensis* was published, shedding light on its genome structure and coumarin biosynthesis gene evolution [5]. A second high-quality genome assembly from another *A. sinensis* variety was recently published, enhancing the genomic landscape and enabling comparative gene content and sequence variation analyses [6]. Alongside genomic resources, extensive transcriptome data for *A. sinensis* across different conditions and stages have been produced. Zhu et al. used RNA-seq and Weighted Gene Co-expression Network Analysis (WGCNA) to identify WRKY6, PAT1, and SCL13 as crucial transcription factors in drought response, finding that the M2 variety showed greater drought tolerance [7]. Zhu et al. used transcriptomic and metabolomic analyses on two *A. sinensis* cultivars to identify key gene-metabolite networks, especially those related to photosynthesis, hormone signaling, and phenolic/flavonoid metabolites, that differentiate early bolting and flowering plants from normal ones [8]. Liu et al. demonstrated that increased quinic acid in *A. sinensis* accelerates flowering by influencing gibberellin and jasmonic acid pathways, using untargeted metabolomics and functional assays [9]. These studies highlight the effectiveness of using transcriptomic and metabolomic methods to connect genes with key traits and metabolites in *A. sinensis*. However, the data are often scattered across supplementary files or public repositories, complicating collective analysis and reuse by researchers.

As genomic and multi-omic data become more accessible, there's a rising demand for an integrated platform for *A. sinensis*. Successful model databases in plant research highlight the benefits of consolidating species-specific data and tools. The Arabidopsis Information Resource (TAIR) is a key center for Arabidopsis thaliana genomics and genetics [10]. To provide a resource for *A. sinensis*, we previously developed the *A. sinensis* Analysis Platform (ASAP) for gene functional analysis [11]. The original ASAP integrated *A. sinensis*'s reference genome and transcriptomes, offering tools for sequence search, functional annotation, and network analysis via a web interface, enabling gene-centric analyses without bioinformatics expertise. However, new data types and omics data necessitated an update. This paper introduces ASAPv2, an upgraded multi-omics database for *A. sinensis*, featuring expanded datasets (new genome assembly, additional transcriptomes, and new proteomic and metabolomic data) and enhanced functions like gene family classification and improved user interaction. We detail ASAPv2's construction, content, and technical implementation, showcasing its utility with examples. ASAPv2 is expected to significantly advance research on *A. sinensis* gene function and medicinal compound biosynthesis.

Construction and content

Data collection and integration

ASAPv2 brings together diverse datasets covering the genome, transcriptome, and metabolome of *A. sinensis*. The genomic component includes two high-quality reference genome assemblies. One genome (2022 version) corresponds to the previously published chromosome-level assembly of *A. sinensis* [5] and was downloaded from the ASAP download page [11]. The second genome (2023 version) offers an independent assembly with an approximate size of 2.16 Gb, accessible from the CNGB database (CNA0050842) [6]. We gathered 97 transcriptome samples of *A. sinensis*, which include 79 public RNA-seq datasets from the NCBI Sequence Read Archive and 18 datasets from the Genome Sequence Archive (GSA) database of National Genomics Data Center (NGDC) in China (Table S1). These transcriptome samples encompass a range of tissues (e.g., root, leaf, flower), developmental stages, and treatment conditions (such as exposures to UV-B), providing broad coverage of gene expression profiles. In addition, we incorporated 3 metabolomic datasets from published analyses of *A. sinensis* [6, 12, 13]. Key metabolites and their measured concentrations or presence in *A. sinensis* are recorded. By integrating genomics, transcriptome, and metabolomic data, ASAPv2 enables researchers to connect genotype to phenotype more directly.

Data processing and functional annotation

We first performed a synteny analysis of the two genome versions using MCScanX software (no version displayed) to determine the correspondence between chromosomes (Table S2). Additionally, a bidirectional BLAST comparison was conducted using DIAMOND blastp (v2.1.11.165) with an e-value threshold of < 1e-10 to obtain the best matches as the correspondence between genes. We then conducted a comprehensive alignment of the predicted proteins of *A. sinensis* against several

protein databases, including the National Center for Biotechnology Information (NCBI) non-redundant (NR) protein database (2024-02-07 downloaded), Universal Protein Resource (Uniprot, release-2024_05) (including TrEMBL and SwissProt) [14], and The Arabidopsis Information Resource (TAIR10) protein database [15], utilizing the DIAMOND BLAST tool (v2.1.11.165) [16]. The most significant matches and their associated descriptions from these databases were aggregated to serve as gene annotations. Furthermore, gene ontology (GO) and protein Pfam domain annotations were performed for each gene using InterProScan (v1.7.0_91) [17]. Enzyme Commission (EC) numbers and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway assignments were obtained via the GhostKOALA server (https://www.kegg.jp/ghostkoala/) [18] and subsequently cross-referenced with the KEGG database [15] for pathway mapping (see Table 1). The annotated GO terms and KEGG pathways provide insights into the potential biological processes, molecular functions, and metabolic pathways in which each gene may be involved.

Identifying gene family membership is crucial for understanding the evolution of gene functions and pinpointing candidates involved in specific processes like transcription regulation or signaling. We used specialized tools to classify *A. sinensis* genes into families as follows:

Transcription factors and kinases

We employed Plant Transcription factor and Protein Kinase Identifier and Classifier (iTAK, v1.7, default parameters) [19] to identify and classify transcription factor and protein kinase families in the *A. sinensis*, discovering 5,594 transcription factors (2,580 from 2022 version and 3,014 from 2023 version) and 2,665 protein kinases (1,269 from 2022 version and 1,396 from 2023 version).

Orthologous groups and known gene families

Using OrthoFinder (v 2.5.2) [20], we analyzed orthologous groups between *A. sinensis* and *A. thaliana*, identifying genes linked to known plant gene families like transporter proteins and Carbohydrate-Active enZymes

Table 1 Detailed statistics for the two genome assemblies integrated in ASAPv2

Data Category	2022 version	2023 version	Description
Basic Genome Information			
Genome Size	2.37 Gb	2.16 Gb	Total genome assembly size
Predicted Genes	43,202	41,040	Total protein-coding genes
Scaffold N50	193.87 Mb	198.27 Mb	Assembly contiguity measure
BUSCO Completeness	98.6% (viridiplantae), 97.1% (eudicots)	>99% completeness	Benchmarking Universal Single-Copy Orthologs
Functional Annotation			
NR Annotation	38,420 Genes	39,391 Genes	NCBI non-redundant database hits
SwissProt Annotation	25,641 Genes	27,131 Genes	SwissProt database matches
TAIR Annotation	28,941 Genes	31,121 Genes	Arabidopsis TAIR database matches
TrEMBL Annotation	36,415 Genes	38,334 Genes	UniProt TrEMBL database matches
GO Annotation	26,441 Genes	16,960 Genes	Gene Ontology term assignments
KEGG Annotation	8,468 Genes	11,125 Genes	KEGG pathway annotations
Pfam Domains	26,441 Genes	28,474 Genes	Protein domain annotations
Gene Family Classification			
Transcription Factors	2,580 Genes	3,014 Genes	iTAK-predicted TF families
Protein Kinases	1,269 Genes	1,396 Genes	iTAK-predicted kinase families
Cytochrome P450	204 Genes	220 Genes	P450 enzyme families
CAZy Enzymes	662 Genes	1,084 Genes	Carbohydrate-active enzymes
Transporters	640 Genes	1,087 Genes	Membrane transport proteins
Ubiquitin System	1,262 Genes	1,508 Genes	Ubiquitin-proteasome system proteins
Sequence Resources			
CDS Sequences	43,202	41,039	Coding sequence records
Protein Sequences	43,202	41,039	Predicted protein sequences
BUSCO	90.3% viridiplantae	96.4% viridiplantae	Benchmarking Universal Single-Copy Orthologs
Transcript Sequences	43,202	41,039	Full-length transcript sequences
Promoter Sequences	43,202	41,039	Upstream regulatory regions (1kb, 2kb, 3kb)
Network Analysis			
Co-expression (Positive)	431,657	623,916	Positive gene co-expression relationships
PPI Interactions	27,094	59,392	Protein-protein interaction pairs

(CAZy) families. We found 1,727 transporter (TP) protein genes (640 from 2022 version and 1,087 from 2023 version) and 1,746 CAZy enzyme genes (662 from 2022 version and 1,084 from 2023 version).

Ubiquitin-proteasome system proteins

Using the integrated annotations for Ubiquitin and Ubiquitin-like Conjugation Database (iUUCD 2.0) Hidden Markov Model (HMM) profiles [21], we classified components of the ubiquitination pathways, including E1, E2, E3 enzymes, and deubiquitinases. By applying hidden Markov models to *A. sinensis* proteins, we identified genes related to the ubiquitin and ubiquitin-like modifier pathways, totaling 2,770 genes—1,262 from the 2022 version and 1,508 from the 2023 version.

Other enzyme families

We also annotated specific families, such as cytochrome P450s (CYPs) by combining keyword search with KEGG pathway annotation. In total, 424 cytochrome P450 genes were identified, including 204 from the 2022 version and 220 from the 2023 version.

Network construction

The RNA-seq data were first subjected to quality control using fastp (v0.20.1) with the command: fastp -i input_1.fastq -I input_2.fastq -o input_1.fastp.fastq -O input_2.fastp.fastq. Subsequently, the clean reads were aligned to the reference genome using hisat2 (v2.2.1) with the following command: hisat2 -x genome.hisat2 –1 input_1.fastp.fastq –2 input_2.fastp.fastq -S input.sam. The resulting SAM files were then converted into BAM format using samtools (v1.18). Gene expression quantification was performed with stringtie (v2.2.1) to calculate TPM values, using the command: stringtie -p 20 -e -G input.gff -o output.gff -A output.table input.sam. Afterward, the raw expression matrix was constructed at the gene level. Batch effects were corrected using the ComBat function of the sva package (v3.56.0) in R, and the data were subsequently normalized by Z-score scaling using the R scale function to obtain the expression matrix used for co-expression network construction.

We calculated the Pearson correlation coefficient (PCC) for all gene pairs using WGCNA packages (v1.73) in R based on expression matrix, identifying genes whose expression levels rise and fall together across conditions. To refine the network and focus on biologically meaningful correlations, we applied a ranking-based filtering method (MR, mutual rank) and determined an optimal PCC cutoff by evaluating the network against known gene sets (using ROC curves as described in Supplementary Materials). The resulting co-expression network connects genes that are likely co-regulated or functionally related. In total, we identified 1,055,573 positive co-expression relationships, including 431,657 from the 2022 version and 623,916 from the 2023 version.

The *A. sinensis* protein–protein interaction network was predicted by mapping its genes to *A. thaliana* orthologs using OrthoFinder. Known *Arabidopsis* (Protein-Protein Interactions) PPIs were then assigned to *A. sinensis*, resulting in 86,486 predicted PPI pairs, with 27,094 from the 2022 version and 59,392 from the 2023 version.

Database implementation and interface

ASAP v2 is a dynamic web database built on a LAMP stack (Linux, Apache, MySQL, PHP). It stores processed data such as genome sequence, gene/protein sequences, annotations, gene family classifications, expression values, network, and metabolomic, as well as several analysis tools for gene functional analysis (Fig. 1).

Home and search

A user-friendly homepage introduces the database and provides quick search functionality. Users can search by gene ID, gene name/keyword, or functional terms to find relevant genes.

Genomes

This section presents information related to genome, including details of the two available genome versions.

Network

The network module provides access to co-expression and protein-protein interaction networks for both genome versions. The search results include dynamic visualizations of these networks. Additionally, the interface allows users to compare gene expression across different transcriptome conditions, with significantly upregulated genes highlighted in red and downregulated genes in blue.

Gene family

This section showcases six different gene families, including Carbohydrate-Active Enzymes Family, Ubiquitin Family, Transporters, Protein Kinases, CYP450, and Transcription Factor Families.

Pathway function

This section presents KEGG annotation results and also includes manually curated genes encoding key enzymes related to the biosynthesis of active compounds in *A. sinensis*.

Analysis tools

A BLAST search tool allows users to perform sequence similarity searches against the *A. sinensis* nucleotide sequence/protein sequence. A Gene Set Enrichment

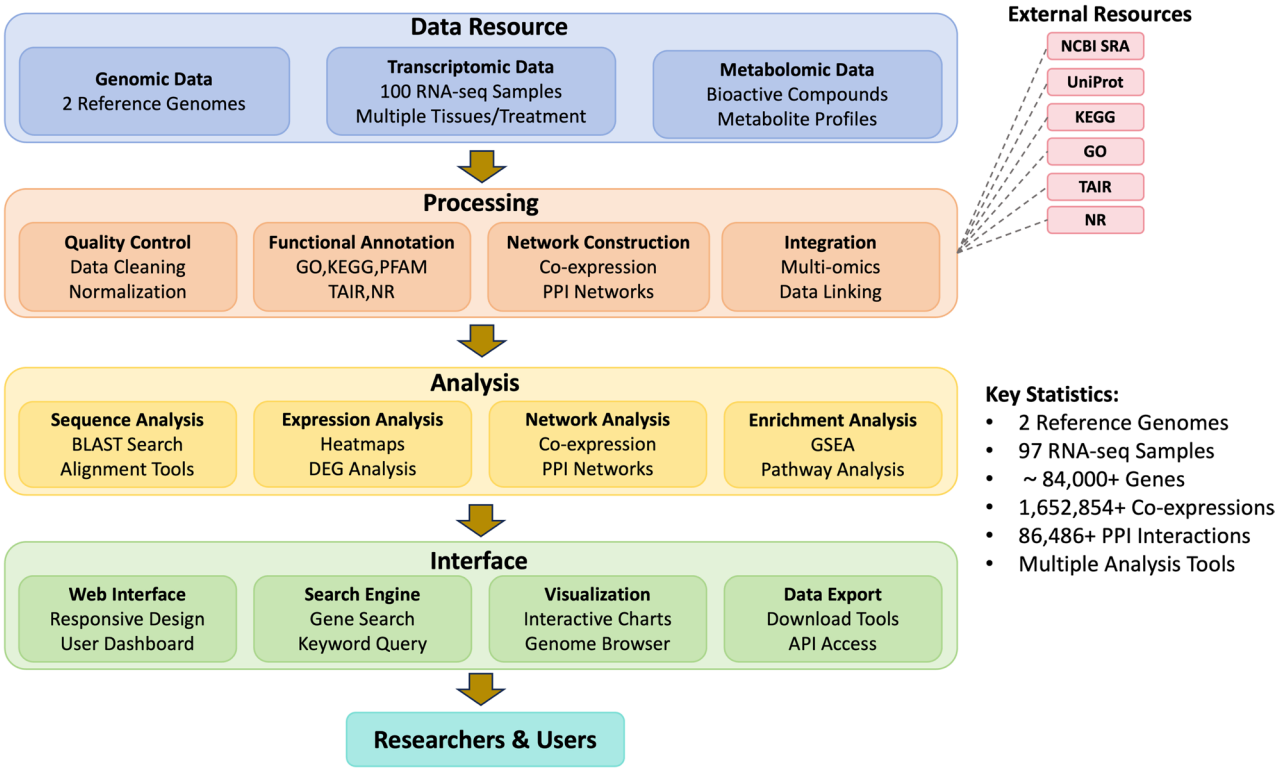

Fig. 1 Schematic overview of the ASAPv2 platform, illustrating its data resources, processing pipeline, analytical capabilities, user interface, and key metrics

Analysis (GSEA) module is provided for enrichment analysis: users can input a list of genes (for example, co-expressed genes or differential expression results) and test for the over-representation of GO terms or KEGG pathways. The results are displayed with statistical scores and linked to functional categories. Additionally, ASAPv2 includes a customizable heatmap tool for visualizing expression patterns of selected genes across transcriptome samples. Finally, a sequence extraction utility allows users to retrieve different kinds of sequences for any gene or a set of genes in FASTA format, which can be useful for further analysis. ASAPv2 integrates JBrowse2, an interactive genome browser, allowing users to visualize genomic features and transcriptome samples mapping peaks. We have also developed differentially expressed gene (DEG) analysis tools for mining target genes, as well as differentially expressed metabolite (DEM) analysis tools for analyzing differentially expressed metabolites. Additionally, we provide primer design tools to facilitate users in obtaining primers required for gene cloning and developed an ID converter tool for converting gene IDs between different versions.

Download

We have also provided download links for bulk data (genome sequences, annotation files, expression matrices, etc.), ensuring that users can obtain the data for offline analysis if needed.

Collectively, the construction and content of ASAPv2 make it a comprehensive repository and analysis platform. By integrating multi-omics data and providing intuitive tools, ASAPv2 enables researchers to not only retrieve information on *A. sinensis* genes but also further mine the genes that function together with target genes through various networks, tools, and diverse visual information.

Gene detail pages

Each gene has a dedicated page summarizing all available information. This includes the gene's description and annotations (with hyperlinks to GO and KEGG), its genomic location (chromosome and coordinates), gene structure (exon-intron organization), and sequences (DNA and protein sequence). Importantly, the page also displays expression data (bar chart of the gene's expression across various samples), gene family labels, and links to any associated proteins. If the gene is part of the co-expression network, a network visualization widget shows the gene connected to its co-expressed partners.

Maintenance and update

Collect *A. sinensis* genomic/transcriptomic research results quarterly to supplement gene sequences and

structural annotations, and expand the basic database; integrate published functional verification data (phenotypes, gene interactions, etc.) annually to improve gene function annotations. Evaluate tool performance every six months, and add functions (such as secondary metabolic pathway enrichment modules, gene network reconstruction).

Utility and discussion

Key advancements in ASAPv2

ASAPv2 introduces several important enhancements that distinguish it from the initial version of the database and significantly improve its value for users compared to ASAP v1 (Table 2).

Usage and case example

To demonstrate the utility of ASAPv2, we investigated genes involved in the flowering process of *A. sinensis*. First, by comparing the transcriptome data (accession ID: SRP232992) of plants before and after bolting (Unbolted vs. Bolted), we identified 1,331 significantly up-regulated genes and 1,058 significantly down-regulated genes (Fig. 2A). GO enrichment analysis of the significantly up-regulated genes showed that genes related to primary meristem tissue development, photosynthesis, light reaction, and polysaccharide metabolic processes were significantly enriched (Fig. 2B). In addition, genes associated with negative regulation of flower development were also found in the up-regulated gene set (p-value = 0.067, Table S2). KEGG enrichment analysis indicated that many genes involved in pathways such as hormone signal transduction and flavonoid synthesis were significantly enriched (Fig. 2C). GO enrichment analysis of the down-regulated genes revealed that many terms related to bolting and flowering were significantly enriched, including maintenance of inflorescence meristem identity, response to jasmonic acid, and leaf senescence (Fig. 2D, Table S3). Moreover, genes associated with positive regulation of flower development were also present in the down-regulated gene set (p-value = 0.061, Table S3). KEGG enrichment analysis for down-regulated genes suggested that Phenylpropanoid biosynthesis and Carotenoid biosynthesis showed significant enrichment (Fig. 2E). We also analyzed the expression of genes related to flowering regulation. For example, the expression levels of genes such as *FT*, *SOC1*, *FT*-like, *AGL6*, and *AGL8* significantly increased after bolting, while the expression of FLC (a gene that inhibits flowering) was significantly down-regulated (Fig. 2F).

Based on the aforementioned transcriptome analysis, we found that a gene numbered AS11G00590 was significantly up-regulated after bolting. Next, we analyzed the detailed information of the gene and found that it is annotated as agamous-like, MADS-box, and suppressor of constans overexpression 1 (SOC1) in different databases (Fig. 3A). This gene is located on the positive strand of chromosome 11, spanning positions 66,909,973 to 66,930,591 bp. JBrowse2 visualization shows that the gene contains 8 exonic regions and is annotated with 5'-UTR and 3'-UTR sequences (Fig. 3B). Additionally, the CDS sequence is provided to facilitate sequence analysis and molecular biology research for researchers (Fig. 3C). On the gene detail page, users can access co-expression network connections and a list of the top 300 genes ranked by Pearson Correlation Coefficient (PCC) (Fig. 3D). Users interested in protein research can also obtain protein sequence and domain information (Fig. 3E and F). Gene family classification indicates that the gene belongs to the MADS-MIKC family within the transcription factor family (Fig. 3G). This gene is not annotated in the KEGG database, suggesting it does not function as a key enzyme in biological pathways (Fig. 3H). GO enrichment analysis reveals that SOC1 exhibits transcription

Table 2 A comprehensive comparison of features and capabilities between the original ASAP platform and the upgraded ASAPv2

Feature Category	ASAP v1.0	ASAP v2.0	Improvement
Genomic Data			
Reference Genomes	1 genome assembly	2 genome assemblies	100% increase
Transcriptomic Data			
RNA-seq Samples	59 samples	97 samples	increase
Tissue Types	Limited coverage	Comprehensive coverage	Enhanced diversity
Metabolomic Data			
Metabolite Profiles	Not available	Available	New integration
Metabolite-microorganism Links	Not available	Pathway-based connections	New functionality
Analysis Tools			
BLAST Search	yes	yes	Improved interface
Gene Set Enrichment	yes	yes	New functionality
Expression Heatmaps	Basic heatmaps	Interactive heatmaps	Enhanced visualization
Genome Browser	yes	yes	Added data sets
multiple sequence alignment	no	yes	new tools
sequence retrieval	yes	yes	Added new genome
differential expressed genes	no	yes	new added tools
differential expressed metabolites	no	yes	new added tools
Primer Design	no	yes	new added tools
User Interface			
Web Interface	Basic HTML interface	Modern responsive design	Complete redesign

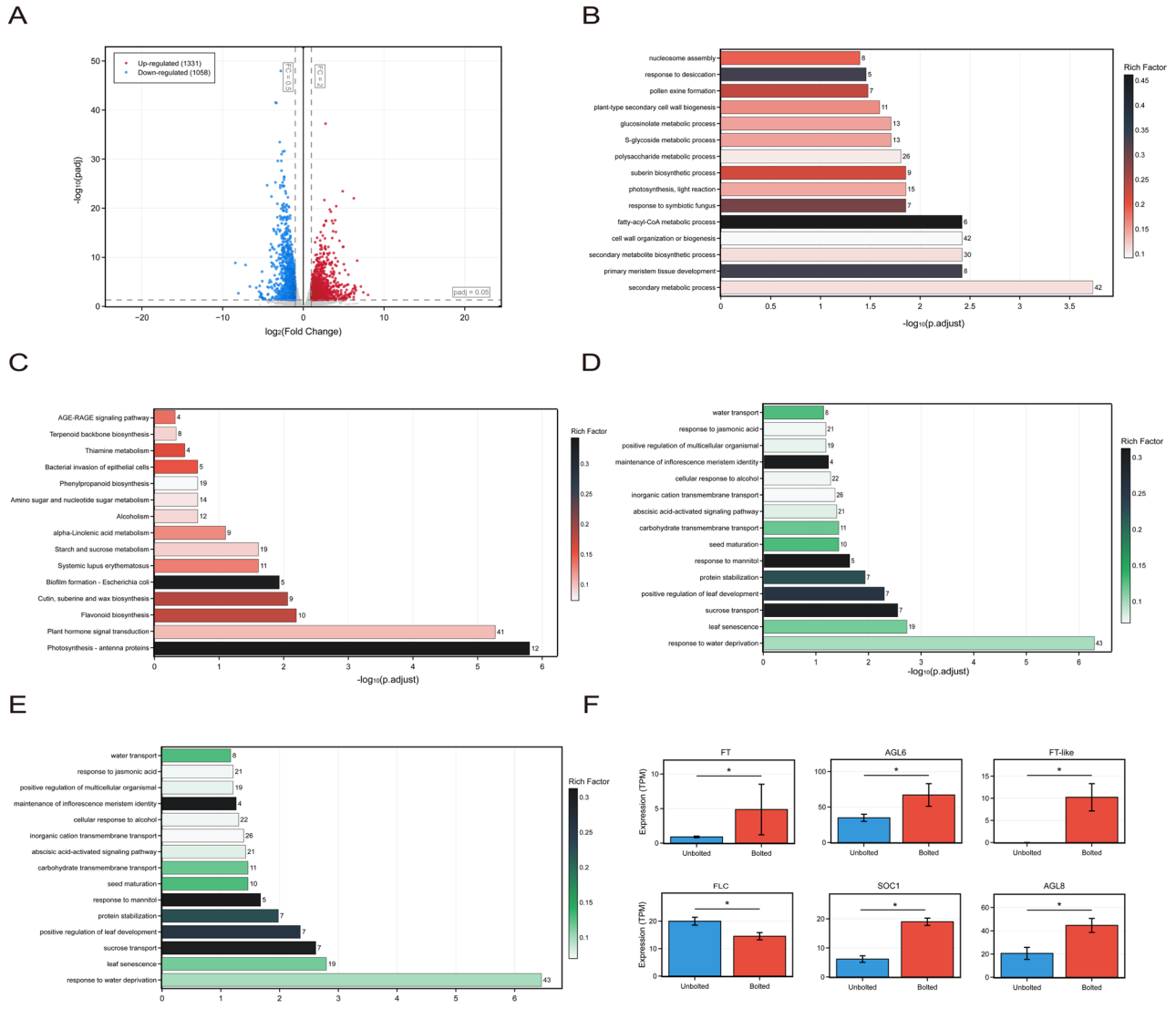

Fig. 2 Elucidating flowering-associated gene regulation in *A. sinensis* via ASAPv2. **A** Transcriptome analysis (accession ID: SRP232992) comparing unbolted versus bolted plants. **B** GO enrichment of up-regulated genes. **C** KEGG pathway enrichment analysis of up-regulated genes. **D** GO enrichment of down-regulated genes. **E** KEGG pathway enrichment analysis of down-regulated genes. **F** Expression dynamics of key flowering regulators: *FT*, *SOC1*, *FT-like*, *AGL6*, *AGL8* and *FLC*

factor activity (Fig. 3I). An expression profile visualization histogram is also provided (Fig. 3J).

Based on the JBrowse2 genome browser, it could be observed that the signal value of this gene after bolting was significantly higher than that before bolting (Fig. 4A), which further indicated that it might be involved in the bolting or flowering regulation process. Co-expression network analysis revealed that this gene had a positive co-expression relationship with 22 genes (Fig. 4B, Table S4). Several genes that are co-expressed with SOC1 have known functions related to flowering time control in Plants, both AS08G00118 and AS08G01965 were also annotated as AGAMOUS-like 20 (also SOC1) and promoter flowering in *Arabidopsis thaliana* and soybean [22, 23], AS08G00735 was annotated as homeobox knotted-like protein and participated in floral organ abscission in *Arabidopsis thaliana* [24], AS05G02661 was annotated as AGAMOUS-like 8 and function in controlling inflorescence arrest [25], and AS04G02159 was annotated as flavanone 3-hydroxylase and functions in formation of flower color [26], which suggest that SOC1 may play a similar role in *A. sinensis*. Comparative transcriptome (unbolted vs. bolted) analysis found that some co-expressed genes, such as AS05G02661 (AGL8), AS11G02180 (myo-inositol oxygenase 2), and AS08G01965 (Transcription activator), also showed significant down-regulation (Fig. 4C). These findings suggested that AS11G00590 might work together with these genes to regulate plant flowering, metabolism, and other processes. GO enrichment analysis of the positively

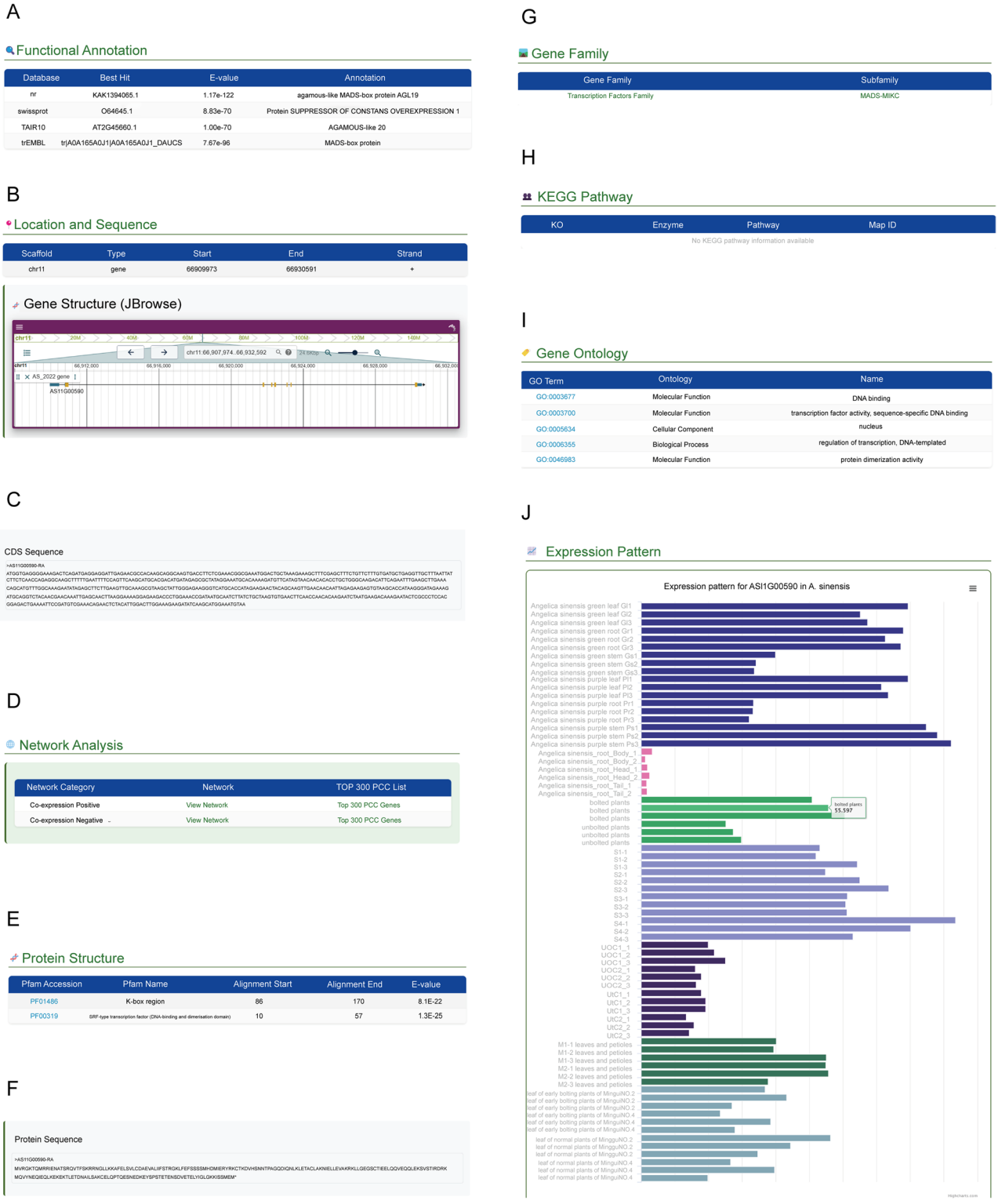

Fig. 3 Detailed characterization of the bolting-induced gene AS11G00590 in *A. sinensis* using ASAPv2. **A** Annotation summary. **B** Genomic localization with JBrowse2 visualization. **C** Coding sequence (CDS). **D** Co-expression network connections. **E** Pfam domain structure. **F** protein sequence. **G** Gene family classification.**H** KEGG annotation. **I** GO annotation. **J** Visualization histogram of the expression profile

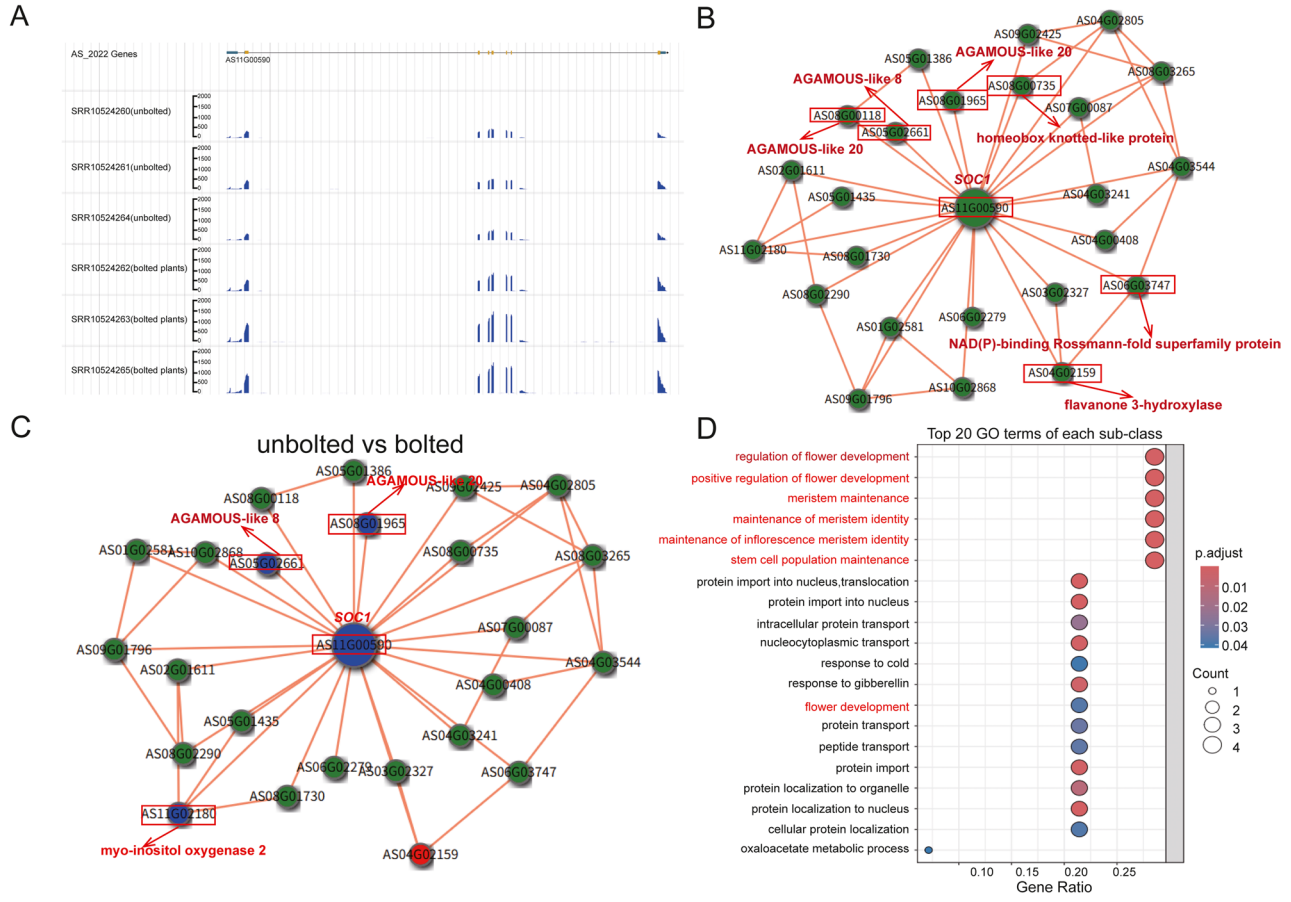

Fig. 4 Co-expression network and functional analysis of the bolting-induced gene AS11G00590 in *A. sinensis*. **A** JBrowse2 genome browser visualization of signal values for bolting and unbolting transcriptome samples. **B** Co-expression network analysis of AS11G00590. **C** Comparative transcriptome analysis (Unbolted vs. Bolted) of AS11G00590's co-expressed genes. **D** GO enrichment analysis of positively co-expressed genes

co-expressed genes showed that they were related to processes such as flower development, shoot system development, regulation of stomatal complex development, and response to hormone and stress (Fig. 4D). These might be the changes that occur during the flowering transition process.

AS11G00590 is annotated as AGAMOUS-like 20 (*AGL20*)/SOC1 in *Arabidopsis*, which has been shown to participate in the regulation of plant flowering in many plant species [27–30], which supports the results of our analysis of AGAMOUS-like 20 (*AGL20*)/SOC1 in *A. sinensis*. This example illustrates how ASAPv2 can be utilized to generate biological insights. The integration of multi-omics data in a single platform accelerates hypothesis generation: researchers can move from a gene query to functional context (annotation, networks, expression, phenotype correlation) within a few clicks. This holistic view is especially valuable in *A. sinensis*, a non-model organism where such information is not readily accessible otherwise. By consolidating data, ASAPv2 saves researchers substantial effort in gathering and analyzing disparate datasets.

Although ASAPv2 improves data integration and analysis for A. sinensis, it still has some limitations. Currently, the platform mainly focuses on a single species, and the coverage of metabolomic and other omics data (e.g., proteome and epigenome) remains limited. Additionally, some analysis tools are relatively basic compared to specialized pipelines, and the capability to process large datasets or user-uploaded datasets needs further improvement. Furthermore, due to the lack of data such as metabolomics data, dedicated tools for secondary metabolism research are still absent. Future development will aim to expand data types, enhance analytical modules, and improve interoperability to support broader applications.

The core gene function analysis results of the platform are derived from computational biology analyses. While such predictions can rely on algorithms to efficiently screen candidate genes, construct metabolic pathway associations, and provide a quick entry point for research, they are limited by the species coverage of model training data and the dynamic complexity of biological systems. Thus, their conclusions are only "predictive" rather than

"definitive" and must be verified through experiments such as molecular genetics and molecular biology to finally confirm the true function and regulatory role of genes. The current platform is hosted on an independent Tencent Cloud server, and users can contact the developers for private data updates.

Conclusion

ASAPv2 is a significant upgrade to the *A. sinensis* multi-omics database, offering an enriched collection of genomic, transcriptomic, and metabolomic data along with advanced analysis tools. The platform's comprehensive annotations and interactive networks allow researchers to uncover gene functions and interactions underpinning the development and phytochemistry of Danggui. By providing user-friendly bioinformatics tools, ASAPv2 lowers the barrier to in-depth analysis of this non-model medicinal plant. ASAPv2 is freely accessible and will be continually updated as new data emerges, ensuring that it remains a valuable and up-to-date resource. We conclude that ASAPv2 will accelerate fundamental research on *A. sinensis* and help unlock the molecular basis of its medicinal properties, ultimately supporting efforts in crop improvement and natural product research.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12362-7.

Supplementary Material 1

Supplementary Material 2

Author contributions

Q. X. and J. Y. conceived the project and supervised its execution. W.Z., S. W., and X. Z. were responsible for collecting and processing genomic and multi-omic data, as well as developing the database's back-end and front-end interfaces with the implementation of new features. W. Z., S.W. drafted the manuscript, and Q. X. and J. Y. revised it. All authors reviewed and approved the final version of the manuscript.

Funding

This work was supported by the National and Provincial Scientific and Technological Innovation Talent Team Cultivation Program of the Guizhou University of Traditional Chinese Medicine (GZYTDHZ[2022]003, GZYTDHZ[2024]002).

Data availability

All source data integrated into ASAPv2 (genome assemblies, sequence datasets, etc.) are from public repositories or literature as cited in the database. Processed data (annotations, networks, etc.) can be downloaded from the website. The ASAPv2 platform is publicly accessible on the website https://www.gzybioinformatics.cn/ASAPv2.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 14 August 2025 / Accepted: 20 November 2025
Published online: 04 December 2025

References

1. Wu YC, Hsieh CL. Pharmacological effects of radix *Angelica sinensis* (Danggui) on cerebral infarction. Chin Med. 2011;6:32.
2. Giacomelli N, et al. *Angelica sinensis* (Oliv.) Diels: influence of value chain on quality criteria and marker compounds ferulic acid and Z-ligustilide. Medicines. 2017. https://doi.org/10.3390/medicines4010014.
3. Wang K, et al. Self-assembled Angelica sinensis polysaccharide nanoparticles with an instinctive liver-targeting ability as a drug carrier for acute alcoholic liver damage protection. Int J Pharm. 2020;577:118996.
4. Qian W, et al. Angelica sinensis polysaccharide suppresses Epithelial-Mesenchymal transition and pulmonary fibrosis via a DANCR/AUF-1/FOXO3 regulatory axis. Aging Dis. 2020;11(1):17–30.
5. Han X, et al. The chromosome-level genome of female ginseng (Angelica sinensis) provides insights into molecular mechanisms and evolution of coumarin biosynthesis. Plant J. 2022;112(5):1224–37.
6. Li S, et al. Integrating genomic and multiomic data for *Angelica sinensis* provides insights into the evolution and biosynthesis of pharmaceutically bioactive compounds. Commun Biol. 2023;6(1):1198.
7. Zhu T, et al. Integrated physiological characterisation and transcriptomics reveals drought tolerance differences between two cultivars of *A. sinensis* at seedling stage. Mol Biol Rep. 2025;52(1):283.
8. Zhu C, et al. Integrated transcriptomic and metabolomic analysis reveals the regulation mechanism of early bolting and flowering in two cultivars of *Angelica sinensis*. Heliyon. 2024;10(7):e28636.
9. Liu P, et al. Untargeted metabolomics and functional analyses reveal that the secondary metabolite Quinic acid associates with *Angelica sinensis* flowering. BMC Plant Biol. 2025;25(1):72.
10. Reiser L, et al. The Arabidopsis Information Resource in 2024. Genetics. 2024. https://doi.org/10.1093/genetics/iyae027.
11. Wu S, et al. ASAP: a platform for gene functional analysis in *Angelica sinensis*. BMC Genomics. 2024;25(1):96.
12. Xu Q, et al. Chromosome-Scale assembly of the dendrobium nobile genome provides insights into the molecular mechanism of the biosynthesis of the medicinal active ingredient of dendrobium. Front Genet. 2022;13:844622.
13. Zhu T, et al. Integrated metabolomic and transcriptomic analysis reveals differential mechanism of flavonoid biosynthesis in two cultivars of *Angelica sinensis*. Molecules. 2022. https://doi.org/10.3390/molecules27010306.
14. UniProt C. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523–31.
15. Reiser L, et al. Using the Arabidopsis information resource (TAIR) to find information about Arabidopsis genes. Curr Protoc. 2022;2(10):e574.
16. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60.
17. Jones P, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.
18. Kanehisa M, Sato Y, Morishima K. BlastKOALA and ghostkoala: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726–31.
19. Zheng Y, et al. iTAK: A program for Genome-wide prediction and classification of plant transcription Factors, transcriptional Regulators, and protein kinases. Mol Plant. 2016;9(12):1667–70.
20. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.
21. Zhou J, et al. iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations. Nucleic Acids Res. 2018;46(D1):D447–53.
22. Barrero-Gil J, et al. Arabidopsis INHIBITOR OF GROWTH 2 promotes flowering by regulating NuA4-dependent H4 acetylation levels at FT and SOC1. Plant Physiol. 2025. https://doi.org/10.1093/plphys/kiaf511.
23. Na X, et al. Cloning and functional analysis of the flowering gene GmSOC1-like, a putative SUPPRESSOR OF OVEREXPRESSION CO1/AGAMOUS-LIKE 20 (SOC1/AGL20) ortholog in soybean. Plant Cell Rep. 2013;32(8):1219–29.

24. Shi CL, et al. Arabidopsis class I KNOTTED-like homeobox proteins act downstream in the IDA-HAE/HSL2 floral abscission signaling pathway. Plant Cell. 2011;23(7):2553–67.
25. Martinez-Fernandez I, et al. Analysis of pea mutants reveals the conserved role of FRUITFULL controlling the end of flowering and its potential to boost yield. Proc Natl Acad Sci U S A. 2024;121(15):e2321975121.
26. Walliser B, et al. The (bio)chemical base of flower colour in *Bidens ferulifolia*. Plants. 2022. https://doi.org/10.3390/plants11101289.
27. Yang Y, et al. Applications of multi-omics technologies for crop improvement. Front Plant Sci. 2021;12:563953.
28. Li H, et al. AP1c and SOC1 form a regulatory feedback loop to regulate flowering time in soybean. Plant Cell Environ. 2024. https://doi.org/10.1111/pce.15190.
29. Shi Y, et al. The SOC1 gene plays an important role in regulating Litchi flowering time. Genomics. 2024;116(2):110804.
30. Sri T, et al. Homeologs of brassica SOC1, a central regulator of flowering time, are differentially regulated due to partitioning of evolutionarily conserved transcription factor binding sites in promoters. Mol Phylogenet Evol. 2020;147:106777.

Publisher's Note