

Comparative Transcriptomics of Floral Transition in *Arabidopsis thaliana* and Ortholog Prediction in *Salvia hispanica*

Saad Khan

Centre for Omic Sciences, Islamia College Peshawar, Pakistan

Juan Philppe Teixeira

juan.bioinformatics@gmail.com

Graduate Program in Biochemistry and Molecular Biology (PMBqBM) <https://orcid.org/0000-0001-7323-2919>

Miklos Maximiliano Bajay

Graduate Program in Biochemistry and Molecular Biology (PMBqBM) <https://orcid.org/0000-0002-8196-3524>

Research Article

Keywords: Comparative transcriptomics, flowering time, *Arabidopsis thaliana*, *Salvia hispanica*, machine learning, bioinformatics

Posted Date: March 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9066517/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Saad Khan¹, Juan Philippe Teixeira², Ghufraan Ullah³, Muhammad Fahim^{*1}, Miklos Maximiliano Bajay^{*2}

KHAN, S; TEIXEIRA, J. P.²; Fahim, M^{*1}; BAJAY, M. M^{*2}

¹ Centre for Omic Sciences, Islamia College Peshawar, Pakistan

² State University of Santa Catarina – Center for Agricultural and Veterinary Sciences – Graduate Program in Biochemistry and Molecular Biology (PMBqBM)

³ Department of Plant Sciences, Quaid-i-Azam University, Islamabad, Pakistan

Comparative Transcriptomics of Floral Transition in *Arabidopsis thaliana* and Ortholog Prediction in *Salvia hispanica*

Abstract

The transition from vegetative growth to reproductive development is a critical phase in the plant life cycle, shaping reproductive success and yield through coordinated responses to internal cues and environmental signals. Here, we integrated comparative transcriptomics and unsupervised machine learning to analyze time-series RNA-seq data from *Arabidopsis thaliana* and to prioritize flowering-time regulators in *Salvia hispanica* (chia). K-means clustering of *Arabidopsis* expression profiles recovered clusters enriched in transposable elements (e.g., AT5G28627) and pseudogenes (e.g., AT2G24110) that peaked during the flowering transition (M4–M6), suggesting participation in flowering-associated regulatory programs. Temperature-stratified profiles (Col10, Col16, Col23, Col27) revealed condition-dependent dynamics among canonical flowering regulators. Using cross-species expression correlation, we identified three high-confidence chia candidates for key *Arabidopsis* flowering genes (FT, FLC, and SOC1), with Pearson correlation coefficients ranging from 0.9943 to 0.9992. Together, these results provide a data-driven shortlist of candidate flowering-time regulators in chia and a reusable computational framework for transferring flowering knowledge from model to non-model plants.

Keywords: Comparative transcriptomics; flowering time; *Arabidopsis thaliana*; *Salvia hispanica*; machine learning; bioinformatics.

INTRODUCTION

The *Arabidopsis thaliana* genome holds a significant place in botanical research as the first complete plant genome to be sequenced (ARABIDOPSIS GENOME INITIATIVE, 2000). This milestone helped establish The Arabidopsis Information Resource (TAIR), a curated platform that integrates gene functions, mutant phenotypes, and expression profiles (BERARDINI et al., 2015). Building on these foundations, the 1001 Genomes Project mapped natural variation across more than one thousand accessions, enabling population-scale analyses of genotype–phenotype diversity (WEIGEL; MOTT, 2009; 1001 GENOMES CONSORTIUM, 2016). More broadly, continued advances in high-throughput sequencing have increased throughput and reduced costs, supporting large-scale genomics and transcriptomics that can be leveraged for comparative, cross-species inference—an approach explored here to prioritize flowering regulators in chia (GOODWIN; MCPHERSON; MCCOMBIE, 2016; VAN DIJK et al., 2018).

Compared with larger, repeat-rich genomes, the compact *Arabidopsis* reference has been especially amenable to short-read resequencing and transcriptome profiling at scale. This practicality accelerated community resources and enabled integrative analyses linking sequence variation to gene expression and phenotypes, which motivates transcriptome-driven discovery pipelines in both model and non-model plants (GOODWIN; MCPHERSON; MCCOMBIE, 2016; VAN DIJK et al., 2018).

Natural accessions of *A. thaliana* display remarkable levels of genetic and phenotypic variation (KOORNNEEF; ALONSO-BLANCO; VREUGDENHIL, 2004). Over the past decades, quantitative trait locus (QTL) mapping has been instrumental in identifying genetic variants that influence a wide range of physiological and developmental traits, including germination, flowering time, and ion content (ALONSO-BLANCO; KOORNNEEF, 2000). While earlier studies often relied on prior knowledge of gene function, later work highlighted key loci such as the flowering regulator *FRIGIDA* (*FRI*) (JOHANSON et al., 2000) and *DELAYED GERMINATION1* (*DOG1*) (BENTSINK et al., 2006), both encoding proteins with previously unknown biochemical roles. These examples illustrate how unbiased genetic mapping can reveal regulators without prior functional annotation, supporting the rationale for transcriptome- and machine-learning-based strategies to prioritize candidate genes in less-characterized species such as chia.

The Science of Flowering Time Genes

phanerogams Flowering is a critical developmental transition that strongly influences reproduction, crop yield, and fitness under diverse environmental conditions (ANDRÉS; COUPLAND, 2012; FORNARA; COUPLAND, 2009). Characterizing flowering time is therefore essential for understanding gene regulatory mechanisms underpinning plant development and for guiding translational efforts to improve agricultural productivity and adaptability (SUNG; AMASINO, 2004).

A. thaliana In *Arabidopsis*, six major genetic pathways regulate floral induction: the age, vernalization, gibberellin, photoperiod-dependent, and autonomous pathways (SUNG; AMASINO, 2005; KOMEDA, 2004). These converge on a set of floral integrator genes, including FLOWERING LOCUS T (FT) and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) (KOMEDA, 2004). Among them, FT plays a central role in integrating signals related to ambient temperature, vernalization, and photoperiodic cues, with its expression inversely correlated with flowering time and leaf number at the reproductive transition (KOMEDA, 2004). The gene CONSTANS (CO) promotes FT expression and mediates day length perception (PARK et al., 1999), whereas FLOWERING LOCUS C (FLC) represses FT, thereby influencing the distinction between summer and winter annual phenotypes (MICHAELS; AMASINO, 1999; SHELDON et al., 2000).

Vernalization, Autonomous Pathways, Temperature, and MicroRNA Regulation of Flowering in *Arabidopsis thaliana*

Vernalization promotes flowering after prolonged cold exposure, largely by epigenetically repressing FLOWERING LOCUS C (FLC), a major MADS-box repressor of flowering. FLC inhibits photoperiodic integrators such as FT and SOC1, and its stable silencing involves chromatin remodeling and long non-coding RNAs (e.g., COOLAIR). Repressors such as SHORT VEGETATIVE PHASE (SVP) can form complexes with FLC to further delay flowering by targeting FT and SOC1 promoters.

The autonomous pathway also reduces FLC activity through RNA processing and chromatin-based mechanisms, whereas ambient temperature provides an additional cue that modulates flowering outputs, particularly under non-inductive photoperiods. Developmental phase change is coordinated by microRNA modules (miR156–SPL and miR172), linking age-dependent programs to reproductive competence. Together, these pathways converge on shared integrators (FT/SOC1) to ensure robust flowering across environments (ANDRÉS; COUPLAND, 2012).

The Science of Flowering Time in *Salvia hispanica* (Chia)

Salvia hispanica, commonly known as chia, is an annual plant from the Lamiaceae family, traditionally cultivated in Mexico, Colombia, and Guatemala (CAHILL, 2003). Its seeds were a staple food of the Aztec civilization and remain valued today as a source of energy and nutrition (AYERZA; COATES, 2005). Chia seeds contain high lipid levels and are among the richest plant sources of omega-3 fatty acids; physicochemical characterization of chia seed oils has been reported using pressing and solvent extraction approaches (IXTAINA et al., 2011). In addition, chia seeds contain phenolic compounds with antioxidant activity, and UHPLC-based profiling has identified and quantified major phenolic acids and isoflavones (MARTÍNEZ-CRUZ; PAREDES-LÓPEZ, 2014). Reviews have summarized the phytochemical profile, isolation approaches, and applications of chia seeds (HRNČIČ et al., 2019).

At the molecular level, transcriptomic studies in chia leaves identified 368 differentially expressed transcripts (DETs), particularly enriched in mature tissues, where both up- and down-regulated genes were observed (GUPTA et al., 2021). During early leaf development, transcription factors such as GROWTH REGULATING FACTORS (GRF2, GRF5) and the bHLH-domain protein SPEECHLESS showed high expression, regulating leaf size, shape, and early floral structures. As flowering progresses, 414 DETs are detected, including members of the MYB family (e.g., MYB54 and MYB52), which regulate floral patterning and morphogenesis. Moreover, more than 2,600 genes are up-regulated during reproductive development, including AGAMOUS-LIKE MADS-box proteins and other MYB transcription factors, highlighting their importance in flower and leaf formation (GUPTA et al., 2021). Notably, genes encoding terpene synthases were up-regulated in floral tissues, supporting their role in the production of scent and flavor compounds.

Machine Learning for Cross-Species Prediction

Machine Learning (ML), a branch of Artificial Intelligence, has become a powerful tool in genomics, enabling applications such as data analytics, gene discovery, functional prediction, and the identification of novel genes (LIBBRECHT; NOBLE, 2015). Recent advances include multi-environment and multi-trait prediction models for flowering plants (HESLOT et al., 2012; CROSSA et al., 2017). For example, Meher et al. (2018) developed an ML model to identify circadian clock genes in plants, while Zhang et al. (2020) applied ML to predict senescence-associated genes. A significant gap, however, remains in developing models based on protein sequence data for flowering-time-associated genes (FTAGs). To address this, Zhang et al. (2020)

trained ML models to recognize proteins encoded by FTAGs, demonstrating that ML can outperform traditional statistical methods by focusing on prediction with fewer assumptions and by handling large, noisy, and high-dimensional datasets (CHING et al., 2018).

In the present study, we applied diverse ML algorithms to analyze and visualize FTAGs in *Arabidopsis thaliana* using time-series RNA-seq data (KLIPIKOVA et al., 2016). Implemented in Python with dedicated ML libraries, our models achieved high accuracy in identifying known FTAGs and predicting novel candidates. We also employed clustering approaches, such as the K-means algorithm (MACQUEEN, 1967), to group genes with similar expression patterns, allowing us to identify putative flowering regulators. Additionally, we explored the role of transposable elements, providing further insight into the complexity of gene regulation during flowering.

Finally, we extended this cross-species approach to *S. hispanica*, leveraging the conservation of flowering pathways between chia and *Arabidopsis*. Our findings suggest that ML can be effectively used to predict flowering-time-associated genes in non-model species, generating valuable insights into the genetic regulation of flowering time in chia.

MATERIAL AND METHODS

Computing Requirements

The computational requirements for this study involved the transcriptome analysis of floral transition in *Arabidopsis thaliana* followed by ortholog prediction in *Salvia hispanica*. All computational tasks were performed in a Linux environment (Ubuntu 22.04.4 LTS), using a combination of programming and bioinformatics tools. The analysis pipeline was managed through Anaconda (Conda 23.7.4) (ANACONDA, 2023) and followed the Tuxedo protocol for RNA-seq analysis (TRAPNELL; PACHTER; SALZBERG, 2009; PERTEA et al., 2016). This pipeline included the following tools: FastQC for quality control, SAMtools for file manipulation, Hisat2 and Bowtie2 for alignment, StringTie for transcript assembly, and Ballgown/Cuffdiff in R for differential expression analysis. Additionally, Google Colab was employed as a cloud-based platform for visualization and machine learning predictions using FPKM datasets.

The minimum computational requirements for running these analyses were as follows:

Component	Specification
Processor	A modern multi-core CPU with at least 4 cores.
RAM	Minimum of 16–32 GB to handle large-scale transcriptomic data.
Storage	SSD with at least 500 GB to 1 TB to support the operating system, software, and temporary files.
Graphics card	Not strictly required for this type of analysis; however, an integrated or entry-level GPU can be useful for visualization and machine learning tasks.
Internet	Stable and fast internet connection for Linux updates, running commands, and accessing online resources.
SRA	Datasets retrieved using the SRA Toolkit via command-line (BROWN et al., 2016).

Source: Created by the author (2025).

Bioinformatics Analysis

A variety of bioinformatics tools and online databases were employed to support each step of the workflow, from data acquisition and quality control to alignment, expression quantification, and downstream functional annotation. The computational environment was set up using R programming language together with the RStudio IDE (Posit, 2023), widely used for statistical analysis and data visualization. Package management and reproducible environments were maintained with Anaconda (Conda 23.7.4) (Anaconda, 2023). Raw sequencing data were retrieved from the NCBI Sequence Read Archive (SRA) using the SRA Toolkit, with RNA-seq datasets from *Arabidopsis thaliana* (Klipikova et al., 2016) and *Salvia hispanica* (Gupta et al., 2021) accessed via their accession numbers. Quality assessment of raw reads was carried out using FastQC, followed by adapter and low-quality trimming with Trimmomatic (Bolger; Lohse; Usadel, 2014).

High-quality reads were then aligned to the respective reference genomes: TAIR10 for *Arabidopsis thaliana* and the Salvia Genome Database (salviagdb.org) for *S. hispanica* (Lamesch et al., 2012) Alignment was performed using HISAT2, an efficient splice-aware aligner for RNA-seq data, with Bowtie2 employed for specific mapping tasks when required (Langmead; Salzberg, 2012; Kim; Paggi; Salzberg, 2019). The resulting BAM files were processed with SAMtools for

sorting, indexing, and file manipulation (Li et al., 2009). Transcript assembly and quantification were conducted using StringTie (Pertea et al., 2016), and differential gene expression analysis was performed with the Ballgown package in R (Frazier; Trapnell, 2015).

Finally, unsupervised machine learning methods were applied to explore co-expression clusters and predict flowering-related genes. Analyses were run in Google Colab, leveraging Python libraries such as pandas, matplotlib, seaborn, and plotly for data handling and visualization, and scikit-learn for implementing clustering algorithms including K-means, hierarchical clustering, DBSCAN (Ester et al., 1996), and principal component analysis (PCA; Jolliffe, 2002). This integrative approach enabled both functional insights and predictive modeling of flowering-time-associated genes.

RESULTS

Transcriptomic Analysis of *Arabidopsis thaliana* and *Salvia hispanica*

Transcriptomic datasets from *Arabidopsis thaliana* and *Salvia hispanica* were analyzed across multiple developmental stages to investigate the dynamics of flowering-time gene expression. For *A. thaliana*, RNA-seq data comprised eight distinct meristem developmental stages (M1–M8), spanning 7 to 14 days post-germination. Across these stages, a total of 33,942 genes were detected. In contrast, *S. hispanica* data included five developmental stages sampled in five biological replicates, encompassing 38,695 genes (Klipikova et al., 2016; Gupta et al., 2021). Differential gene expression analysis in *A. thaliana* revealed variable changes in gene counts between developmental transitions, ranging from only 13 genes (M1–M2) to as many as 4,890 genes (M7–M8). Early stages (M1–M4), associated with leaf primordia formation, were characterized by activation of *SOC1*, concurrent with reduced *FLC* expression and low *LFY* levels, reflecting the onset of flowering primordia initiation (Parcy, 2005).

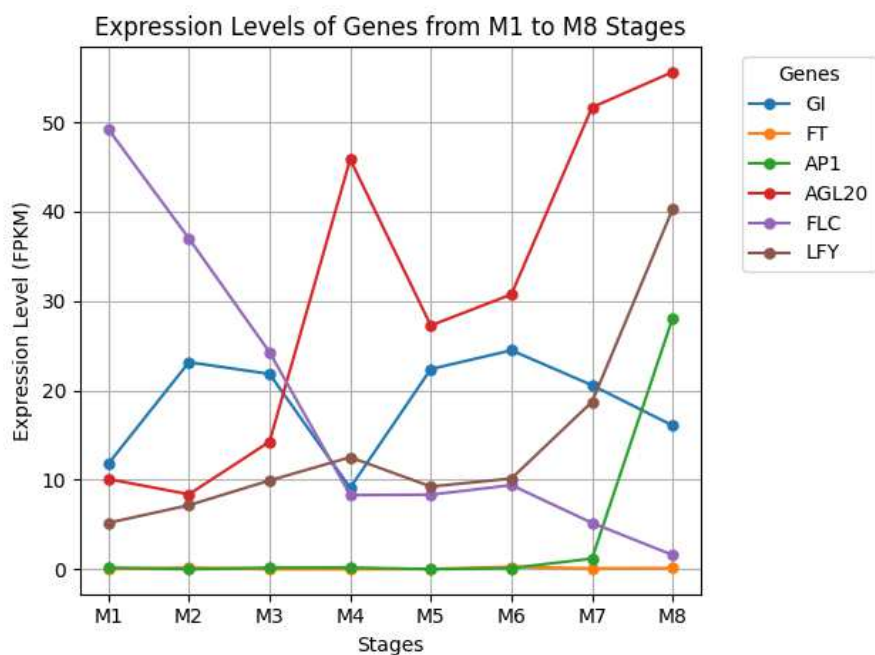
Table 1 – Differentially expressed genes at different meristem stages in *Arabidopsis thaliana* (Klepikova et al., 2016)

Stages	DE genes
M1-M2	13
M4-M5	3552
M5-M6	3019

M6-M7	770
M7-M8	4890

Source: Created by the author (2025).

Figure 2 – Reproducibility of flowering time gene expression in *Arabidopsis thaliana* using Klepikova et al. (2016) dataset and machine learning approaches



This figure shows the temporal expression profiles of key flowering regulators, including *FT*, *FLC*, *GI*, *AP1*, and *AGL20*, across eight developmental stages (7–14 days after germination) based on the public RNA-seq dataset of Klepikova et al. (2016). Unlike the original analysis conducted with the proprietary CLC Genomics Workbench, these results were reproduced using open-source Python and R libraries, in combination with the scikit-learn machine learning package and the K-means clustering algorithm. The comparable expression dynamics confirm the integrity of the open-source workflow and the robustness of the dataset, while highlighting the enhanced reproducibility, transparency, and scalability achievable through advanced computational methods.

Source: Adapted from Klepikova et al. (2016).

Reproducibility of Klepikova et al. Dataset

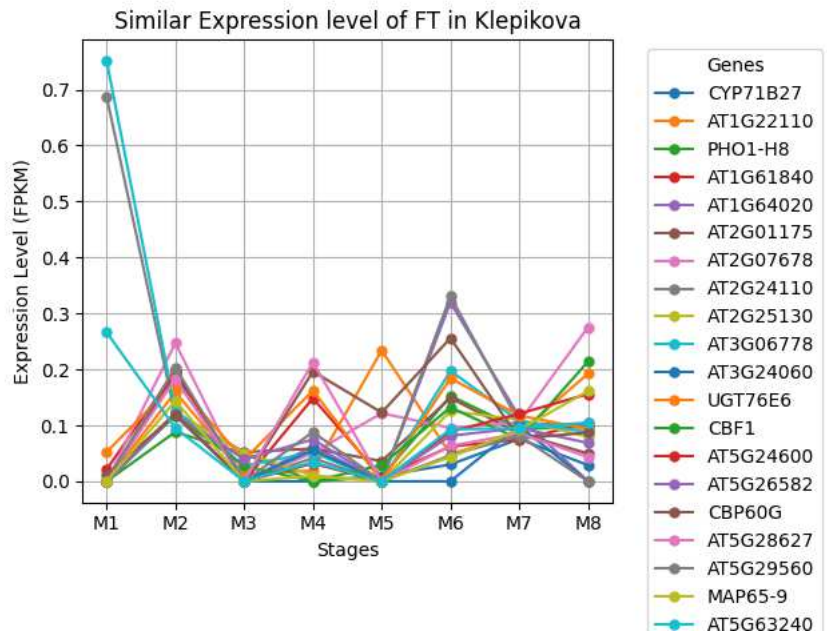
A primary outcome of this study was the rigorous reproduction and extension of the key findings reported by Klepikova et al. (2016), with a focus on demonstrating methodological robustness and computational advancement. While the original analysis employed CLC Genomics Workbench 6.5.1 (Qiagen Digital Insights, 2013) with a fixed quality score threshold of 0.005, our research implemented a fully open-source pipeline based on Python and R environments. This framework

provided greater transparency, flexibility, and reproducibility. Specifically, we integrated machine learning approaches by applying the scikit-learn library (Pedregosa et al., 2011) and the K-means clustering algorithm to systematically analyze and visualize gene expression patterns in a time-series context. Using this approach, Figure 2 presents the temporal expression profiles of pivotal flowering regulators, including *FLOWERING LOCUS T* (*FT*), *FLOWERING LOCUS C* (*FLC*), *GIGANTEA* (*GI*), *APETALA1* (*API*), *LEAFY* (*LFY*), and *AGAMOUS-LIKE 20* (*AGL20*), across developmental stages from 7 to 14 days post-germination. Our findings confirmed the reproducibility of the previously published expression trajectories, while providing a more scalable and transparent bioinformatics framework for future transcriptomic studies.

Clustering of Known Genes and Cross-Species Prediction

The clustering analysis of *FLOWERING LOCUS T* (*FT*) and its associated network revealed distinct expression dynamics across developmental and environmental contexts. *FT*, recognized as a key florigen (CORBESIER et al., 2007), plays a central role in floral induction by transmitting flowering signals from the leaves to the shoot apical meristem (SAM), where it interacts with *FD* to activate downstream developmental pathways (ABE et al., 2005). As shown in **Figure 3**, our clustering identified co-expressed genes such as *CYP71B27*, *UGT76E6*, *CBF1*, and *MAP65-9*, all of which displayed stage-specific expression during the growth and development of *Arabidopsis thaliana*. Within the temperature-responsive dataset (Col10–Col27; LI et al., 2018), *FT* expression peaked sharply at Col16 and Col23, reflecting its sensitivity to thermal cues and underscoring its role as a regulatory nexus between temperature perception and floral induction.

Figure 3 – Clustering of *FLOWERING LOCUS T* (*FT*) and Co-Expressed Genes Across Developmental and Temperature-Responsive Stages in *Arabidopsis thaliana*

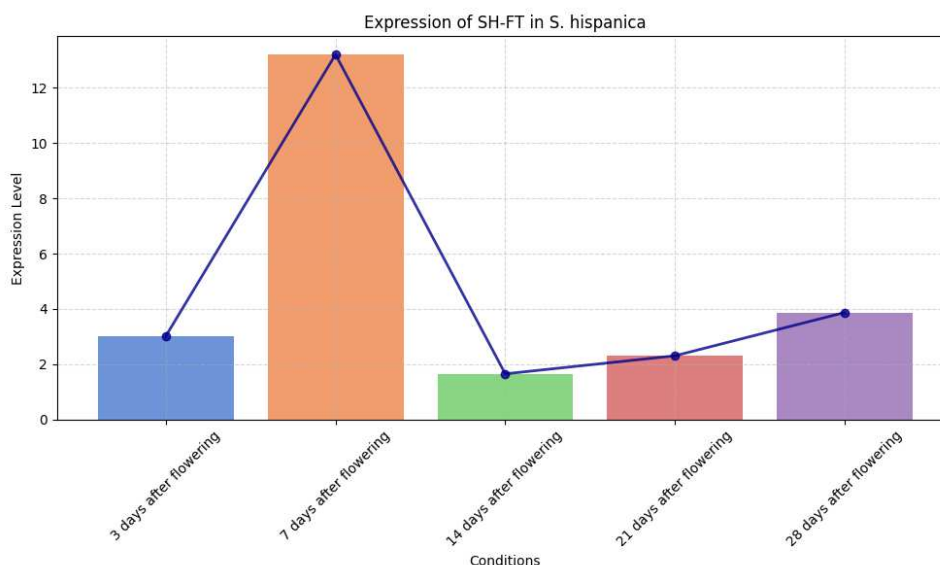


This figure shows the clustering of *FT* together with co-expressed genes (*CYP71B27*, *UGT76E6*, *CBF1*, *MAP65-9*) across developmental stages (M1–M8) and temperature-responsive conditions (Col10–Col27). The results highlight stage-specific expression patterns and demonstrate the sensitivity of *FT* expression to thermal cues, with peaks observed at Col16 and Col23. These findings illustrate the regulatory role of *FT* at the interface between environmental perception and floral induction.

Among the predicted genes in *Salvia hispanica*, the candidate exon-XM_048087101.1-1 was identified as a robust *FT* ortholog (*shFT*), suggesting a significant role as a flowering time regulator (Figure 4). The Pearson correlation coefficient (PCC) between the expression profiles of this *Salvia FT* candidate and the canonical *Arabidopsis FT* gene was 0.9943, indicating an exceptionally strong positive correlation. This value, approaching a perfect correlation, strongly supports the functional conservation of *FT* between the model species and chia. Given that PCC values above 0.9 are typically considered very strong, our results validate the predicted homology and suggest that *shFT* likely performs a conserved role in flowering induction in *Salvia hispanica*.

This finding highlights the effectiveness of our comparative transcriptomics and machine learning framework in accurately identifying key flowering regulators in non-model plants, reinforcing the translational value of model–non-model cross-species predictions.

Figure 4 – Identification of a Putative *FLOWERING LOCUS T* (*FT*) Ortholog in *Salvia hispanica* Based on Expression Correlation with *Arabidopsis thaliana*



This figure presents the identification of the *Salvia hispanica* gene **exon-XM_048087101.1-1** as a putative *FT* ortholog (*shFT*). The expression profile of *shFT* showed a Pearson correlation coefficient (PCC) of 0.9943 with canonical *FT* in *Arabidopsis thaliana*, indicating a highly conserved regulatory role. The results confirm the functional homology between species and validate the predictive power of comparative transcriptomics combined with machine learning.

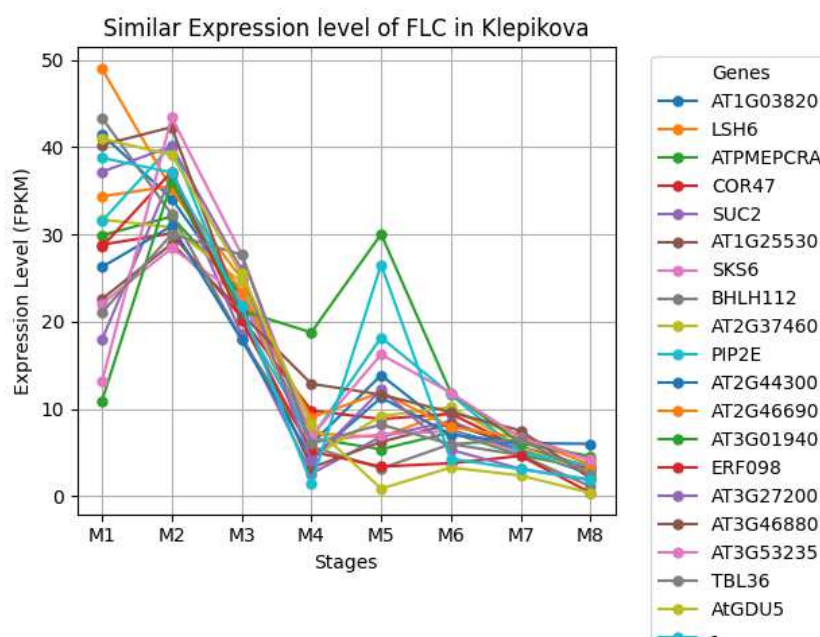
Source: Generated from transcriptomic datasets of *Arabidopsis thaliana* (Klepikova et al., 2016) and *Salvia hispanica* (Gupta et al., 2021), with additional interpretation based on clustering analysis.

Among the predicted genes in *Salvia hispanica*, exon-XM_048087101.1-1 gene was identified as a robust FT ortholog (sh. FT), predicting a highly significant role as a flowering time regulator Figure 4. The Pearson Correlation Coefficient (PCC) between the expression profiles of this *Salvia* FT candidate and the canonical *Arabidopsis* FT gene was calculated at 0.9943, indicating an exceptionally strong positive correlation. This value, approaching a perfect positive correlation, strongly supports the functional conservation of FT between the model species and *Salvia*. Given that correlation coefficients greater than 0.9 are typically classified as very strong, these results validate the predicted homology and suggest that this sh. FT gene likely performs a conserved role in flowering induction within *Salvia*. This finding underscores the effectiveness of our comparative transcriptomics and ML approach in accurately identifying key flowering regulators in non-model plants.

Clustering and Temperature Sensitivity of *FLOWERING LOCUS C (FLC)*

FLOWERING LOCUS C (FLC) is a central MADS-box transcription factor that integrates the vernalization pathway and functions as a major floral repressor (MICHAELS; AMASINO, 1999). Its expression decreases during vernalization, thereby fine-tuning the timing of flowering in response to prolonged cold (SUNG; AMASINO, 2004). Despite extensive characterization of individual flowering regulators, the broader transcriptome-wide networks and their temperature-dependent dynamics remain incompletely understood. Our clustering analysis of *FLC* expression in *Arabidopsis thaliana* revealed a co-expressed gene set that included *LSH6*, *SUC2*, *SKS6*, *PIP2E*, *ERF098*, and *TBL36*, all of which mirrored the *FLC* expression profile, particularly during flower development (**Figure 5**). These genes are candidate members of regulatory modules influencing the floral transition. When examining temperature-responsive datasets, *FLC* expression was elevated at intermediate conditions (Col10, Col16, and Col23) but declined at higher temperatures (**Figure 7**). This temperature sensitivity aligns with the notion that elevated temperatures can disrupt *FLC* function, potentially through mechanisms that impair its repressor activity and facilitate floral induction (KIM; SCHNEIDER; HONG, 2009). Furthermore, clustering under variable thermal regimes consistently grouped *FLC* with genes such as *ATWRKY60*, *ATDTX31*, *AT3G57780*, *UCC3*, and *AT3G54120*, reinforcing the hypothesis that *FLC* operates within a broader temperature-responsive transcriptional network governing flowering readiness and plant adaptation.

Figure 5 – Clustering of *FLOWERING LOCUS C (FLC)* and Co-Expressed Genes During Flower Development in *Arabidopsis thaliana*

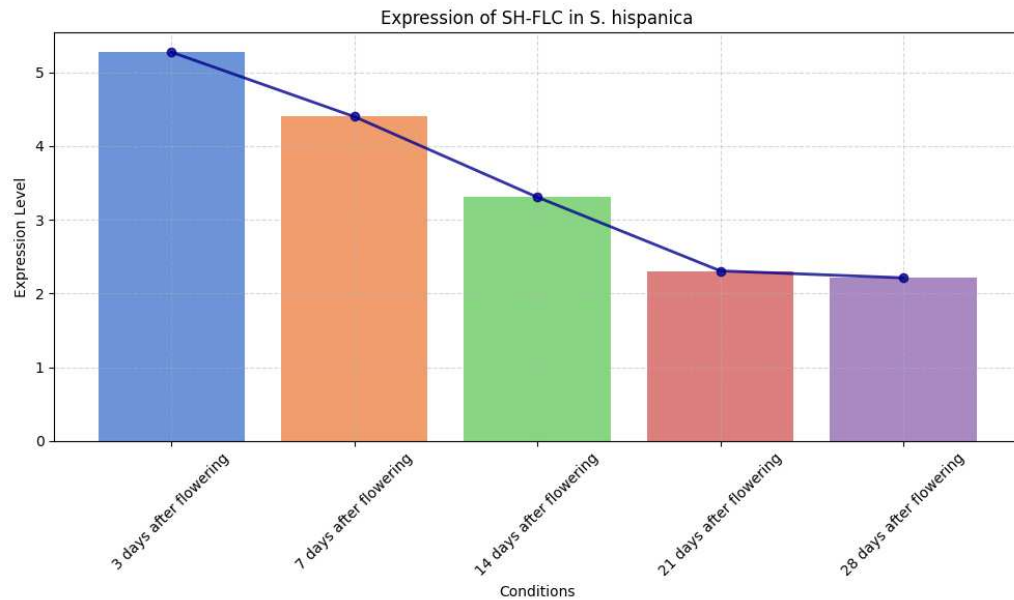


Source: Generated from transcriptomic datasets of *Arabidopsis thaliana* (Klepikova et al., 2016), with additional interpretation based on clustering analysis.

Ortholog of *FLC* in *Salvia hispanica*

Interrogation of the *Salvia hispanica* transcriptome identified the gene exon-XM_048090739.1-3 (*shFLC*) as a robust ortholog of *Arabidopsis thaliana FLC* (**Figure 6**). This gene displayed a highly synchronized expression pattern with *Arabidopsis FLC*, supported by a strong correlation coefficient. Notably, *shFLC* showed marked expression as early as three days post-germination, suggesting a specialized role in seedling establishment or initial developmental transitions. This transient early-stage expression implies that *shFLC* may act as a temporal regulator, integrating developmental and environmental signals during seedling growth. The closely aligned expression profiles of *shFLC* and *Arabidopsis FLC* highlight conserved functional importance, underscoring the preservation of regulatory strategies for flowering time and possibly stress adaptation across distantly related plant species.

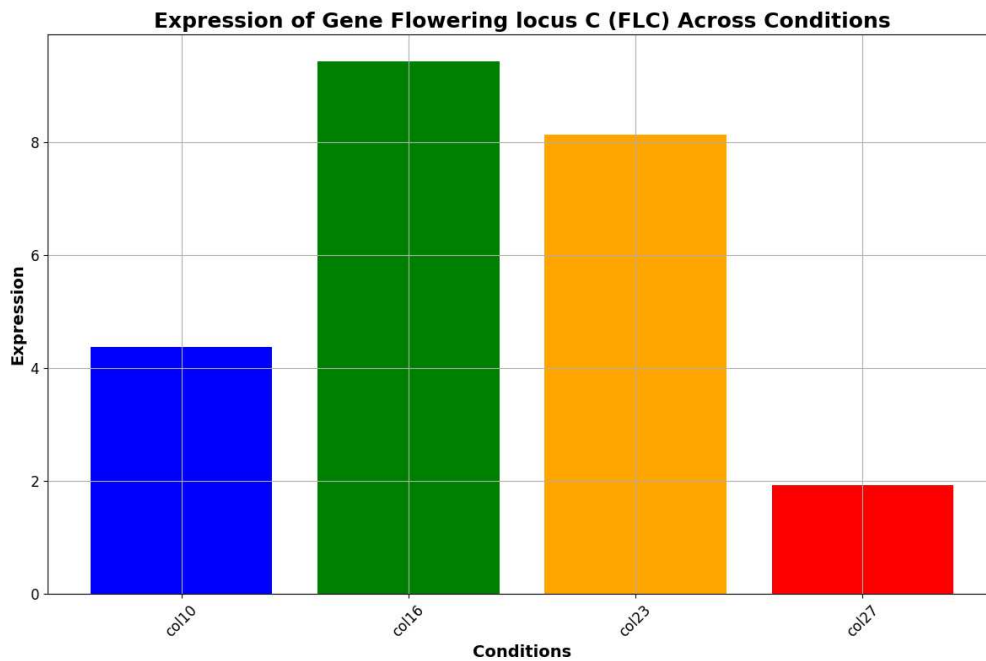
Figure 6 – Ortholog of FLC in *Salvia hispanica* Dataset



This figure shows the expression profile of *shFLC* (exon-XM_048090739.1-3), identified as a strong ortholog of *Arabidopsis thaliana FLC*. The synchronized expression pattern across early developmental stages suggests a role in seedling establishment and early flowering regulation.

Source: Generated from transcriptomic datasets of *Salvia hispanica* (Gupta et al., 2021) and *Arabidopsis thaliana* (Klepikova et al., 2016).

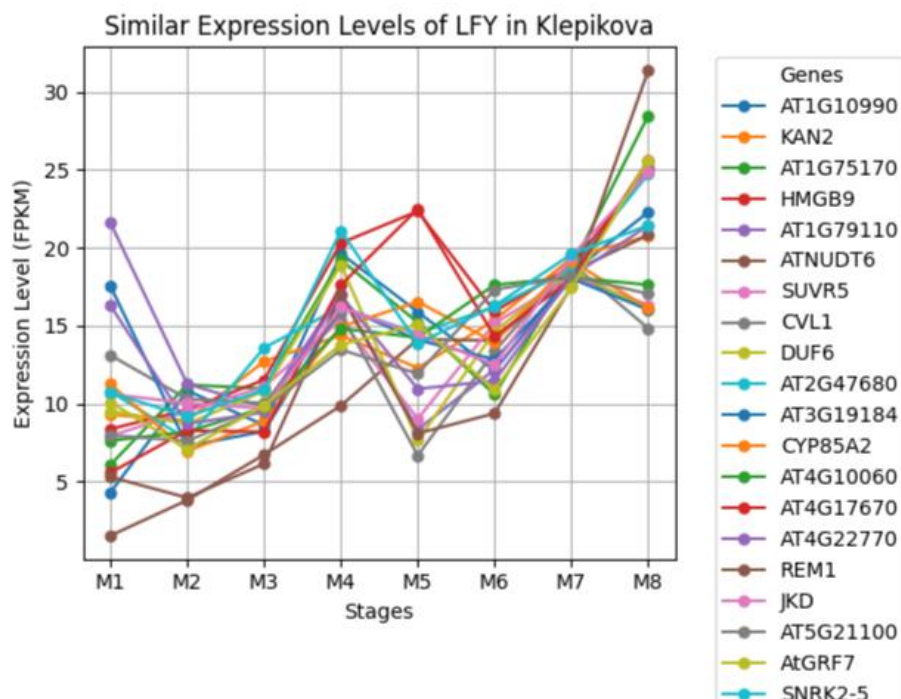
Figure 7 - Expression of FLC at Different Temperature



Dynamics of *FLC* and *LFY* During Floral Transition

In *Arabidopsis thaliana*, three distinct phases of *FLC* and *LFY* dynamics were observed. The first phase (M1–M4) corresponded to vegetative meristem growth, characterized by a decline in *FLC* expression accompanied by low *LFY* levels. The second phase (M4–M6) represented a transitional stage with roughly equivalent expression of *FLC* and *LFY*. Finally, the third phase (M6–M8) marked the activation of flowering, where *LFY* expression rose sharply while *FLC* expression further decreased. Clustering analysis of the *LFY* gene identified a co-expressed module of approximately 20 genes with similar temporal expression patterns (**Figure 8**). This cluster included established flowering regulators such as *KANADI2*, *HMGB9*, *ATNUDT6*, *SUVR5*, *CVL1*, *DUF6*, *CYP85A2*, and *REMI*, reflecting their coordinated roles during floral transition. Temperature-dependent analysis revealed that *LFY* expression peaked at moderate ranges (Col16–Col23), with a sharp decline beyond Col23 (**Figure 10**). This pattern underscores *LFY* sensitivity to thermal conditions, suggesting its fine-tuned role in environmentally modulated flowering. Moreover, the temperature-responsive cluster encompassed additional genes (*MIOX4*, *REM22*, *ATWIP5*, *RPP27*, *MAP4K9*, and *AT2G24760*), all paralleling *LFY*'s expression profile. Together, these findings emphasize the intricate regulatory network integrating developmental and environmental signals to orchestrate flowering time.

Figure 8 – Clustering of *LEAFY* (*LFY*) and Co-Expressed Genes in *Arabidopsis thaliana*



This figure presents the co-expression cluster of *LFY*, including ~20 genes such as *KANADI2*, *HMGB9*, *ATNUDT6*, *SUVR5*, *CVL1*, *CYP85A2*, and *REM1*. The cluster reflects coordinated regulatory activity during the floral transition and highlights genes associated with developmental patterning.

Source: Adapted from transcriptomic datasets of *Arabidopsis thaliana* (Klepikova et al., 2016)

Figure 9 - Ortholog of *LFY* Identified in the *Salvia hispanica* Dataset

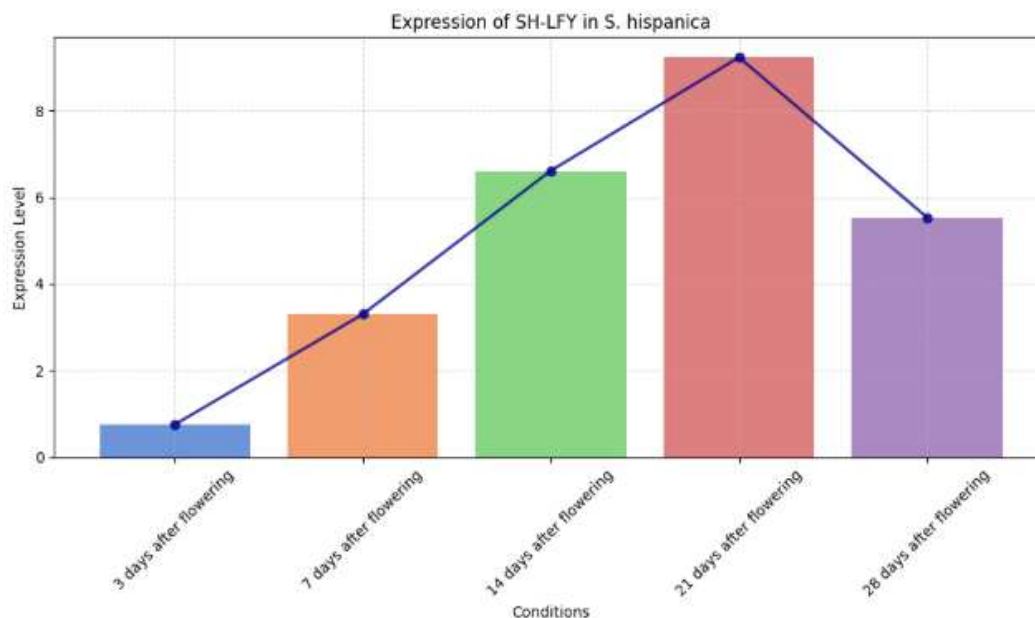
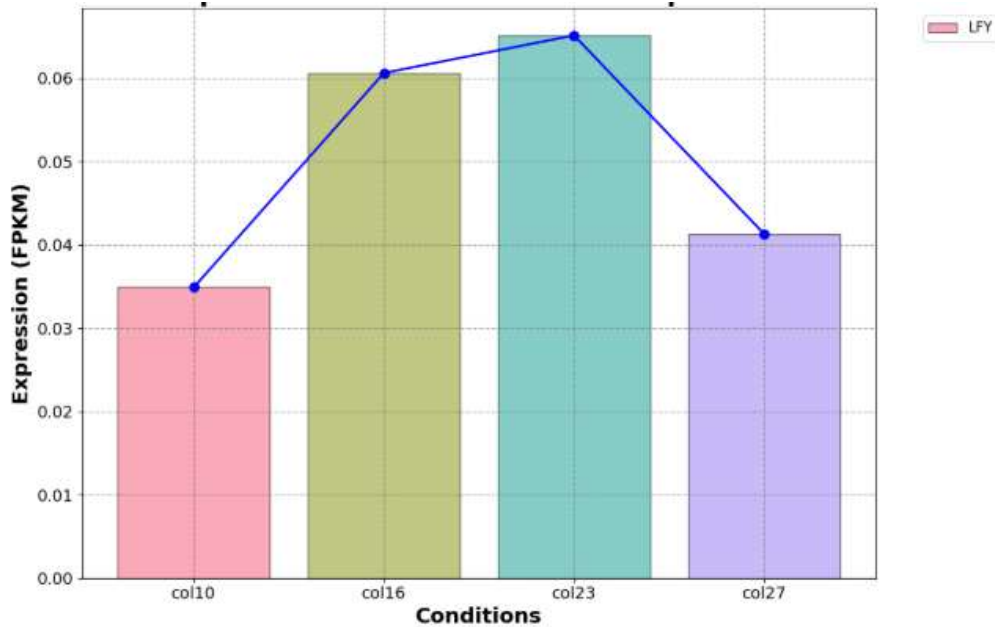


Figure 10 – Temperature-Dependent Expression of *LEAFY* (*LFY*) in *Arabidopsis thaliana*



This figure illustrates the expression dynamics of *LFY* under temperature-responsive datasets (Col10–Col27), with peaks at Col16 and Col23 and decline beyond Col23. The results demonstrate *LFY* sensitivity to thermal cues, reinforcing its role in environmentally modulated flowering induction.

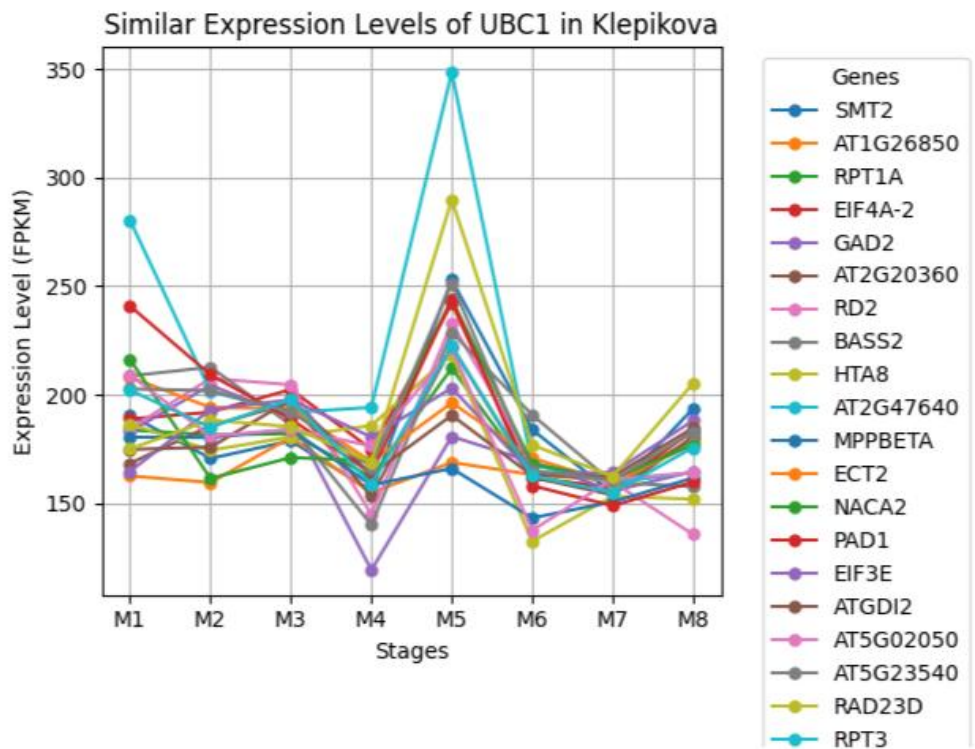
Source: Adapted from Li et al. (2018).

Role of Ubiquitin-Conjugating Enzyme (UBC) Family

The Ubiquitin-Conjugating Enzyme (UBC) family is central to the ubiquitination pathway, catalyzing the attachment of ubiquitin to target proteins and thereby regulating protein degradation and function (177,178). In *Arabidopsis thaliana*, multiple UBC isoforms show tissue-specific expression patterns, supporting roles in growth, development, and stress responses. For example, *UBC1* and *UBC2* are constitutively expressed across most tissues, while *UBC3* and *UBC4* display reproductive tissue specificity, indicating functional diversification (179). Clustering analysis centered on *UBC1* (**Figure 11**) revealed a co-expression module containing both characterized genes (*SMT2*, *RPT1A*, *EIF4A-2*, *GAD2*, *RPT3*) and uncharacterized loci (*AT1G26850*, *AT2G20360*, *AT2G47640*, *AT5G02050*, *AT5G23540*). These genes are likely involved in shared processes related to ubiquitination and protein turnover. Time-series profiles from Klepikova et al. (2016) demonstrated a decline in *UBC1* expression during later developmental stages, consistent with

reduced ubiquitination activity as reproduction is initiated. Temperature-dependent analysis further revealed an inverse pattern, with *UBC1* expression peaking at lower temperatures (Col10) and progressively diminishing at higher conditions, suggesting thermosensitive regulation of ubiquitination. Similarly, *UBC22* exhibited a dynamic and tightly regulated expression pattern in the Klepikova dataset, underscoring its role in early development and floral transition. *UBC22* was highly expressed during initial stages, declined gradually, then showed a pronounced peak coinciding with floral development, before diminishing during late stages. Clustering identified a co-expressed group of ~20 genes, including *RPT4B*, *FVE*, *ATPPC2*, *HSP60*, *SCL30*, *PRL1*, *HDI*, *BTR1*, *FH6*, and *ATHB5*, along with several unannotated loci. Temperature-dependent analysis revealed elevated *UBC22* expression under lower temperatures, implying roles in cold adaptation and stress response. Conversely, its expression declined under heat stress, suggesting a temperature-sensitive regulatory mechanism. The parallel profiles within this cluster indicate conserved regulatory pathways shared with *UBC22*.

Figure 11 - Similar Expression pattern of UBC1 at Klepikova RNA-seq Dataset



Transposable Elements and Flowering Transition

Although transposable elements (TEs) have traditionally received limited attention in flowering regulation, our bioinformatics and machine learning analyses revealed a compelling role for *AT5G28627* (Figure 13). Analysis of the Klepikova et al. (2016) dataset uncovered a dynamic bi-phasic expression profile for *AT5G28627*: elevated transcripts during early vegetative stages, followed by a decline, and a strong reactivation during floral transition (M3–M5). This expression coincided with key developmental milestones, suggesting involvement of the TE in modulating the shift from vegetative to reproductive growth.

Clustering analysis further identified co-expressed genes with similar trajectories (Figure 14), including *ATSTP9*, *ANAC025*, *ASK9*, *RFR1*, *ZIP9*, *LEA4-5*, and *MAP65-9*, alongside several uncharacterized loci (*ATGO2965*, *AT1G22110*, *AT1G51910*, *AT1G61840*, *AT3G24060*, *AT5G29560*). These findings point to novel candidate genes not yet annotated in TAIR, expanding the landscape of potential regulators. The stage-specific activation of *AT5G28627* during floral transition provides evidence for a previously underappreciated role of TEs in plant development, suggesting their integration into genetic and epigenetic regulatory networks orchestrating flowering

Figure 12- Similar Expression Pattern of *UBC22*

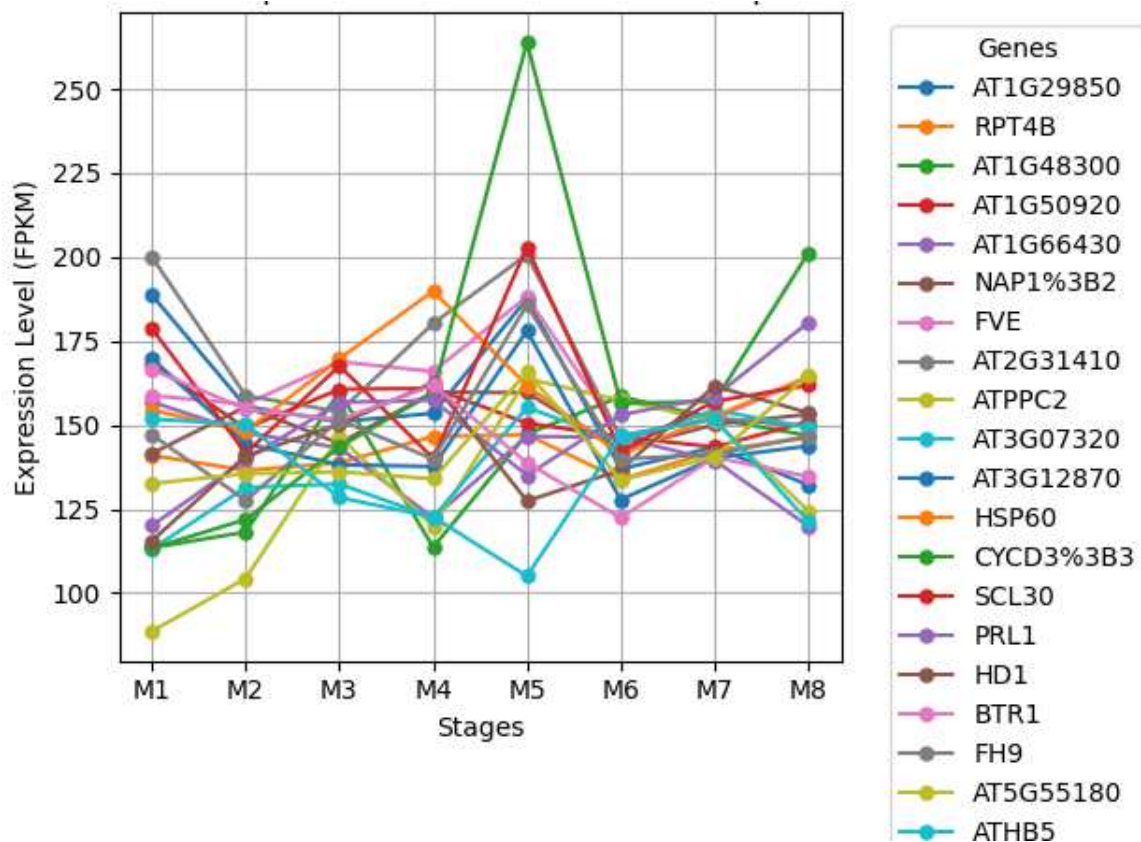


Figure 13 - Expression Profiles of the Selected Genes in the Klepikova RNA-Seq Dataset

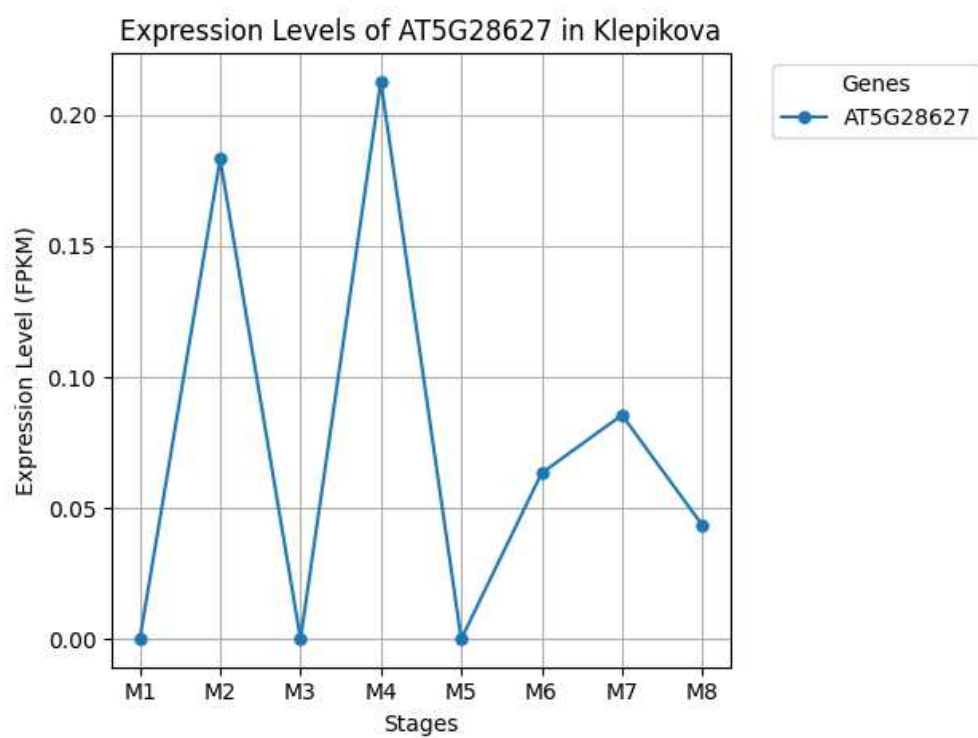


Figure 14 - Similar Expression Pattern pf TEs gene in Klepikova RNAseq Dataset

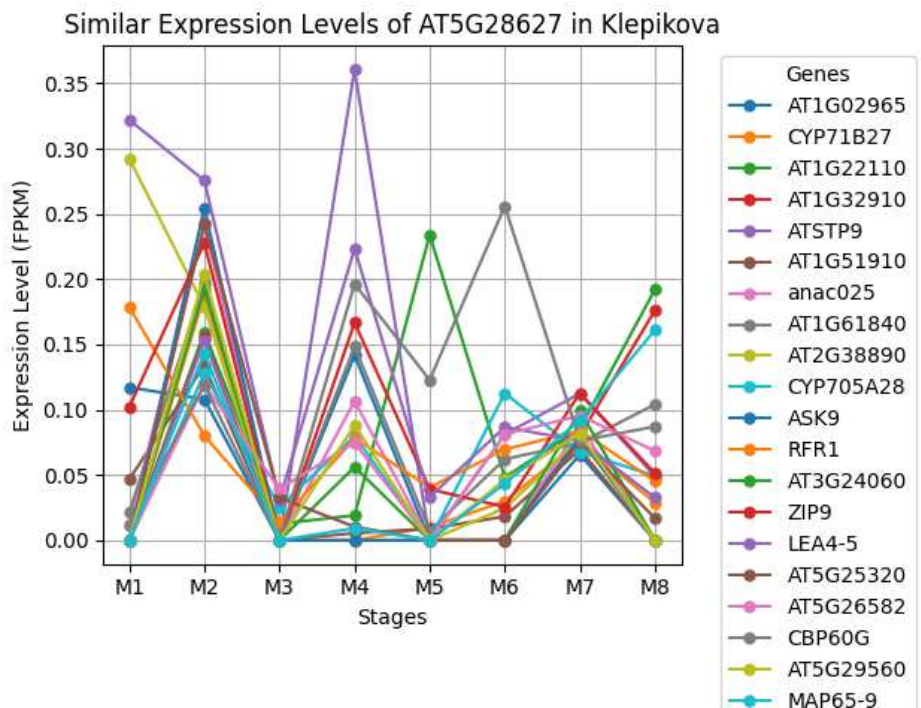


Figure 15 - Clustering analysis of AGL20/SCO1 at RNA-seq

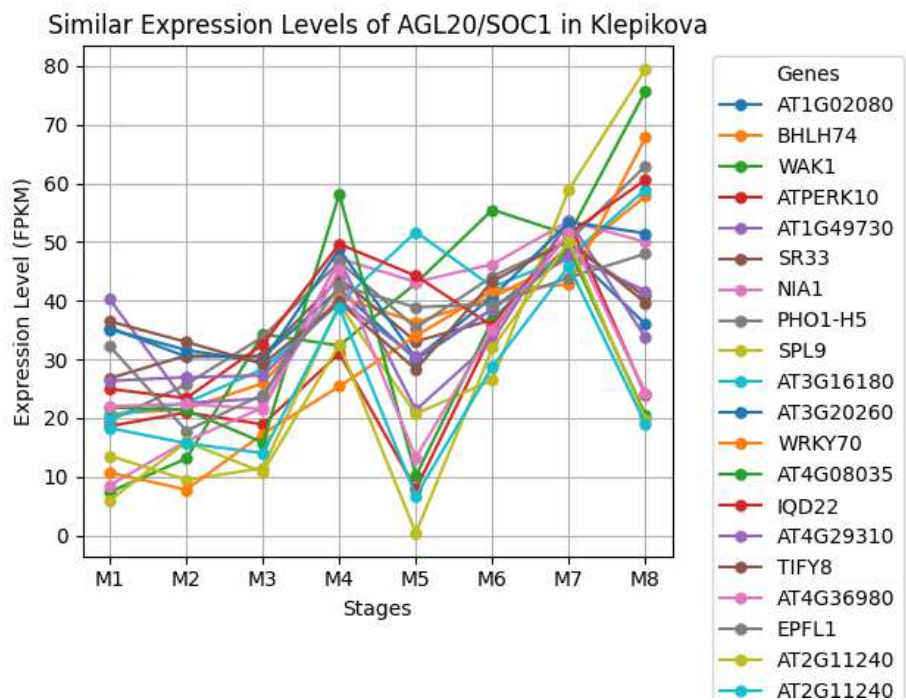


Figure 16 - Clustering analysis of SVP at RNA-seq Dataset of *Arabidopsis thaliana*

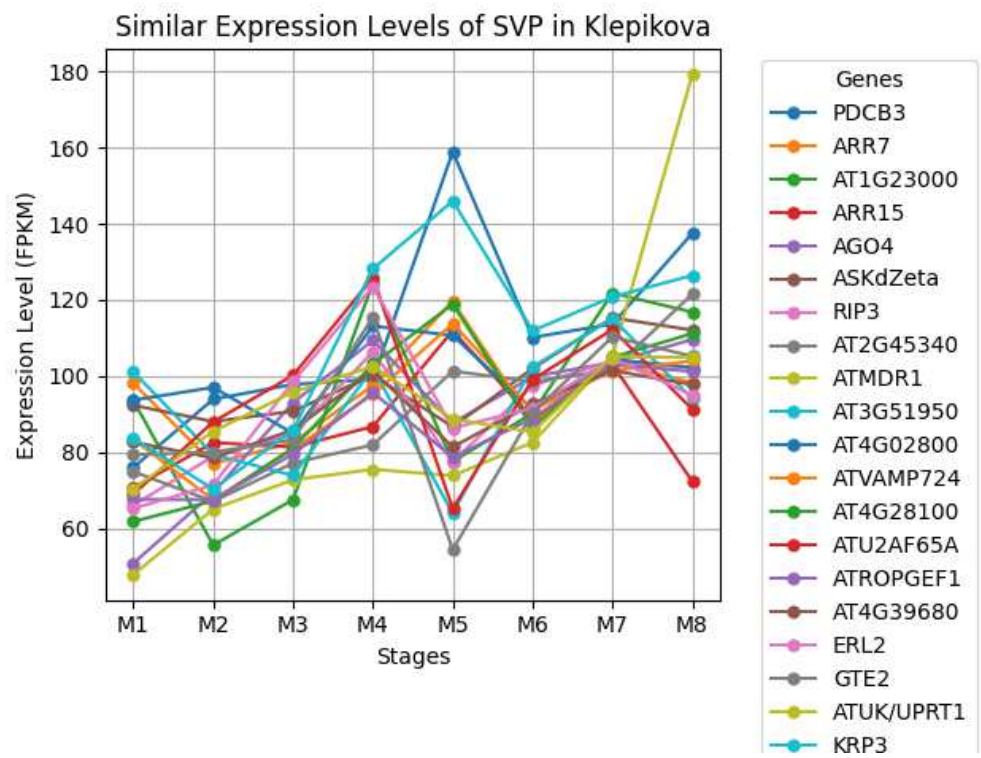


Figure 17 - Ortholog of SOC1 in *Salvia*

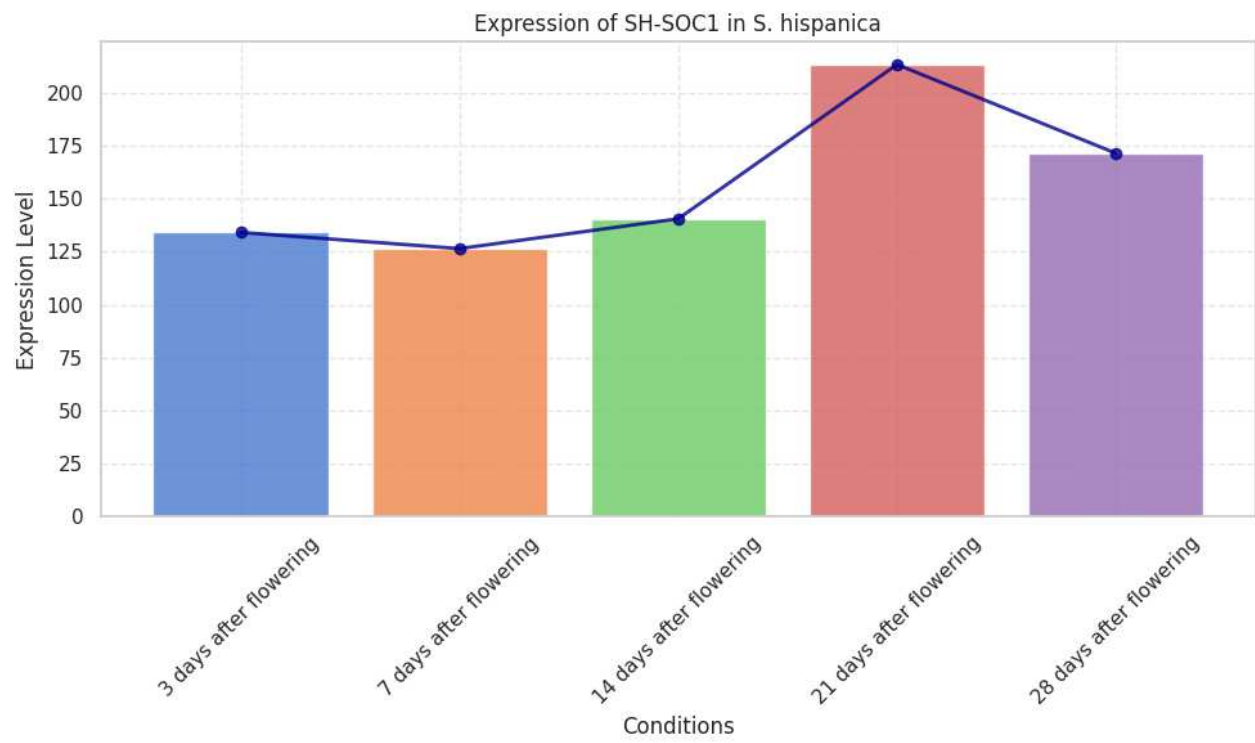
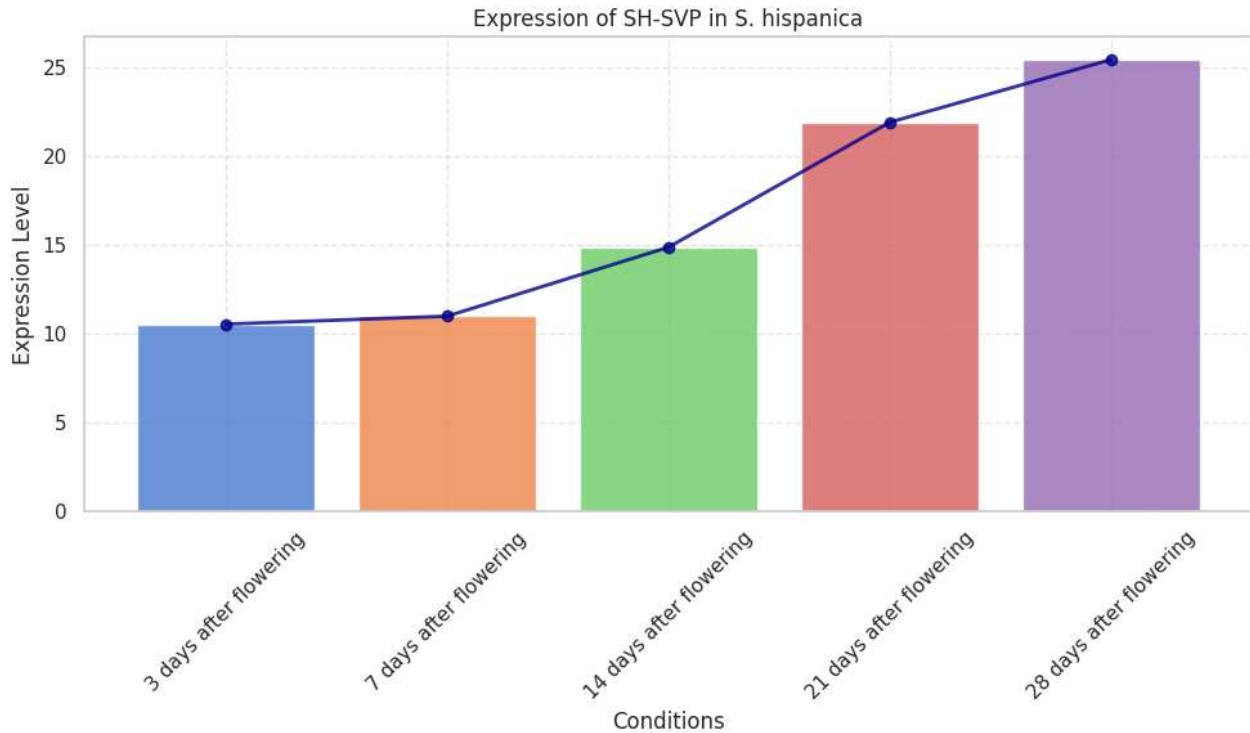


Figure 18- Ortholog of *SVP* in *Salvia hispanica*



DISCUSSION

A comprehensive analysis of the public RNA-seq dataset by Klepikova et al. revealed expression profiles for >21,000 genes across critical developmental stages in *Arabidopsis thaliana*, with the number of differentially expressed genes varying widely ranging from 14 (M1 vs. M2) to 4,890 (M7 vs. M8; Table 1) (Klepikova et al., 2015). Consistent with prior studies, FLC expression decreased markedly during the floral transition (Ietswaart et al., 2017), whereas LFY expression increased correspondingly, underscoring their well-established antagonistic roles in flowering regulation (Blázquez et al., 1998; Michaels & Amasino, 1999). In addition, SVP, a negative regulator of flowering, was highly expressed at early stages and increased over time (Hartmann et al., 2000). The floral integrator SOC1 was significantly induced from stages M1 to M4, in line with its function as an early marker of the flowering transition (Qi et al., 2022).

We used a suite of Python libraries (Pandas, NumPy, SciPy, scikit-learn, Seaborn, and Matplotlib) to perform robust gene-expression analyses across two time-series RNA-seq datasets from *A. thaliana* (Klepikova et al., 2015) and *Salvia hispanica* (Gupta et al., 2021). A key objective was to identify *S. hispanica* genes whose expression profiles show the highest Pearson correlation coefficients with target flowering regulators from *Arabidopsis*, leveraging machine-learning methodology to quantify these cross-species relationships (De Paola et al., 2023). This integrative computational approach not only validated conserved expression patterns but also enabled the prediction of putative flowering-time regulators in a non-model species, highlighting the value of combining transcriptomics with advanced computational tools (Blázquez et al., 1998; Michaels & Amasino, 1999).

The FT gene is a pivotal promoter of flowering, primarily responsive to changes in day length (Kardailsky et al., 1999). Structurally, FT contains a PEBP (phosphatidylethanolamine-binding protein) domain, extensively characterized in animals for roles in intracellular signaling mediated by protein–protein interactions (Kardailsky et al., 1999; Blázquez et al., 1998). In the Klepikova et al. dataset, FT expression increased sharply during the floral transition, particularly between stages M4 and M7 (Figure 2), corroborating its central role in flowering induction (Kardailsky et al., 1999). Our Python-based clustering identified several genes—CYP71B27, PHO1-H8, UGT76E6, CBF1, CBP6OG, and MAP65-9—with expression patterns closely paralleling FT during these stages (Figure 3), suggesting coordinated regulatory networks that facilitate floral development. Temperature profiling further showed that FT expression peaks around moderate temperature points (Col16–Col23) (Nasim et al., 2022) and declines beyond Col23, indicating temperature-dependent regulation consistent with thermal sensitivity of flowering pathways (Capovilla et al., 2015; Brightbill & Sung, 2022). Interestingly, a pseudogene in *Arabidopsis* clustered with FT and displayed a comparable expression profile, hinting at possible regulatory or evolutionary roles that merit investigation. Extending to *S. hispanica*, we identified a strong FT ortholog (XM_048087101.1-1; sh.FT) whose expression pattern was highly correlated with *Arabidopsis* FT (PCC = 0.9943), indicating remarkable conservation (Figure 4). sh.FT showed elevated expression ~7 days after flowering initiation, aligning with early floral development and suggesting a conserved mechanism in this non-model species.

The MADS-box transcription factors SOC1 (also AGL20) and SVP act as central integrators in the complex flowering network in *Arabidopsis*. SOC1 shows early and sustained activation as the vegetative meristem transitions to inflorescence, marking the developmental switch to flowering (Cheng et al., 2017; Qi et al., 2022). Clustering of SOC1 in our time-series dataset revealed a tightly co-expressed module including BHLH74, WAK1, SR33, NIA1, SPL9, WRKY70, IQD22, TIFY8,

and EPFL1 (Figure 15). The co-expression with known timing and hormone-signaling regulators (e.g., SPL9, WRKY70) reinforces the relevance of this module for floral induction (Shalom et al., 2015). In parallel, SVP, a key MADS-box repressor, cooperates with FLC to delay flowering by suppressing floral integrators (Hartmann et al., 2000; Mateos et al., 2015). SVP clustering identified a distinct co-expressed set— PDCB3, ARR7, ARR15, AGO4, RIP3, ATMDR1, ERL2, GTE2, KRP3 (Figure 16)—implicated in hormonal responses, chromatin remodeling, and cell-cycle control, highlighting SVP's multifaceted role in balancing development with environmental adaptation

Across species, orthologs of SOC1 and SVP were predicted in *S. hispanica* via comparative transcriptomics plus machine learning (Figures 17–18). The candidate sh.SOC1 (exon-XM_048120816.1-4; Figure 17) displayed a strikingly high correlation with *Arabidopsis* SOC1 (PCC = 0.9992), indicating near-identical temporal dynamics and suggesting strong functional conservation. Similarly, sh.SVP (exon-XM_048108698.1-2) showed a high correlation with *Arabidopsis* SVP, supporting evolutionary preservation of this flowering-repression module.

FLC, a MADS-box transcription factor, is a key negative regulator that represses floral promoters and integrates vernalization signals (Michaels & Amasino, 1999; Searle et al., 2006; Sung & Amasino, 2004). The transition to flowering is marked by a decrease in FLC alongside an increase in LFY. Analysis of Klepikova et al. confirms elevated FLC during early vegetative stages (M1–M3), followed by a progressive decline as flowering proceeds, whereas LFY increases during the vegetative-to-reproductive transition (Figure 2) (Blázquez et al., 1998). Spatial LFY expression in para-clade primordia aligns with sites of floral induction, consistent with its role in meristem identity (Blázquez et al., 1998). Genes co-expressed with FLC—including LSH6, SUC2, PIP2E, ERF098, TBL36, and AtGDU5 (Figure 5)—shared similar temporal profiles, suggesting involvement in the FLC-centered repression network. Temperature-dependent data showed that FLC decreases with rising temperature, aligning with thermo-responsive flowering control (Capovilla et al., 2015; Michaels & Amasino, 1999). In *S. hispanica*, sh.FLC (exon XM_048090739.1-3) was the top predictive ortholog, with PCC = 0.9986 (Figure 6), reinforcing evolutionary conservation of this floral repressor.

Protein ubiquitination is pivotal in plant development and stress responses, mediating protein turnover and signaling (Liu et al., 2020; Liao et al., 2024). Although the *Arabidopsis* genome encodes ~37 ubiquitin-conjugating enzymes (UBCs), many remain poorly characterized (Yao et al., 2023). In the Klepikova dataset, multiple UBC family members were elevated during the floral-transition window (M3–M6). UBC1 peaked between M4–M6 and declined with higher temperatures (Figure 11), suggesting temperature-sensitive regulation linked to flowering timing and stress adaptation. Clustering identified ~20 UBC1 co-expressed genes, including AT2G38810

(HTA8) (Figure 11), implicated in flower development. UBC22 also showed a pronounced peak during M4–M6 (Figure 12), further supporting a coordinated role of ubiquitination in flowering regulation.

Transposable elements (TEs) are increasingly recognized as drivers of plant evolution, domestication, and environmental adaptation, shaping genome structure and gene regulation (Hassan et al., 2024; Huang et al., 2025; Tao et al., 2025). In *Arabidopsis*, AT5G28627 showed elevated expression during M3–M5, coincident with the floral transition (Figure 13), and a peak at Col23, suggesting environmental responsiveness. Co-expression with CYP71B27, ATSTP9, ANAC025, ASK9, RFR1, ZIP9, and LEA4-5 (Figure 14) points to TE-associated regulatory roles in flowering time. Extending to *S. hispanica*, strong expression correlations between predicted orthologs—sh-AGO (Figure 15), sh-SVP (Figure 18), sh-SOC1 (Figure 17)—and their *Arabidopsis* counterparts reinforce functional conservation. While these computational predictions are compelling, experimental validation remains essential to confirm orthology and function in *S. hispanica*. Future work should prioritize functional genomics (e.g., gene knockouts, CRISPR, and expression modulation) to test candidate roles in floral transition. Deeper investigation into TEs such as AT5G28627 may reveal how genome plasticity contributes to developmental regulation and environmental adaptability.

CONCLUSIONS

This study demonstrates that comparative transcriptomics, coupled with advanced bioinformatics and machine-learning approaches, is a powerful strategy to investigate gene-expression dynamics in both the model species *Arabidopsis thaliana* and the non-model species *Salvia hispanica*. Unsupervised algorithms (e.g., K-means, DBSCAN) identified gene-expression patterns that reliably recapitulate known flowering regulators in *Arabidopsis*. Notably, we observed a pseudogene (AT2G24110) with an FT-like expression profile and a high-expression transposable element (AT5G28627) during floral-transition stages, underscoring the complexity and potential novelty of flowering networks. By integrating multiple machine-learning libraries, we predicted *S. hispanica* genes highly correlated with *Arabidopsis* counterparts, highlighting cross-species conservation and enabling gene-function inference in a non-model context. Collectively, our findings validate an integrative transcriptomics–machine learning framework to expand understanding of flowering regulation and open avenues for functional characterization and crop improvement.

REFERENCES

- 1000 GENOMES PROJECT CONSORTIUM. A map of human genome variation from population-scale sequencing. *Nature*, 2010. DOI: 10.1038/nature09534.
- ANDRÉS, F.; COUPLAND, G. The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics*, 2012. DOI: 10.1038/nrg3291.
- AYERZA, R.; COATES, W. Chia: Rediscovering a Forgotten Crop of the Aztecs. Tucson: University of Arizona Press, 2005.
- BERNIER, G. The control of floral evocation and morphogenesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, 1988. DOI: 10.1146/annurev.pp.39.060188.001135.
- BLÁZQUEZ, M. A. Flower development pathways. *Trends in Plant Science*, 2000. DOI: 10.1016/S1360-1385(00)01790-6.
- CAHILL, J. P. Ethnobotany of *Salvia hispanica* L. (Lamiaceae) in Mexico. *Economic Botany*, 2003. DOI: 10.1663/0013-0001(2003)057[0604:EOSHLL]2.0.CO;2.
- FORNARA, F.; COUPLAND, G. Plant phase transitions. *Current Opinion in Plant Biology*, 2009. DOI: 10.1016/j.pbi.2008.10.004.
- KOMEDA, Y. Genetic regulation of time to flower in *Arabidopsis thaliana*. *Annual Review of Plant Biology*, 2004. DOI: 10.1146/annurev.arplant.55.031903.141644.
- LANG, A. Physiology of flowering. In: *Encyclopedia of Plant Physiology*. Berlin: Springer, 1965. p. 1380–1536.
- PARK, D. H. et al. GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 1999. DOI: 10.1073/pnas.96.17.12162.
- SATO, S. et al. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 2012. DOI: 10.1038/nature11119.
- Abdul Aziz M, Masmoudi K. Molecular breakthroughs in modern plant breeding techniques. *Horticultural Plant Journal*. 2025;11(1):15–41. doi: 10.1016/j.hpj.2024.01.004
- Advance AI with Open Source. Anaconda. [accessed 3 Aug 2025] Available from: <https://www.anaconda.com/>
- Agarwal A, Rizwana, Tripathi AD, Kumar T, Sharma KP, Patel SKS. Nutritional and Functional New Perspectives and Potential Health Benefits of Quinoa and Chia Seeds. *Antioxidants*. 2023;12(7):1413. doi: 10.3390/antiox12071413

Alasalvar C, Chang SK, Bolling B, Oh WY, Shahidi F. Specialty seeds: Nutrients, bioactives, bioavailability, and health benefits: A comprehensive review. *Comprehensive Reviews in Food Science and Food Safety*. 2021;20(3):2382–427. doi: 10.1111/1541-4337.12730

Aldino AA, Darwis D, Prastowo AT, Sujana C. Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency. *J Phys: Conf Ser*. 2021;1751(1):012038. doi: 10.1088/1742-6596/1751/1/012038

Almoselhy RIM, Usmani A, Aftab Siddiqui M. Evolution of Chia Seeds from Antiquity to Modern Superfood. Rochester, NY: Social Science Research Network; 2024. [accessed 3 Aug 2025] Available from: <https://papers.ssrn.com/abstract=5072585>

Amasino R. Vernalization, Competence, and the Epigenetic Memory of Winter. *Plant Cell*. 2004;16(10):2553–9. doi: 10.1105/tpc.104.161070

Arraouadi S, Badri M, Abdelly C, Huguet T, Aouani ME. QTL mapping of physiological traits associated with salt tolerance in *Medicago truncatula* Recombinant Inbred Lines. *Genomics*. 2012;99(2):118–25. doi: 10.1016/j.ygeno.2011.11.005

Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A. Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics. *Life*. 2022;12(1):30. doi: 10.3390/life12010030

Ausín I, Alonso-Blanco C, Jarillo JA, Ruiz-García L, Martínez-Zapater JM. Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nat Genet*. 2004;36(2):162–6. doi: 10.1038/ng1295

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the ‘gold standard’ annotated reference plant genome. *Genesis*. 2015;53(8):474–85. doi: 10.1002/dvg.22877

Bernier G, Kinet J-M, Sachs RM. *The Physiology of Flowering: Volume II: Transition to Reproductive Growth*. Boca Raton: CRC Press; 2018. 241 p. doi: 10.1201/9781351075688

Biswas S, Islam F, Imran A, Zahoor T, Noreen R, Fatima M, et al. Phytochemical profile, nutritional composition, and therapeutic potentials of chia seeds: A concise review. *Cogent Food & Agriculture*. 2023;9(1):2220516. doi: 10.1080/23311932.2023.2220516

Bitá CE, Gerats T. Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Front Plant Sci*. 2013;4:273. doi: 10.3389/fpls.2013.00273

Blog R-S. RNA-seq QC and Data Analysis using the Tuxedo Suite | RNA-Seq Blog. 2015 June 10. [accessed 3 Aug 2025] Available from: <https://www.rna-seqblog.com/rna-seq-qc-and-data-analysis-using-the-tuxedo-suite/>

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. doi: 10.1093/bioinformatics/btu170

Bouché F, Lobet G, Tocquin P, Périlleux C. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res*. 2016;44(D1):D1167–71. doi: 10.1093/nar/gkv1054

Bowman JL, Smyth DR, Meyerowitz EM. Genes directing flower development in *Arabidopsis*. *Plant Cell*. 1989;1(1):37–52. doi: 10.1105/tpc.1.1.37

Brightbill CM, Sung S. Temperature-mediated regulation of flowering time in *Arabidopsis thaliana*. *aBIOTECH*. 2022;3(1):78–84. doi: 10.1007/s42994-022-00069-2

Buescher E, Achberger T, Amusan I, Giannini A, Ochsenfeld C, Rus A, et al. Natural Genetic Variation in Selected Populations of *Arabidopsis thaliana* Is Associated with Ionomics Differences. *PLOS ONE*. 2010;5(6):e11081. doi: 10.1371/journal.pone.0011081

Capovilla G, Schmid M, Posé D. Control of flowering by ambient temperature. *J Exp Bot*. 2015;66(1):59–69. doi: 10.1093/jxb/eru416

Carrillo-Barral N, Rodríguez-Gacio M del C, Matilla AJ. Delay of Germination-1 (DOG1): A Key to Understanding Seed Dormancy. *Plants*. 2020;9(4):480. doi: 10.3390/plants9040480

Castaigns L, Bergonzi S, Albani MC, Kemi U, Savolainen O, Coupland G. Evolutionary conservation of cold-induced antisense RNAs of FLOWERING LOCUS C in *Arabidopsis thaliana* perennial relatives. *Nat Commun*. 2014;5(1):4457. doi: 10.1038/ncomms5457

Castelán-Muñoz N, Herrera J, Cajero-Sánchez W, Arrizubieta M, Trejo C, García-Ponce B, et al. MADS-Box Genes Are Key Components of Genetic Regulatory Networks Involved in Abiotic Stress and Plastic Developmental Responses in Plants. *Front Plant Sci*. 2019;10. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2019.00853/full>

Chen M, Zhang J-Z. Flowering and Flower Development in Plants. *Agronomy*. 2024;14(2):256. doi: 10.3390/agronomy14020256

Chen Q, Zheng Y, Luo L, Yang Y, Hu X, Kong X. Functional FRIGIDA allele enhances drought tolerance by regulating the P5CS1 pathway in *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications*. 2018;495(1):1102–7. doi: 10.1016/j.bbrc.2017.11.149

Cheng J-Z, Zhou Y-P, Lv T-X, Xie C-P, Tian C-E. Research progress on the autonomous flowering time pathway in *Arabidopsis*. *Physiol Mol Biol Plants*. 2017;23(3):477–85. doi: 10.1007/s12298-017-0458-3

Chávez-Hernández EC, Quiroz S, García-Ponce B, Álvarez-Buylla ER. The flowering transition pathways converge into a complex gene regulatory network that underlies the phase changes of the shoot apical meristem in *Arabidopsis thaliana*. *Front Plant Sci*. 2022;13. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.852047/full>

Coolen S, Van Pelt JA, Van Wees SCM, Pieterse CMJ. Mining the natural genetic variation in *Arabidopsis thaliana* for adaptation to sequential abiotic and biotic stresses. *Planta*. 2019;249(4):1087–105. doi: 10.1007/s00425-018-3065-9

Corbesier L, Coupland G. Photoperiodic flowering of *Arabidopsis*: integrating genetic and physiological approaches to characterization of the floral stimulus. *Plant, Cell & Environment*. 2005;28(1):54–66. doi: 10.1111/j.1365-3040.2005.01283.x

Crossa J, Martini JWR, Vitale P, Pérez-Rodríguez P, Costa-Neto G, Fritsche-Neto R, et al. Expanding genomic prediction in plant breeding: harnessing big data, machine learning, and advanced software. *Trends in Plant Science*. 2025;30(7):756–74. doi: 10.1016/j.tplants.2024.12.009

Crossa J, Montesinos-Lopez OA, Costa-Neto G, Vitale P, Martini JWR, Runcie D, et al. Machine learning algorithms translate big data into predictive breeding accuracy. *Trends in Plant Science*. 2025;30(2):167–84. doi: 10.1016/j.tplants.2024.09.011

Csorba T, Questa JI, Sun Q, Dean C. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proceedings of the National Academy of Sciences*. 2014;111(45):16160–5. doi: 10.1073/pnas.1419030111

Cullerne DP, Fjellheim S, Spriggs A, Eamens AL, Trevaskis B, Wood CC. A Vernalization Response in a Winter Safflower (*Carthamus tinctorius*) Involves the Upregulation of Homologs of FT, FUL, and MAF. *Front Plant Sci*. 2021;12. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.639014/full>

Datta SD, Islam M, Rahman Sobuz MdH, Ahmed S, Kar M. Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review. *Heliyon*. 2024;10(5):e26888. doi: 10.1016/j.heliyon.2024.e26888

De Paola C, Garcia-Carpintero V, Vazquez-Vilar M, Kaminski K, Fernandez-del-Carmen A, Sierro N, et al. Comparative analysis of the Squamosa Promoter Binding-Like (SPL) gene family in *Nicotiana benthamiana* and *Nicotiana tabacum*. *Plant Science*. 2023;335:111797. doi: 10.1016/j.plantsci.2023.111797

Deng D. DBSCAN Clustering Algorithm Based on Density. In: 2020 7th International Forum on Electrical Engineering and Automation (IFEEA). 2020. p. 949–53. [accessed 3 Aug 2025] Available from: <https://ieeexplore.ieee.org/document/9356727>

Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES. FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. *Proc Natl Acad Sci U S A*. 2011;108(16):6680–5. doi: 10.1073/pnas.1103175108

Ding Y, Shi Y, Yang S. Molecular Regulation of Plant Responses to Environmental Temperatures. *Molecular Plant*. 2020;13(4):544–64. doi: 10.1016/j.molp.2020.02.004

Dong Q, Hu B, Zhang C. microRNAs and Their Roles in Plant Development. *Front Plant Sci*. 2022;13:824240. doi: 10.3389/fpls.2022.824240

Eom H, Park SJ, Kim MK, Kim H, Kang H, Lee I. TAF15b, involved in the autonomous pathway for flowering, represses transcription of FLOWERING LOCUS C. *The Plant Journal*. 2018;93(1):79–91. doi: 10.1111/tpj.13758

Fan M, Chen L, wang E, Xue X, Guo Q, Guo D, et al. Identification and characterization of flowering time regulatory gene FLC of *Paeonia ostii* ‘Fengdan’. *Scientia Horticulturae*. 2023;310:111748. doi: 10.1016/j.scienta.2022.111748

Force E, Lamy D, Debernard S, Sauré A, Dachet M. Developmental transitions involve common biological processes across living beings. *Heliyon*. 2025;11(6):e42995. doi: 10.1016/j.heliyon.2025.e42995

Franklin KA. Light and temperature signal crosstalk in plant development. *Current Opinion in Plant Biology*. 2009;12(1):63–8. doi: 10.1016/j.pbi.2008.09.007

Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol*. 2015;33(3):243–6. doi: 10.1038/nbt.3172

Freytes SN, Canelo M, Cerdán PD. Regulation of Flowering Time: When and Where? *Current Opinion in Plant Biology*. 2021;63:102049. doi: 10.1016/j.pbi.2021.102049

Geraldo N, Bäurle I, Kidou S, Hu X, Dean C. FRIGIDA Delays Flowering in Arabidopsis via a Cotranscriptional Mechanism Involving Direct Interaction with the Nuclear Cap-Binding Complex. *Plant Physiol*. 2009;150(3):1611–8. doi: 10.1104/pp.109.137448

González R, Butković A, Rivarez MPS, Elena SF. Natural variation in Arabidopsis thaliana rosette area unveils new genes involved in plant development. *Sci Rep*. 2020;10(1):17600. doi: 10.1038/s41598-020-74723-4

Goshisht MK. Machine Learning and Deep Learning in Synthetic Biology: Key Architectures, Applications, and Challenges. *ACS Omega*. 2024;9(9):9921–45. doi: 10.1021/acsomega.3c05913

Gray SB, Brady SM. Plant developmental responses to climate change. *Developmental Biology*. 2016;419(1):64–77. doi: 10.1016/j.ydbio.2016.07.023

Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet*. 2012;13(5):343–57. doi: 10.1038/nrg3173

Gregis V, Sessa A, Colombo L, Kater MM. AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in *Arabidopsis*. *The Plant Journal*. 2008;56(6):891–902. doi: 10.1111/j.1365-313X.2008.03648.x

Guo Y-L, Todesco M, Hagmann J, Das S, Weigel D. Independent FLC Mutations as Causes of Flowering-Time Variation in *Arabidopsis thaliana* and *Capsella rubella*. *Genetics*. 2012;192(2):729–39. doi: 10.1534/genetics.112.143958

Gupta P, Geniza M, Naithani S, Phillips JL, Haq E, Jaiswal P. Chia (*Salvia hispanica*) Gene Expression Atlas Elucidates Dynamic Spatio-Temporal Changes Associated With Plant Growth and Development. *Front Plant Sci*. 2021;12:667678. doi: 10.3389/fpls.2021.667678

H H, S W, Kj B. DELAY OF GERMINATION1 (DOG1) regulates both seed dormancy and flowering time through microRNA pathways. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(15). [accessed 2 Aug 2025] Available from: <https://pubmed.ncbi.nlm.nih.gov/27035986/>

Hagmann J. Characterizing short-term evolution of DNA methylation in *A. thaliana* using next-generation sequencing. 2017; [accessed 2 Aug 2025] Available from: <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/77512>

Hartmann U, Höhmann S, Nettekheim K, Wisman E, Saedler H, Huijser P. Molecular cloning of SVP: a negative regulator of the floral transition in *Arabidopsis*. *The Plant Journal*. 2000;21(4):351–60. doi: 10.1046/j.1365-313x.2000.00682.x

Hasanuzzaman M, Nahar K, Alam MdM, Roychowdhury R, Fujita M. Physiological, Biochemical, and Molecular Mechanisms of Heat Stress Tolerance in Plants. *Int J Mol Sci*. 2013;14(5):9643–84. doi: 10.3390/ijms14059643

Hassan AH, Mokhtar MM, El Allali A. Transposable elements: multifunctional players in the plant genome. *Front Plant Sci*. 2024;14:1330127. doi: 10.3389/fpls.2023.1330127

He Y, Chen T, Zeng X. Genetic and Epigenetic Understanding of the Seasonal Timing of Flowering. *Plant Communications*. 2020;1(1):100008. doi: 10.1016/j.xplc.2019.100008

Hollister JD. Genomic variation in *Arabidopsis*: tools and insights from next-generation sequencing. *Chromosome Res*. 2014;22(2):103–15. doi: 10.1007/s10577-014-9420-1

Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, et al. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. [accessed 2 Aug 2025] Available from: <https://dx.doi.org/10.1093/nar/29.1.102>

Huang Y, Sahu SK, Liu X. Deciphering recent transposition patterns in plants through comparison of 811 genome assemblies. *Plant Biotechnol J*. 2025;23(4):1121–32. doi: 10.1111/pbi.14570

Huo H, Wei S, Bradford KJ. DELAY OF GERMINATION1 (DOG1) regulates both seed dormancy and flowering time through microRNA pathways. *Proceedings of the National Academy of Sciences*. 2016;113(15):E2199–206. doi: 10.1073/pnas.1600558113

Ibañez C, Poeschl Y, Peterson T, Bellstädt J, Denk K, Gogol-Döring A, et al. Ambient temperature and genotype differentially affect developmental and phenotypic plasticity in *Arabidopsis thaliana*. *BMC Plant Biology*. 2017;17(1):114. doi: 10.1186/s12870-017-1068-5

Ietswaart R, Rosa S, Wu Z, Dean C, Howard M. Cell-Size-Dependent Transcription of FLC and Its Antisense Long Non-coding RNA COOLAIR Explain Cell-to-Cell Expression Variation. *Cell Systems*. 2017;4(6):622–635.e9. doi: 10.1016/j.cels.2017.05.010

Irwin JA, Lister C, Soumpourou E, Zhang Y, Howell EC, Teakle G, et al. Functional alleles of the flowering time regulator FRIGIDA in the Brassica oleraceagenome. *BMC Plant Biology*. 2012;12(1):21. doi: 10.1186/1471-2229-12-21

Janni M, Maestri E, Gulli M, Marmioli M, Marmioli N. Plant responses to climate change, how global warming may impact on food security: a critical review. *Front Plant Sci*. 2024;14. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1297569/full>

Jia J, Deng P, Li T, Wang K, Gao L, Zhao G, et al. Transposable elements drive the subgenomic divergence of homoeologous genes to shape wheat domestication and improvement. *Plant Communications*. 2025;6(6):101341. doi: 10.1016/j.xplc.2025.101341

Joshi HJ, Christiansen KM, Fitz J, Cao J, Lipzen A, Martin J, et al. 1001 Proteomes: a functional proteomics portal for the analysis of *Arabidopsis thaliana* accessions. *Bioinformatics*. 2012;28(10):1303–6. doi: 10.1093/bioinformatics/bts133

Jung H, Winefield C, Bombarely A, Prentis P, Waterhouse P. Tools and Strategies for Long-Read Sequencing and De Novo Assembly of Plant Genomes. *Trends in Plant Science*. 2019;24(8):700–24. doi: 10.1016/j.tplants.2019.05.003

Kang M, Choi Y, Kim H, Choi M-S, Lee S, Hyun Y, et al. Loss-of-function variants of CYP706A3 in two natural accessions of *Arabidopsis thaliana* increase floral sesquiterpene emission. *BMC Plant Biology*. 2025;25(1):275. doi: 10.1186/s12870-025-06283-6

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, et al. Activation Tagging of the Floral Inducer FT. *Science*. 1999;286(5446):1962–5. doi: 10.1126/science.286.5446.1962

Kaur S, Bains K. Chia (*Salvia hispanica* L.) – a rediscovered ancient grain, from Aztecs to food laboratories: A review. *Nutrition & Food Science*. 2019;50(3):463–79. doi: 10.1108/NFS-06-2019-0181

Kerwin R, Feusier J, Corwin J, Rubin M, Lin C, Muok A, et al. Natural genetic variation in *Arabidopsis thaliana* defense metabolism genes modulates field fitness. Kant MR, editor. *eLife*. 2015;4:e05604. doi: 10.7554/eLife.05604

Khalid W, Arshad MS, Aziz A, Rahim MA, Qaisrani TB, Afzal F, et al. Chia seeds (*Salvia hispanica* L.): A therapeutic weapon in metabolic disorders. *Food Sci Nutr*. 2022;11(1):3–16. doi: 10.1002/fsn3.3035

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15. doi: 10.1038/s41587-019-0201-4

Kim D-H, Sung S. Genetic and Epigenetic Mechanisms Underlying Vernalization. *Arabidopsis Book*. 2014;12:e0171. doi: 10.1199/tab.0171

Kisvarga S, Horotán K, Wani MA, Orlóci L. Plant Responses to Global Climate Change and Urbanization: Implications for Sustainable Urban Landscapes. *Horticulturae*. 2023;9(9):1051. doi: 10.3390/horticulturae9091051

Klepikova AV, Logacheva MD, Dmitriev SE, Penin AA. RNA-seq analysis of an apical meristem time series reveals a critical point in *Arabidopsis thaliana* flower initiation. *BMC Genomics*. 2015;16(1):466. doi: 10.1186/s12864-015-1688-9

Knez Hrnčič M, Ivanovski M, Cör D, Knez Ž. Chia Seeds (*Salvia Hispanica* L.): An Overview—Phytochemical Profile, Isolation Methods, and Application. *Molecules*. 2020;25(1):11. doi: 10.3390/molecules25010011

Kong X, Luo L, Zhao J, Chen Q, Chang G, Huang J, et al. Expression of FRIGIDA in root inhibits flowering in *Arabidopsis thaliana*. *J Exp Bot*. 2019;70(19):5101–14. doi: 10.1093/jxb/erz287

Koornneef M, Hanhart CJ, van der Veen JH. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Molec Gen Genet*. 1991;229(1):57–66. doi: 10.1007/BF00264213

Krivmane B, Girgžde E, Samsone I, Ruņģis D. Expression of juvenility related microRNAs and target genes during micropropagation of silver birch (*Betula pendula* Roth.). *Plant Cell Tiss Organ Cult*. 2023;152(3):455–69. doi: 10.1007/s11240-022-02419-w

Kulczyński B, Kobus-Cisowska J, Taczanowski M, Kmiecik D, Gramza-Michałowska A. The Chemical Composition and Nutritional Value of Chia Seeds—Current State of Knowledge. *Nutrients*. 2019;11(6):1242. doi: 10.3390/nu11061242

Kumar R, Das SP, Choudhury BU, Kumar A, Prakash NR, Verma R, et al. Advances in genomic tools for plant breeding: harnessing DNA molecular markers, genomic selection, and genome editing. *Biological Research*. 2024;57(1):80. doi: 10.1186/s40659-024-00562-6

Lahlali R, Taoussi M, Laasli S-E, Gachara G, Ezzouggari R, Belabess Z, et al. Effects of climate change on plant pathogens and host-pathogen interactions. *Crop and Environment*. 2024;3(3):159–70. doi: 10.1016/j.crope.2024.05.003

Lang A. Die photoperiodische Regulation von Förderung und Hemmung der Blütenbildung. *Berichte der Deutschen Botanischen Gesellschaft*. 1984;97(1):293–314. doi: 10.1111/j.1438-8677.1984.tb03470.x

Li D, Liu C, Shen L, Wu Y, Chen H, Robertson M, et al. A Repressor Complex Governs the Integration of Flowering Signals in Arabidopsis. *Developmental Cell*. 2008;15(1):110–20. doi: 10.1016/j.devcel.2008.05.002

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. doi: 10.1093/bioinformatics/btp352

Lian H, Wang L, Ma N, Zhou C-M, Han L, Zhang T-Q, et al. Redundant and specific roles of individual MIR172 genes in plant development. *PLoS Biol*. 2021;19(2):e3001044. doi: 10.1371/journal.pbio.3001044

Liao Y, Zhang W, Liu Y, Zhu C, Zou Z. The role of ubiquitination in health and disease. *MedComm* (2020). 2024;5(10):e736. doi: 10.1002/mco2.736

Lin S-I, Wang J-G, Poon S-Y, Su C, Wang S-S, Chiou T-J. Differential Regulation of FLOWERING LOCUS C Expression by Vernalization in Cabbage and Arabidopsis. *Plant Physiol*. 2005;137(3):1037–48. doi: 10.1104/pp.104.058974

Liu C, Xi W, Shen L, Tan C, Yu H. Regulation of Floral Patterning by Flowering Time Genes. *Developmental Cell*. 2009;16(5):711–22. doi: 10.1016/j.devcel.2009.03.011

Liu F, Quesada V, Crevillén P, Bäurle I, Swiezewski S, Dean C. The Arabidopsis RNA-Binding Protein FCA Requires a Lysine-Specific Demethylase 1 Homolog to Downregulate FLC. *Molecular Cell*. 2007;28(3):398–407. doi: 10.1016/j.molcel.2007.10.018

Liu M, Jiang J, Han Y, Shi M, Li X, Wang Y, et al. Functional Characterization of the Lysine-Specific Histone Demethylases Family in Soybean. *Plants (Basel)*. 2022;11(11):1398. doi: 10.3390/plants11111398

Liu W, Tang X, Qi X, Fu X, Ghimire S, Ma R, et al. The Ubiquitin Conjugating Enzyme: An Important Ubiquitin Transfer Platform in Ubiquitin-Proteasome System. *Int J Mol Sci.* 2020;21(8):2894. doi: 10.3390/ijms21082894

Lu Y, Meng Y, Zeng J, Luo Y, Feng Z, Bian L, et al. Coordination between GROWTH-REGULATING FACTOR1 and GRF-INTERACTING FACTOR1 plays a key role in regulating leaf growth in rice. *BMC Plant Biology.* 2020;20(1):200. doi: 10.1186/s12870-020-02417-0

Madrid E, Chandler JW, Coupland G. Gene regulatory networks controlled by FLOWERING LOCUS C that confer variation in seasonal flowering and life history. *J Exp Bot.* 2021;72(1):4–14. doi: 10.1093/jxb/eraa216

Mahood EH, Kruse LH, Moghe GD. Machine learning: A powerful tool for gene function prediction in plants. *Appl Plant Sci.* 2020;8(7):e11376. doi: 10.1002/aps3.11376

Maple R, Zhu P, Hepworth J, Wang J-W, Dean C. Flowering time: From physiology, through genetics to mechanism. *Plant Physiol.* 2024;195(1):190–212. doi: 10.1093/plphys/kiae109

Marquardt S, Raitskin O, Wu Z, Liu F, Sun Q, Dean C. Functional Consequences of Splicing of the Antisense Transcript COOLAIR on FLC Transcription. *Molecular Cell.* 2014;54(1):156–65. doi: 10.1016/j.molcel.2014.03.026

Matar S, Kumar A, Holtgräwe D, Weisshaar B, Melzer S. The transition to flowering in winter rapeseed during vernalization. *Plant, Cell & Environment.* 2021;44(2):506–18. doi: 10.1111/pce.13946

Mateos JL, Madrigal P, Tsuda K, Rawat V, Richter R, Romera-Branchat M, et al. Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. *Genome Biol.* 2015;16(1):31. doi: 10.1186/s13059-015-0597-1

Maćkiewicz A, Ratajczak W. Principal components analysis (PCA). *Computers & Geosciences.* 1993;19(3):303–42. doi: 10.1016/0098-3004(93)90090-R

Mbebi AJ, Mercado F, Hobby D, Tong H, Nikoloski Z. Advances in multi-trait genomic prediction approaches: classification, comparative analysis, and perspectives. *Brief Bioinform.* 2025;26(3):bbaf211. doi: 10.1093/bib/bbaf211

McClung CR, Lou P, Hermand V, Kim JA. The Importance of Ambient Temperature to Growth and the Induction of Flowering. *Front Plant Sci.* 2016;7. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2016.01266/full>

Meher PK, Mohapatra A, Satpathy S, Sharma A, Saini I, Pradhan SK, et al. PredCRG: A computational method for recognition of plant circadian genes by employing support vector machine with Laplace kernel. *Plant Methods.* 2021;17(1):46. doi: 10.1186/s13007-021-00744-3

Melo D, Machado TB, Oliveira MBPP. Chia seeds: an ancient grain trending in modern human diets. *Food Funct.* 2019;10(6):3068–89. doi: 10.1039/C9FO00239A

Michaels SD, Amasino RM. FLOWERING LOCUS C Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. *Plant Cell.* 1999;11(5):949–56. doi: 10.1105/tpc.11.5.949

Michaels SD, Amasino RM. Loss of FLOWERING LOCUS C Activity Eliminates the Late-Flowering Phenotype of FRIGIDA and Autonomous Pathway Mutations but Not Responsiveness to Vernalization. *Plant Cell.* 2001;13(4):935–42. doi: 10.1105/tpc.13.4.935

Mishra S, Spaccarotella K, Gido J, Samanta I, Chowdhary G. Effects of Heat Stress on Plant-Nutrient Relations: An Update on Nutrient Uptake, Transport, and Assimilation. *Int J Mol Sci.* 2023;24(21):15670. doi: 10.3390/ijms242115670

Mondal S, Karmakar S, Panda D, Pramanik K, Bose B, Singhal RK. Crucial plant processes under heat stress and tolerance through heat shock proteins. *Plant Stress.* 2023;10:100227. doi: 10.1016/j.stress.2023.100227

Mora-Poblete F, Maldonado C, Henrique L, Uhdre R, Scapim CA, Mangolim CA. Multi-trait and multi-environment genomic prediction for flowering traits in maize: a deep learning approach. *Front Plant Sci.* 2023;14. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1153040/full>

Mora-Poblete F, Maldonado C, Henrique L, Uhdre R, Scapim CA, Mangolim CA. Multi-trait and multi-environment genomic prediction for flowering traits in maize: a deep learning approach. *Front Plant Sci.* 2023;14:1153040. doi: 10.3389/fpls.2023.1153040

Motyka S, Koc K, Ekiert H, Blicharska E, Czarnek K, Szopa A. The Current State of Knowledge on *Salvia hispanica* and *Salviae hispanicae* semen (Chia Seeds). *Molecules.* 2022;27(4):1207. doi: 10.3390/molecules27041207

Motyka S, Skala E, Ekiert H, Szopa A. Health-promoting approaches of the use of chia seeds. *Journal of Functional Foods.* 2023;103:105480. doi: 10.1016/j.jff.2023.105480

Naik YD, Bahuguna RN, Garcia-Caparrós P, Zwart RS, Reddy MSS, Mir RR, et al. Exploring the multifaceted dynamics of flowering time regulation in field crops: Insight and intervention approaches. *The Plant Genome.* 2025;18(2):e70017. doi: 10.1002/tpg2.70017

Nasim Z, Susila H, Jin S, Youn G, Ahn JH. Polymerase II–Associated Factor 1 Complex-Regulated FLOWERING LOCUS C-Clade Genes Repress Flowering in Response to Chilling. *Front Plant Sci.* 2022;13. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.817356/full>

Niu F, Rehmani MS, Yan J. Multilayered regulation and implication of flowering time in plants. *Plant Physiology and Biochemistry.* 2024;213:108842. doi: 10.1016/j.plaphy.2024.108842

- Ortuño-Miquel S, Rodríguez-Cazorla E, Zavala-Gonzalez EA, Martínez-Laborda A, Vera A. *Arabidopsis* HUA ENHANCER 4 delays flowering by upregulating the MADS-box repressor genes FLC and MAF4. *Sci Rep*. 2019;9(1):1478. doi: 10.1038/s41598-018-38327-3
- Osnato M, Cota I, Nebhnani P, Cereijo U, Pelaz S. Photoperiod Control of Plant Growth: Flowering Time Genes Beyond Flowering. *Front Plant Sci*. 2022;12. [accessed 2 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.805635/full>
- Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D. Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res*. 2008;18(12):2024–33. doi: 10.1101/gr.080200.108
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650–67. doi: 10.1038/nprot.2016.095
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5. doi: 10.1038/nbt.3122
- Poethig RS. Vegetative phase change and shoot maturation in plants. *Curr Top Dev Biol*. 2013;105:125–52. doi: 10.1016/B978-0-12-396968-2.00005-1
- Poethig RS. Small RNAs and developmental timing in plants. *Curr Opin Genet Dev*. 2009;19(4):374–8. doi: 10.1016/j.gde.2009.06.001
- Pral W, Gregory BD. The fold makes all the difference in COOLAIR-mediated regulation of plant flowering time. *Developmental Cell*. 2022;57(19):2254–6. doi: 10.1016/j.devcel.2022.09.008
- Putterill J, Robson F, Lee K, Simon R, Coupland G. The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell*. 1995;80(6):847–57. doi: 10.1016/0092-8674(95)90288-0
- Qi P-L, Zhou H-R, Zhao Q-Q, Feng C, Ning Y-Q, Su Y-N, et al. Characterization of an autonomous pathway complex that promotes flowering in *Arabidopsis*. *Nucleic Acids Res*. 2022;50(13):7380–95. doi: 10.1093/nar/gkac551
- Raza A, Razzaq A, Mehmood SS, Zou X, Zhang X, Lv Y, et al. Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. *Plants*. 2019;8(2):34. doi: 10.3390/plants8020034
- Reddy INBL, Kim B-K, Yoon I-S, Kim K-H, Kwon T-R. Salt Tolerance in Rice: Focus on Mechanisms and Approaches. *Rice Science*. 2017;24(3):123–44. doi: 10.1016/j.rsci.2016.09.004
- Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT SCI*. 2021;2(3):160. doi: 10.1007/s42979-021-00592-x

Schönrock N, Bouveret R, Leroy O, Borghi L, Köhler C, Gruissem W, et al. Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway. *Genes Dev.* 2006;20(12):1667–78. doi: 10.1101/gad.377206

Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, et al. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes Dev.* 2006;20(7):898–912. doi: 10.1101/gad.373506

Shalom L, Shlizerman L, Zur N, Doron-Faigenboim A, Blumwald E, Sadka A. Molecular characterization of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) gene family from Citrus and the effect of fruit load on their expression. *Front Plant Sci.* 2015;6. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2015.00389/full>

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 2003;13(11):2498–504. doi: 10.1101/gr.1239303

Shi M, Wang C, Wang P, Yun F, Liu Z, Ye F, et al. Role of methylation in vernalization and photoperiod pathway: a potential flowering regulator? *Hortic Res.* 2023;10(10):uhad174. doi: 10.1093/hr/uhad174

Silva JCF, Teixeira RM, Silva FF, Brommonschenkel SH, Fontes EPB. Machine learning approaches and their current application in plant molecular biology: A systematic review. *Plant Science.* 2019;284:37–47. doi: 10.1016/j.plantsci.2019.03.020

Silva L de A, Verneque BJF, Mota APL, Duarte CK. Chia seed (*Salvia hispanica* L.) consumption and lipid profile: a systematic review and meta-analysis. *Food Funct.* 2021;12(19):8835–49. doi: 10.1039/D1FO01287H

Simpson GG. The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Current Opinion in Plant Biology.* 2004;7(5):570–4. doi: 10.1016/j.pbi.2004.07.002

Singh NK, Yadav M, Singh V, Padhiyar H, Kumar V, Bhatia SK, et al. Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. *Bioresource Technology.* 2023;369:128486. doi: 10.1016/j.biortech.2022.128486

Song YH, Ito S, Imaizumi T. Flowering time regulation: photoperiod- and temperature-sensing in leaves. *Trends Plant Sci.* 2013;18(10):575–83. doi: 10.1016/j.tplants.2013.05.003

Srikanth A, Schmid M. Regulation of flowering time: all roads lead to Rome. *Cell Mol Life Sci.* 2011;68(12):2013–37. doi: 10.1007/s00018-011-0673-y

Sun L, Nie T, Chen Y, Yin Z. From Floral Induction to Blooming: The Molecular Mysteries of Flowering in Woody Plants. *Int J Mol Sci.* 2022;23(18):10959. doi: 10.3390/ijms231810959

Sun Y, Li H, Wu J, Zhang K, Tang W, Cong L, et al. Genome-wide identification of growth-regulating factor transcription factor family related to leaf and stem development in alfalfa. *Front Plant Sci.* 2022;13:964604. doi: 10.3389/fpls.2022.964604

Sung S, Amasino RM. Vernalization and epigenetics: how plants remember winter. *Current Opinion in Plant Biology.* 2004;7(1):4–10. doi: 10.1016/j.pbi.2003.11.010

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, et al. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. [accessed 2 Aug 2025] Available from: <https://dx.doi.org/10.1093/nar/gkm965>

Takagi H, Hempton AK, Imaizumi T. Photoperiodic flowering in Arabidopsis: Multilayered regulatory mechanisms of CONSTANS and the florigen FLOWERING LOCUS T. *Plant Communications.* 2023;4(3):100552. doi: 10.1016/j.xplc.2023.100552

Tang M, Bai X, Niu L-J, Chai X, Chen M-S, Xu Z-F. miR172 Regulates both Vegetative and Reproductive Development in the Perennial Woody Plant *Jatropha curcas*. *Plant and Cell Physiology.* 2018;59(12):2549. doi: 10.1093/pcp/pcy175

Tao X-Y, Feng S-L, Yuan L, Li Y-J, Li X-J, Guan X-Y, et al. Harnessing transposable elements for plant functional genomics and genome engineering. *Trends in Plant Science.* 2025 Apr 15; [accessed 3 Aug 2025] Available from: <https://www.sciencedirect.com/science/article/pii/S1360138525000676>

The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature.* 2000;408(6814):796–815. doi: 10.1038/35048692

The molecular basis of heat stress responses in plants. *Molecular Plant.* 2023;16(10):1612–34. doi: 10.1016/j.molp.2023.09.013

Torti S, Fornara F, Vincent C, Andrés F, Nordström K, Göbel U, et al. Analysis of the Arabidopsis Shoot Meristem Transcriptome during Floral Transition Identifies Distinct Regulatory Patterns and a Leucine-Rich Repeat Protein That Promotes Flowering. *Plant Cell.* 2012;24(2):444–62. doi: 10.1105/tpc.111.092791

TUBEROSA R, SALVI S, SANGUINETI MC, LANDI P, MACCAFERRI M, CONTI S. Mapping QTLs Regulating Morpho-physiological Traits and Yield: Case Studies, Shortcomings and Perspectives in Drought-stressed Maize. *Ann Bot.* 2002;89(7):941–63. doi: 10.1093/aob/mcf134

Ullah R, Nadeem M, Khalique A, Imran M, Mehmood S, Javid A, et al. Nutritional and therapeutic perspectives of Chia (*Salvia hispanica* L.): a review. *J Food Sci Technol.* 2016;53(4):1750–8. doi: 10.1007/s13197-015-1967-0

- Vadapalli S, Abdelhalim H, Zeeshan S, Ahmed Z. Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine. *Brief Bioinform.* 2022;23(5):bbac191. doi: 10.1093/bib/bbac191
- van Leeuwen H, Kliebenstein DJ, West MAL, Kim K, van Poecke R, Katagiri F, et al. Natural Variation among *Arabidopsis thaliana* Accessions for Transcriptome Response to Exogenous Salicylic Acid. *Plant Cell.* 2007;19(7):2099–110. doi: 10.1105/tpc.107.050641
- Vollmeister E, Phokas A, Meyberg R, Böhm CV, Peter M, Kohnert E, et al. A DELAY OF GERMINATION 1 (DOG1)-like protein regulates spore germination in the moss *Physcomitrium patens*. *The Plant Journal.* 2024;117(3):909–23. doi: 10.1111/tpj.16537
- Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, et al. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. *Plant Cell.* 2017;29(8):1806–21. doi: 10.1105/tpc.17.00073
- Wang B, Yang X, Jia Y, Xu Y, Jia P, Dang N, et al. High-quality *Arabidopsis thaliana* Genome Assembly with Nanopore and HiFi Long Reads. *Genomics, Proteomics & Bioinformatics.* 2022;20(1):4–13. doi: 10.1016/j.gpb.2021.08.003
- Wang J, Wang Q, Gao J, Lei Y, Zhang J, Zou J, et al. Genetic regulatory pathways of plant flowering time affected by abiotic stress. *Plant Stress.* 2025;15:100747. doi: 10.1016/j.stress.2025.100747
- Wang Z, Li X, Xu J, Yang Z, Zhang Y. Effects of ambient temperature on flower initiation and flowering in saffron (*Crocus sativus* L.). *Scientia Horticulturae.* 2021;279:109859. doi: 10.1016/j.scienta.2020.109859
- Weigel D. Natural Variation in *Arabidopsis*: From Molecular Genetics to Ecological Genomics1[W][OA]. *Plant Physiol.* 2012;158(1):2–22. doi: 10.1104/pp.111.189845
- Weigel D, Mott R. The 1001 Genomes Project for *Arabidopsis thaliana*. *Genome Biol.* 2009;10(5):107. doi: 10.1186/gb-2009-10-5-107
- Wu Z, Fang X, Zhu D, Dean C. Autonomous Pathway: FLOWERING LOCUS C Repression through an Antisense-Mediated Chromatin-Silencing Mechanism1[CC-BY]. *Plant Physiol.* 2020;182(1):27–37. doi: 10.1104/pp.19.01009
- Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, et al. Artificial intelligence: A powerful paradigm for scientific research. *The Innovation.* 2021;2(4):100179. doi: 10.1016/j.xinn.2021.100179
- Xu Y, Zhang B, Ma N, Liu X, Qin M, Zhang Y, et al. Quantitative Trait Locus Mapping and Identification of Candidate Genes Controlling Flowering Time in *Brassica napus* L. *Front Plant Sci.* 2021;11:626205. doi: 10.3389/fpls.2020.626205

Yao S, Xie M, Hu M, Cui X, Wu H, Li X, et al. Genome-wide characterization of ubiquitin-conjugating enzyme gene family explores its genetic effects on the oil content and yield of *Brassica napus*. *Front Plant Sci.* 2023;14. [accessed 3 Aug 2025] Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1118339/full>

Yuan J, Cheng L, Wang Y, Zhang F. Genome-Wide Association Studies for Key Agronomic and Quality Traits in Potato (*Solanum tuberosum* L.). *Agronomy.* 2024;14(10):2214. doi: 10.3390/agronomy14102214

Zare T, Fournier-Level A, Ebert B, Roessner U. Chia (*Salvia hispanica* L.), a functional ‘superfood’: new insights into its botanical, genetic and nutraceutical characteristics. *Ann Bot.* 2024;134(5):725–46. doi: 10.1093/aob/mcae123

Zeller G, Clark RM, Schneeberger K, Bohlen A, Weigel D, Rättsch G. Detecting polymorphic regions in *Arabidopsis thaliana* with resequencing microarrays. *Genome Res.* 2008;18(6):918–29. doi: 10.1101/gr.070169.107

Zhang B, Tong Y, Luo K, Zhai Z, Liu X, Shi Z, et al. Identification of GROWTH-REGULATING FACTOR transcription factors in lettuce (*Lactuca sativa*) genome and functional analysis of LsaGRF5 in leaf size regulation. *BMC Plant Biology.* 2021;21(1):485. doi: 10.1186/s12870-021-03261-6

Zhang J, He S, Wang W, Chen F, Li Z, Zhang J, et al. FTGD: a machine learning method for flowering-time gene prediction. *T.* 2023;2(1). [accessed 3 Aug 2025] Available from: <https://www.maxapress.com/article/doi/10.48130/TP-2023-0023>

Zhang L, Jiménez-Gómez JM. Functional analysis of FRIGIDA using naturally occurring variation in *Arabidopsis thaliana*. *The Plant Journal.* 2020;103(1):154–65. doi: 10.1111/tpj.14716

Zhu Q-H, Helliwell CA. Regulation of flowering time and floral patterning by miR172. *J Exp Bot.* 2011;62(2):487–95. doi: 10.1093/jxb/erq295

ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 2000. DOI: 10.1038/35048692.

BERARDINI, T. Z. et al. The *Arabidopsis* information resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis*, 2015. DOI: 10.1002/dvg.22877.

WEIGEL, D.; MOTT, R. The 1001 Genomes Project for *Arabidopsis thaliana*. *Genome Biology*, 2009. DOI: 10.1186/gb-2009-10-5-107.

1001 GENOMES CONSORTIUM. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, 2016. DOI: 10.1016/j.cell.2016.05.063.

GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 2016. DOI: 10.1038/nrg.2016.49.

VAN DIJK, E. L. et al. The Third Revolution in Sequencing Technology. Trends in Genetics, 2018. DOI: 10.1016/j.tig.2018.05.008.

KOORNNEEF, M.; ALONSO-BLANCO, C.; VREUGDENHIL, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. Annual Review of Plant Biology, 2004. DOI: 10.1146/annurev.arplant.55.031903.141605.

ALONSO-BLANCO, C.; KOORNNEEF, M. Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. Trends in Plant Science, 2000. DOI: 10.1016/S1360-1385(99)01510-1.

JOHANSON, U. et al. Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. Science, 2000. DOI: 10.1126/science.290.5490.344.

BENTSINK, L. et al. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. Proceedings of the National Academy of Sciences of the USA, 2006. DOI: 10.1073/pnas.0607877103.

LIBBRECHT, M. W.; NOBLE, W. S. Machine learning applications in genetics and genomics. Nature Reviews Genetics, 2015. DOI: 10.1038/nrg3920.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 2011;12:2825–2830.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967.

PARCY, F. Flowering: a time for integration. The International Journal of Developmental Biology, 2005. DOI: 10.1387/ijdb.041930fp.

KLEPIKOVA, A. V. et al. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. The Plant Journal, 2016. DOI: 10.1111/tpj.13312.

IXTAINA, V. Y. et al. Characterization of chia seed oils obtained by pressing and solvent extraction. Journal of Food Composition and Analysis, 2011. DOI: 10.1016/j.jfca.2010.08.006.

MARTÍNEZ-CRUZ, O.; PAREDES-LÓPEZ, O. Phytochemical profile and nutraceutical potential of chia seeds (*Salvia hispanica* L.) by ultra high performance liquid chromatography. Journal of Chromatography A, 2014. DOI: 10.1016/j.chroma.2014.04.007.

HRNČIČ, M. K. et al. Chia Seeds (*Salvia hispanica* L.): An Overview—Phytochemical Profile, Isolation Methods, and Application. Molecules, 2019. DOI: 10.3390/molecules25010011.

QIAGEN DIGITAL INSIGHTS. CLC Genomics Workbench (version 6.5.1). Aarhus: QIAGEN, 2013.