Biological Invasions (2025)

Molecular evolutionary analysis of genes potentially associated with invasiveness in two species of the genus *Thunbergia* (Acanthaceae)

Santiago Ruiz-Calero[1,2*], Harold Suárez-Baron[1,2], Pablo A. Pérez-Mesa[3, 4]

[1] Programa de Biología, Departamento de Ciencias Naturales y Matemáticas, Pontificia Universidad Javeriana Cali, Colombia

[2] Semillero Eco-Evo-Devo, Facultad de Ingeniería y Ciencias, Pontificia Universidad Javeriana Cali, Colombia

[3] Max Planck Tandem Group, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá

[4] Institute for Bio- and Geosciences (IBG-4: Bioinformatics), CEPLAS, Forschungszentrum Jülich, Wilhelm Johnen Straße, Jülich, Germany

*Correspondence: santiagoruizc@javerianacali.edu.co

Abstract

Invasive species present a critical threat to global biodiversity and ecosystems, with significant economic impacts and human health risks. In Colombia, *Thunbergia alata* and *Thunbergia grandiflora* (Acanthaceae) are introduced species that exhibit contrasting patterns of invasiveness. *T. alata* has rapidly expanded in Andean forests and highly disturbed areas, whereas *T. grandiflora* shows no evident invasive behavior in the country. This study aimed to analyze molecular evolutionary patterns of genes potentially associated with invasiveness by identifying selection signals in genes involved in the adaptation of both species to new environment. We selected 32 candidate genes through differential expression analysis, gene ontology annotations, and literature review, focusing on processes such as development, reproduction, metabolism, and stress response. We conducted a gene co-expression analysis which revealed functionally enriched modules exclusively in *T. alata*. In parallel, we carried out a positive selection analysis based on codon substitution models identifying specific amino acid sites under selective pressure in five genes (*POK*, *ANJ1*, *ACLB1*, *NAC92*, and *DEGP1*), associated with vesicle trafficking, protein folding, metabolism, and stress tolerance. These findings suggest that positive selection on specific sites may be linked to mechanisms of gene neofunctionalization or subfunctionalization, enhancing adaptive capacity and invasive potential. The comparative analysis between both species contribute to a better understanding of the genetic mechanisms underlying plant invasiveness. However, further functional validation of these genes and also an exploration of alternative molecular pathways are essential to fully elucidate the genetic basis driving adaptive and invasive traits in invasive species.

Keywords

Invasive species   *Thunbergia alata*   *Thunbergia grandiflora*   positive selection   co-expression

Acronyms

BEB: Bayes Empirical Bayes
BI: Bayesian Inference
CDS: Coding DNA sequence
DEG: Differentially Expressed Genes
DigiCGA: Digital Candidate Gene Approach
ER: Early Reproductive

EV: Early Vegetative
eQTL: Expression Quantitative Trait Loci
FDR: False Discovery Rate
GTR+I+G: General Time Reversible model + Invariant sites + Gamma-distributed rate variation
GO: Gene Ontology
Ka or *dN*: Non-synonymous Substitution Rate
Ks or *dS*: Synonymous Substitution Rate
logFC: Log Fold Change
LRT: Likelihood Ratio Test
LR: Late Reproductive
LV: Late Vegetative
M: Mixed tissues
ME: Module Eigengene
ML: Maximum Likelihood
n/c: Not Calculated/Not Classified
NJ: Neighbor-Joining
ORF: Open Reading Frame
R: Reproductive
RNA-seq: RNA sequencing
V: Vegetative
WGCNA: Weighted Gene Co-expression Network Analysis
ω: *Ka/Ks* or *dN/dS* Ratio (Omega)
(2): Following a gene name indicates a gene copy
2*|ΔlnL|: Twice the Absolute Value of the Change in Log-Likelihood
ΔlnL: Delta Log-Likelihood (Change in Log-Likelihood)

Introduction

Biological invasions result from human's movement of plants, animals, and other organisms across the globe (Di Castri 1989). These species are intentionally or accidentally introduced (Meyerson & Money 2007). Species are taken from their native habitats, where their populations are regulated by a specific set of factors, and transported beyond their natural dispersal barriers. Upon arrival in new areas, their abundance and distribution are influenced by different biotic and abiotic conditions (Hierro et al. 2005). In recent years, globalization has significantly increased international trade, technological exchange, and tourism, which in turn has facilitated the spread of invasive species in non-native ecosystems. This has altered natural biogeographical barriers, creating negative impacts on both the environment and the economy, as well as posing a threat to human health (Meyerson & Money 2007).

In this context, a biological invasion can be defined as the success of a species in establishing, developing, and maintaining populations outside its native range (Theoharides & Dukes 2007; Marin et al. 2020). The success of such an invasion depends on factors like the species' establishment success, the connectivity of available zones for colonization and expansion, its mode and pattern of reproduction and dispersal, and the competitive relationships it develops with other species (Hellman et al. 2008) (Fig. 1).

A species that invades a new region must possess high physiological tolerance and plasticity, which allows it to disperse, establish, and naturalize in geographical areas different from its native range. Otherwise, it must undergo genetic differentiation (either through adaptive phenotypic plasticity or adaptive evolution) to reach the necessary levels of adaptation for the new region (Pyšek & Richardson 2010; Marin et al. 2020). During its initial establishment, and throughout its expansion and colonization of new areas, these species can evolve in response to the pressures of

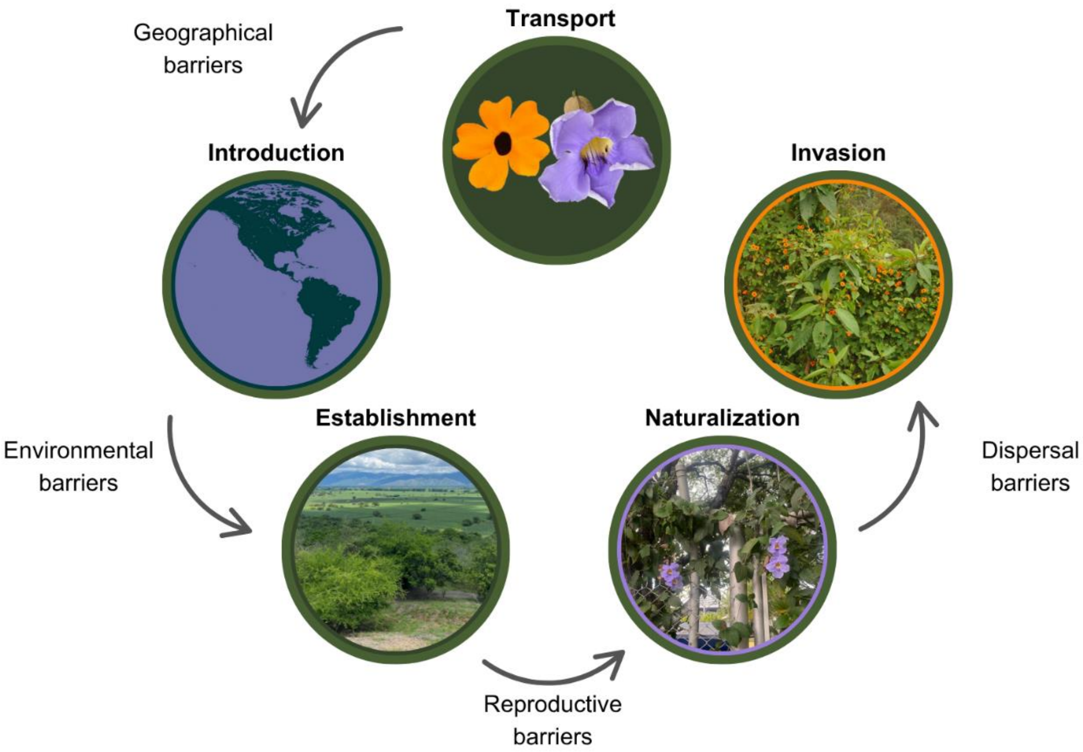

Fig. 1 Plant invasion model based on *Thunbergia* species. Plant invasion in an exogenous ecosystem typically involves several stages through different biotic and abiotic barriers: transport, introduction, establishment, naturalization, and ecosystem integration by invasion. Initially, the plant is transported and introduced to a new environment, where it adapts to local conditions and begins to reproduce. Over time, it spreads to new areas, competing with native species for resources. As the plant establishes dominance, it may alter habitat structures and disrupt ecological relationships, leading to a decline in biodiversity. Factors such as high reproductive rates, stress resistance, and lack of natural predators facilitate the plant's spread. Ultimately, the ecosystem may adapt to the presence of invasive species, but the balance of the environment is often significantly changed.

the new environment (Sakai et al. 2021). The subsequent diversification that occurs during colonization promotes rapid speciation, mainly through selection processes driven by factors such as hybridization, genetic variation, polyploidy, the role of specific genes, and environmental gradients. These factors are crucial to understanding the evolution of invasiveness (Lee 2002; Schierenbeck & Ainouche 2006; Prentis et al. 2008). These evolutionary processes and genetic mechanisms are key to studying accelerated evolution and rapid adaptation in invasive species (Sakai et al. 2021).

With the onset of climate change, significant alterations have occurred in phenology, distribution ranges, species interactions, ecosystem processes, and the genetic composition of organisms (Hellman et al. 2008). These changes may create more favorable conditions for the establishment of invasive species that exhibit high physiological tolerance, phenotypic plasticity, and broad distribution ranges (Stachowicz et al. 2002). Moreover, it is important to consider that climatic factors impose strong selection pressures on the evolution of traits in plants. This can lead invasive species to develop greater climatic tolerance in the invaded environments than in their regions of origin (Clements & Ditomasso 2011). As a result, even species previously considered non-invasive may become invasive in response to climate change (McNeely 2001; Hellman et al. 2008). Today, invasive species have significantly impacted biodiversity and ecosystems worldwide, affecting their structure and composition and often leading to the exclusion or displacement of native species (Simberloff et al. 2013).

A large portion of research on plant invasion focuses on understanding the stages of transport, establishment, and dispersal of species. Most studies emphasize physiological and ecological aspects to develop management strategies aimed at mitigating the impact of invasive species on native populations (North et al. 2021). The development of sequencing technologies and their decreasing costs have made these tools essential for tracking, managing, and reducing the effects of invasive species (Chown et al. 2015; Hamelin & Roe 2020; Sherpa & Després 2021). These technologies enable the identification of the genetic factors and mechanisms that underlie a successful biological invasion by comparing invasive species with their non-invasive counterparts (Cai et al. 2020).

Among invasive plant species, *Thunbergia alata* Bojer ex Sims, commonly known as the Black-eyed Susan or Clock vine, is a climbing plant native to the island of Mauritius in East Africa. This species has been introduced to various tropical regions around the world, where it has become a highly prolific invasive plant (Quijano-Abril et al. 2021). Its distribution spans across Latin America and the Caribbean, from Argentina to Guatemala, as well as some areas of the Greater and Lesser Antilles and the Bahamas (Quijano-Abril et al. 2021) (Fig. 2.A). In Colombia, *T. alata* has adapted particularly well to the edges of Andean forest fragments and areas with high anthropogenic disturbance (CAR 2020). The species propagates through seeds and cuttings, and its invasive growth, along with its dispersal mechanisms, allows it to climb large trees and cover extensive areas of vegetation in Andean forests (Quijano-Abril et al. 2019). The fruits of *T. alata* are dehiscent, ranging from 1.5 to 2.5 cm in size, and consist of valves that facilitate seed release. During dehiscence, the seeds are launched distances of 2 to 3 meters, allowing for dispersal without the need for pollinators or natural dispersal agents (Quijano-Abril et al. 2019). Due to its rapid growth and propagation ability, *T. alata* is considered a high-risk invasive species in Colombia, prompting several control and management initiatives (Cárdenas et al. 2010; Cárdenas et al. 2017). Despite the research and control efforts made so far, little attention has been given to analyzing the genetic component of the species or identifying the possible molecular bases responsible for its successful adaptation to various ecosystems in Colombia. Understanding the genetic, molecular, and evolutionary mechanisms underlying the traits of invasiveness can help develop more effective strategies for the control and management of this invasive species.

Sky vine (*Thunbergia grandiflora* Roxb), also known as blue thunbergia or Bengal clock vine, is a perennial, semi-woody, robust, and very vigorous vine. It belongs to the Acanthaceae family and is native to India, southern China, and Myanmar (Mela et al. 2022). *T. grandiflora* is a widely distributed dicotyledonous plant in tropical and subtropical regions (Yuan et al. 2021) (Fig. 2.B), exhibiting a similar woody climbing habit as *T. alata* (Adhikari et al. 2013). Despite similarities in morphology and habit, they exhibit significant ecological differences, such as niche requirements, adaptability, and invasive potential. These differences, along with their phylogenetic proximity, provide an opportunity for comparative transcriptomic studies to better understand the genetic mechanisms underlying the extreme invasiveness observed in *T. alata* and, potentially, in other invasive plant species.

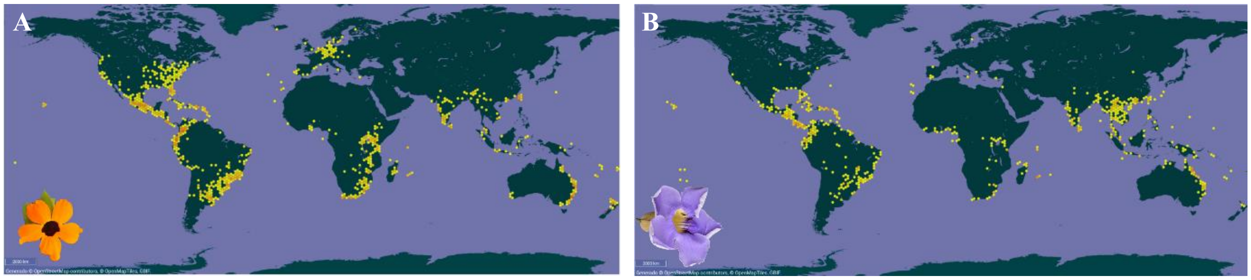

Fig. 2 Geographic distribution of the species included in this study A. Geographic areas where *T. alata* has successfully established populations. **B.** Geographic areas where *T. grandiflora* has successfully established populations. Both sets of information are based on data retrieved from (https://www.gbif.org/es/).

In recent years, the rapid accumulation of large transcriptomic datasets, particularly from high-throughput RNA sequencing in both model and non-model plants, has created a growing need for researchers to analyze and interpret this vast amount of data to better understand biological processes at the systems level. Co-expression network analysis is a highly effective method for interpreting large transcriptomic datasets. It allows researchers to identify modules of co-expressed genes, often functionally related (Serin et al. 2016; Rao & Dixon 2019). These networks provide an initial framework for exploring functional associations based on gene expression profiles and can be applied across various areas of plant biology (Rao & Dixon 2019). In this context, co-expression networks are powerful tools in plant molecular evolution, helping interpret functional modules, detect key genes involved in adaptive pathways, and explore gene plasticity in response to environmental changes. When applied to comparative analyses, these networks provide a valuable approach for generating hypotheses and gaining insights into the evolution of biological processes and gene interactions across multiple species (Ovens et al. 2021). Their application is particularly relevant in the study of invasive species, as this tool could reveal genes or gene clusters potentially involved in rapid adaptation, expansion, and invasion.

Similarly, many comparative transcriptomic studies focus on identifying the genetic mechanisms underlying the traits of invasiveness observed in invasive species. A commonly employed approach is to identify genes that have undergone positive selection. Positive selection refers to the process by which advantageous gene variants become prevalent in a population or species over time due to their contribution to increased fitness. This plays a key role in evolution by driving the fixation of new traits and adaptation to environmental changes (Sahm et al. 2017). Since synonymous substitutions do not result in amino acid changes due to the redundancy of the genetic code, they are generally considered neutral and unaffected by selection. As a result, they are commonly used as a reference to estimate the neutral rate of evolution (Zhang 2022). Therefore, the ratio of non-synonymous substitution rate (*Ka* or *dN*) to synonymous substitution rate (*Ks* or *dS*), known as $\omega = Ka/Ks$ (or dN/dS), is widely used as an indicator of positive selection. This ratio also helps distinguish between neutral mutations ($\omega \approx 1$), negative (purifying) selection ($\omega < 1$), and positive (adaptive) selection ($\omega > 1$) (Álvarez-Carretero et al. 2023). This ratio serves as a valuable tool for exploring the molecular evolution of coding sequences (Zhang 2022). For the selection analyses, we employed a software tool based on the comparison of specific codon substitution models to detect signatures of positive selection. Specifically, we applied both site-specific and branch-site models. Site-specific models detect positive selection by allowing variation in the ω (dN/dS) ratio across codon sites, while branch-site models extend this by also allowing variation across lineages to identify selection on specific codons within defined evolutionary branches (Gao et al. 2019).

In this study, we used the transcriptomic dataset that Ruiz-Londoño and collaborators (2024) generated to identify and propose key candidate genes underlying putative traits of invasiveness and study their molecular evolution by detecting selection signals in genes potentially involved in the adaptation of *T. alata* and *T. grandiflora* to new environments. We hypothesize that *T. alata*, having successfully colonized new environments, will exhibit signals of positive selection in genes related to reproduction, metabolism, and stress tolerance, in contrast to *T. grandiflora*, which hasn't shown invasive behavior. We used different approaches to identify a list of candidate genes by using 1) a targeted search for genes associated with invasiveness according to the literature, 2) differential expression analysis based on *T. alata* and *T. grandiflora* transcriptomes, and 3) weighted gene co-expression network analysis (WGCNA) of the RNA-sequencing (RNA-seq) data for *T. alata* and *T. grandiflora*. Overall, we identified 32 candidate genes based on the previous methodologies mentioned to perform phylogenetic and selection analyses. Our findings reveal representative specific sites under positive selection in *T. alata POKY POLLEN TUBE* (*POK*) and *DnaJ protein homolog* (*ANJ1*) genes, while for *T. grandiflora* we found specific sites under positive selection in *ATP-citrate lyase B-1* (*ACLB1*) and *NAC domain-containing protein 92* (*NAC92*) genes. Meanwhile, in both species we found specific sites under positive selection in *DegP protease 1* (*DEGP1*) gene and one of its copies. These genes are associated with key biological functions, including development and reproduction *(POK*, *NAC92*), metabolism (*ACLB1*), and stress response and development respectively (*ANJ1*, *DEGP1*) suggesting their potential relevance to adaptive and invasive processes. These findings in these genes suggest that positive selection in specific amino acids could be associated with functional mechanisms of adaptation and gene specialization, specifically neofunctionalization or subfunctionalization. Our discoveries highlight the importance of integrating comparative expression and selection analyses to understanding molecular evolution and the genetic mechanisms of invasiveness in plants. Furthermore, considering that transcriptomic resources are a fundamental source of information for identifying gene networks associated with biological processes, their implementation in the context of climate change may be a crucial tool for finding candidate genes related to plant invasiveness processes in *Thunbergia*. Additionally, the implementation of different expression and selection analyses can help to clarify the evolutionary mechanisms of invasiveness in plants.

Methodology

Data and information source

For the present study, previously generated transcriptomic data by Ruiz-Londoño and collaborators (2024) from vegetative and reproductive tissues of *T. alata* and *T. grandiflora* were used. The analysis of the transcriptomic data enabled the assessment of sequence integrity and completeness, quantification of transcript abundance, identification of differentially expressed genes between tissues and species, and detailed functional annotation. Additionally, targeted search for candidate genes related to adaptation and invasiveness were conducted, along with the enrichment analysis of functional categories by Gene ontology (GO) associated with metabolism, reproduction, development, stress response and pathogen defense.

Phylogenetic analysis

We constructed gene trees and species trees phylogenies using transcriptomic and genomic data. We selected 32 genes present across 21 angiosperms species (Table S1). MAFFT alignments for each gene were used to generate individual phylogenetic trees with RAxML (Revell & Harrison, 2008), applying the GTR+G+I substitution model, number Bootstrap analysis (1000 bootstraps), and searching for the best-scoring ML tree in a single program run. The phylogenetic trees of angiosperms were then analyzed with IQ-TREE2 for multi-copy gene tree inference (Minh et al. 2020). We used the default IQ-TREE2 parameters and selected the best tree. It's important to note that the phylogenetic analysis with IQ-TREE2 was conducted using a distinct set of parameters and methodological approaches compared to those phylogenetic trees employed in the selection pressure analyses.

Network analyses - Co-expression

We used the WGCNA package from the R software (Langfelder & Horvath 2008) to identify co-expression modules related to genes of invasiveness using expression data (TPM10K) from transcriptomes of *T. alata* and *T. grandiflora* species. WGCNA was run with the following parameters: power = 6, TOMType = "signed", minModuleSize = 20, maxBlockSize = 30000, deepSplit=2.5, reassignThreshold = 0, mergeCutHeight = 0.25, corType = "pearson". To identify genes associated with co-expression modules we searched for significant modules containing hub genes potentially associated with the following key annotation words: development, reproduction, metabolism, and stress responses.

We also searched for functions enriched by using the “Functional Annotation Table” tool from the DAVID server (https://david.ncifcrf.gov/home.jsp). Specifically, we used the Refseq_RNA codes of the genes from the 2 modules as gene lists while we used as backgrounds the Refseq_RNA codes from all the genes imputed to WGCNA.

Candidate and hub genes identification

We identified candidate genes based on their expression levels, statistical significance, functional annotation (i.e., Gene Ontology), and relevance to adaptation and invasiveness processes as supported by the scientific literature. These genes were prioritized for the main selection pressure analyses. Therefore, we employed a target search which it's an important approach to identify candidate genes based on previous scientific reports of validated genes associated with several biological processes and functional categories related with invasiveness (Table S2). For instance, through transcriptomic data source and information, and functional annotation, we could have broader insights into the identification and selection of candidate genes. To further characterize candidate genes, we categorized the genes according to their origin within the transcriptome of *T. alata* and *T. grandiflora*. Genes were grouped by target search, developmental stage (late vs. early), tissue type (vegetative vs. reproductive), and the interspecific comparison between *T. grandiflora* and *T. alata*. Based on the above, we identified a total of 38 genes through target search (Table S3), and defined three sets of the top 40 differentially expressed genes (DEGs) (each consisting of 20 upregulated and 20 down regulated genes) derived on three comparisons: developmental stage (Table S4), tissue type (Table S5), and the interspecific comparison (Table S6). These candidate genes were identified by local alignment using NCBI

BLAST+ software (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) in both vegetative and reproductive transcriptomes. Genes were determined for their potential involvement in the invasive capacity of the species, based on statistically significant differential expression levels filtered by FDR. This analysis was complemented with functional annotation and supported by relevant scientific literature.

Additionally, through co-expression network analysis of the tissue dataset, we identified 10 hub genes for each comparison (developmental stage, tissue type, and interspecific comparison) and for the target search (Table S7). Through this approach, we could identify potentially candidate genes and elucidate their functional categories potentially related to adaptive and invasive processes. However, while these hub genes provide broader insight into gene co-expression and regulatory potential, the candidate genes of this study remain central to the core investigation of the adaptive and invasive evolutionary mechanisms.

Selection pressure analyses

To identify the candidate genes under selection, we scanned the corresponding orthologues genes to the *Thunbergia* species, alongside with a selection of 18 angiosperms species comprising different lineages and one outgroup (*Arabidopsis thaliana*) using the software EasyCodeML (Gao et al. 2019). This tool facilitated the computation and visualization of the results of non-synonymous (*Ka* or *dN*) and synonymous (*Ks* or *dS*) substitution rates, as well as their ratios ($\omega = Ka/Ks$ or dN/dS). In this study, we used the acronyms Ka/Ks or dN/dS for ω in the EasyCodeML analyses and just Ka/Ks for ω in the further use of MEGA 11 analyses.

The alignments were carried out in PAML format and a phylogenetic tree in Newick format from the sequences of each candidate gene were needed. Firstly, the coding sequences (CDS) of the *Thunbergia*'s species were obtained using the "Complete ORF predict (Batch mode)" function in TBtools-II (Chen et al. 2023), while for the rest of the species the CDSs were obtained from the NCBI database using the BLAST tool (Sayers et al. 2021). To avoid ambiguity in sequence quality and annotation, we downloaded the orthologues sequences of the 18 species and the outgroup with the specific tag "Reference Sequence" (RefSeq) from the NCBI Genbank. The selected coding sequences for each candidate gene according to their corresponding species were saved in FASTA format. The "MAFFT" function (Katoh & Standley 2013) in PhyloSuite (Zhang et al. 2020) was utilized to perform multiple gene alignment of the CDSs. The alignment mode was set to "Normal", with the standard code and an auto strategy (-auto). The aligned CDSs were then revised in BioEdit (Hall & Carlsbad 2011), to improve the results of the alignment. The resulting aligned sequences by this method were saved in FASTA format. Then, we used the "IQ-TREE" function in Phylosuite (Minh et al. 2020) for constructing the ML tree. The new aligned sequences file was imported, with *A. thaliana* as the outgroup. For the corresponding parameters we used the model "GTR + I + G" with a "ML-optimized" state of frequency of 4 categories, and for the bootstraps we used the "Ultrafast" strategy of 10000 bootstraps and 10000 maximum iterations. The default values were used for the remaining parameters. Subsequently, we used the "Converted Sequence Format" function in PhyloSuite to convert the alignment files in FASTA format to PAML format. Then, to visualize the phylogenetic tree files (TreeFile format) we used the software FigTree (Rambaut 2018) and exported the resulting trees in Newick format.

For the selection analyses using EasyCodeML, we employed a site-specific model with seven site models (M0, M1a & M2a, M3, M7, M8 & M8a) to detect adaptation signatures across candidate genes for *Thunbergia*'s transcriptomes. This model allowed ω ratio to vary among sites while maintaining a fixed ω ratio across all branches. Specifically, the site-specific models, M1a (nearly neutral) vs. M2a (positive selection), and M7 (β) vs. M8 (β & ω) were computed to detect positive selection. Likelihood ratio tests (LRT) comparing M0 vs. M3, M1a vs. M2a, M7 vs. M8, and M8a vs. M8 were employed to assess the strength of selection (Yang & Nielsen 2002, cited by Feng et al. 2024). Likewise, we employed a branch-site model with two branch-site models (Model A & Model A Null). This model allowed ω ratio to vary among sites and lineages by assigning foreground and background branches *a priori* and testing for positive selection on specific codons. In particular, the branch-site models, model A (positive selection) vs. model A null (no positive selection) were compared to identify codon sites under positive selection along the foreground lineages (Zhang et al. 2005). To assess the strength of selection, we also employed an LRT to compare the Model A vs. Model A Null. To conduct the selection analyses in *T. alata* and *T. grandiflora*, the sequences of both species for each candidate gene were designed as the foreground branches under both models, while the remaining species were assigned as the background. For both models, the Bayes Empirical Bayes (BEB) method (Yang et al. 2005) was utilized to calculate posterior probabilities and filtered the false positives. In this analysis, sites with posterior probabilities greater than 0.95 (*) were considered to be under positive selection, while those exceeding 0.99 (**) were classified as experiencing strong positive selection.

We performed additional evolutionary and selection analyses using MEGA 11 (Tamura et al. 2021). Analyses were conducted using the Nei-Gojobori model (Nei & Gojobori 1986). The rate variation among sites was modeled with a gamma distribution (shape parameter = 1) and Invariant Sites (I). We found the number of synonymous substitutions per synonymous site (*Ks*) and the number of nonsynonymous substitutions per nonsynonymous site (*Ka*) from between sequences to calculate the *Ka*/*Ks* ratio. All ambiguous positions were removed for each sequence pair (pairwise deletion option). The presence of n/c in the results denotes cases in which it wasn't possible to estimate evolutionary distances.

Results

Phylogenetic analysis

To visualize the distribution and representation of the different angiosperms taxon in our study, we generated a phylogeny by using the transcriptomic sequences of the two *Thunbergia* species obtained by Ruiz-Londoño and collaborators (2024). We additionally included 18 CDSs sequences of angiosperms transcriptomes available in databases like Phytozome (https://phytozome-next.jgi.doe.gov/) and the NCBI (https://www.ncbi.nlm.nih.gov/). Both CDSs and genomic and transcriptomic data available were employed in the phylogenetic analyses. Moreover, *A. thaliana* was used as an outgroup to establish the relationships among the species (Fig. 3). The ML tree based on both CDSs and transcriptomic data exhibited strong branch support (bootstrap values >80%), indicating the reliability of the constructed tree.

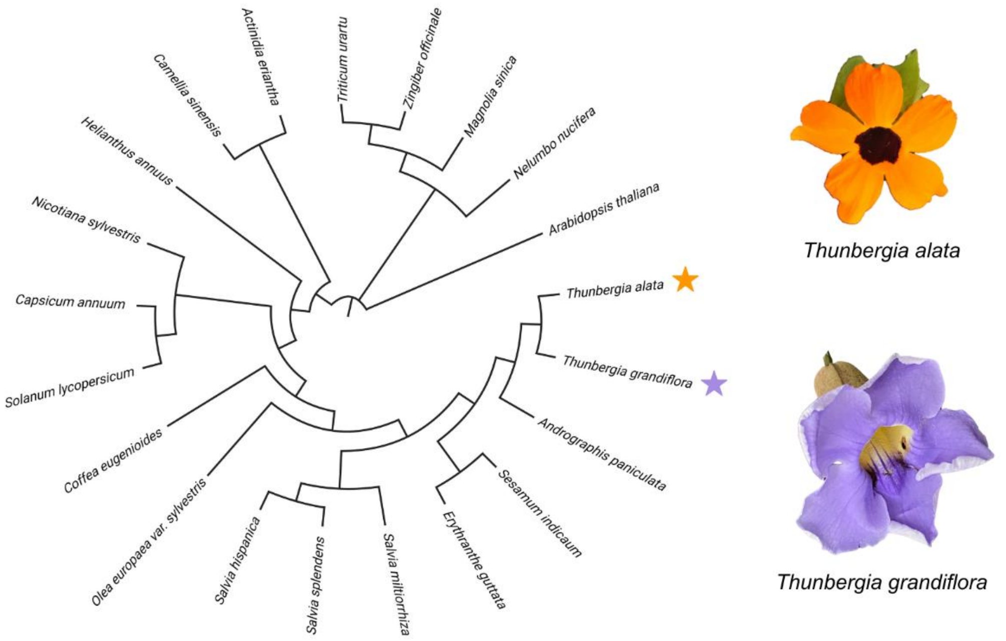

Fig. 3 Phylogeny from Maximum Likelihood analysis (RAxML) of angiosperm species considered in the selection analysis. The phylogeny has been rooted with the outgroup species used in the selection analyses *A. thaliana*. Orange star highlights *T. alata*, while purple star highlights *T. grandiflora*.

Co-expression network hub genes

To identify co-expressed networks of genes potentially associated with invasiveness in *T. alata* transcriptome (Fig. 4.A), we employed weighted correlation network analysis (Langfelder & Horvath 2008). Our fundamental assumption is that genes involved in the diversification and evolutionary signatures of invasiveness exhibit correlated expression patterns across *Thunbergia* species. We identified two modules containing the most significant correlated genes based on their expression levels (genes with more significant intergenic correlations). Specifically, the modules corresponding to green and magenta color assignments (Fig. 4.B) were identified as the most relevant based on their gene connectivity and functional enrichment profiles (Fig. S1). Notably, these modules are enriched in functions related to stress responses, pathogenesis defenses, mitochondrial metabolism, and protein synthesis. According to module detection, we verified the scale-free topology fit index to ensure that the selected soft-thresholding power ($\beta = 6$) resulted in a biologically meaningful network architecture (Fig. S2). No significant co-expression modules could be recovered for *T. grandiflora*, and consequently, only *T. alata* was maintained for the network analysis.

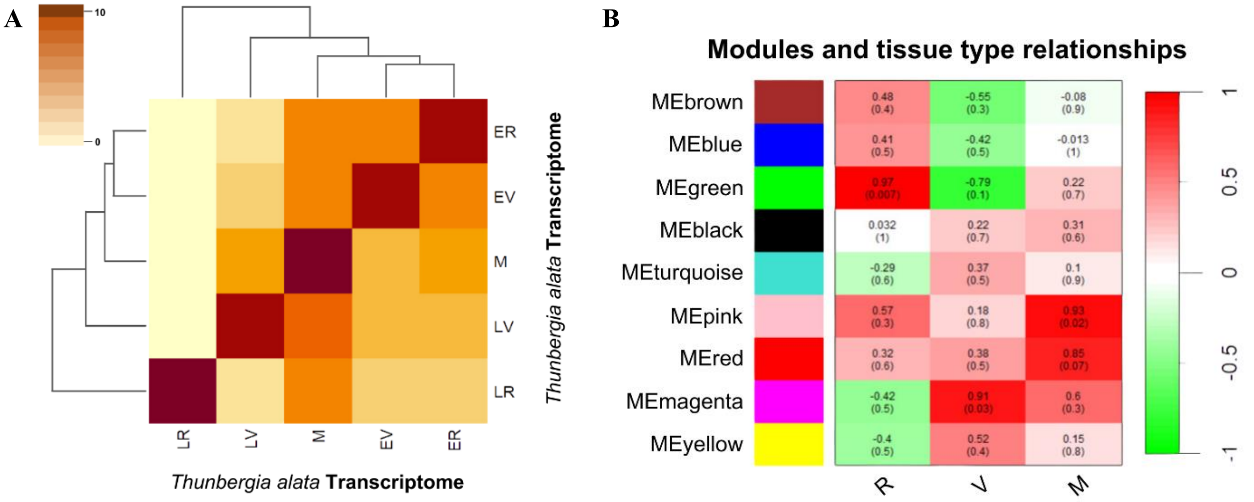

Fig. 4 Gene expression analysis and module-tissue correlation in *T. alata* transcriptomes A. Correlation analysis of the co-expression of differentially expressed genes for the *T. alata* transcriptome, comparing transcriptomes from the following categories: early reproductive (ER), early vegetative (EV), late reproductive (LR), late vegetative (LV), and mixed (M). **B.** Figure of co-expressed gene modules in *T. alata* transcriptome for the categories of reproductive (R), vegetative (V), and mixed tissues (M). In terms of expression levels, the green module is the most significant for reproductive tissues, while the magenta module is the most significant for vegetative tissues.

Considering the above, to reveal genes potentially associated with invasiveness in *T. alata*, we searched for genes enriched in functions related to development and reproduction; and defense and stress response contained in the two candidate modules and displayed differential expression in *Thunbergia* species (Table S8). In the green candidate module associated with reproductive tissue, we identified an interesting gene, *SWEET1*, which is linked to both development and stress responses (Table 1). Meanwhile, in the magenta candidate module corresponding to vegetative tissue, we found two notable genes, *RPL21C* and *RAP*, involved in protein synthesis and stress responses respectively, which could potentially be linked to the invasiveness of the invasive species (Table 1).

Table 1 Correlation of co-expressed candidate genes modules in *T. alata* reproductive, vegetative, and mixed transcriptomes, based on a WGCNA direct search under the following comparison categories: target search, developmental stage, tissue type, and interspecific comparison between *T. alata* and *T. grandiflora*. However, only the comparisons where genes located in highly co-expressed modules (specifically the green and magenta modules with bold selection) are reported here.

Comparison	Gene Name	Reproductive_Module	Vegetative_Module	Mixed_Module
Tissue type	*CYP450*	MEyellow	MEyellow	MEyellow
	PDC	MEgreen	MEblue	MEblue
	UGT	MEturquoise	MEturquoise	MEturquoise
	PSAH2	MEturquoise	MEturquoise	MEturquoise
	FBN12	MEturquoise	MEturquoise	MEturquoise
	RGLG2	MEblue	MEblue	MEblue
	PCC13-62	MEbrown	MEbrown	MEbrown
	SWEET1	**MEgreen**	**MEgreen**	**MEgreen**
	FLA21	MEbrown	MEbrown	MEbrown
	FDH1			
Interspecific comparison	***RPL21C***	**MEmagenta**	**MEmagenta**	**MEmagenta**
	HSCP70-1	MEbrown	MEbrown	MEbrown
	GLP1-16	MEblue	MEblue	MEblue
	rpsl	MEturquoise	MEturquoise	MEturquoise
	EXPA8	MEblue	MEblue	MEblue
	RAP	**MEmagenta**	**MEmagenta**	**MEmagenta**
	AGP18	MEred	MEred	MEred
	PMA3	MEbrown	MEbrown	MEbrown
	CYP1	MEblue	MEblue	MEblue
	psbY	MEblue	MEblue	MEblue

Candidate genes for selection pressure analyses

To investigate the adaptation signatures for positive selection of genes potentially associated with invasiveness in *T. alata* and also in its congener *T. grandiflora*, we identified a total of 38 genes through the target search. Additionally, three distinct sets of top 40 DEGs (20 upregulated and 20 downregulated, each one) were identified, each corresponding to a specific comparison: developmental stage (late vs. early), tissue type (vegetative vs. reproductive), and the interspecific comparison between *T. grandiflora* and *T. alata*. The selection of the three distinct sets of top 40 DEGs, was based on statistical significance (p-value ≤ 0.05 and FDR ≤ 0.05) and log fold change (logFC) thresholds. For the developmental stage and interspecific comparisons, a stringent logFC threshold of > 5 or < -5 was applied. While for the tissue type comparison, a slightly more relaxed threshold of > 2 or < -5 was used. In some cases (particularly in the tissue type comparison) genes with strong biological relevance were retained based on p-value alone when FDR filtering was too stringent. In addition, genes were prioritized based on functional annotation from Ruiz-Londoño and collaborators (2024), focusing on GO terms associated with metabolic processes, pollen, fruit and stamen development; reproduction; cold and heat acclimatation; response to oxidative, abiotic and xenobiotic stresses; defense against pathogens; and other processes potentially linked to adaptive capacity and invasiveness. In addition, these genes exhibited differential expression when comparing the vegetative and reproductive tissues, as well as the late and early developmental stages in *T. alata*. Additionally, the comparison between *T. grandiflora* and *T. alata* is important to reveal the insights on the genetic basis of their ecological and morphological differences.

Based on this selection, we carried out a final selection of the top 10 genes in each set (including the target search) corresponding to their specific comparison. This final selection aimed to identify the most relevant and biologically interesting genes potentially associated with invasiveness in both *Thunbergia* species, providing a focused set of candidate genes for downstream selection pressure analyses for each gene. An additional filtering step was applied to the top genes from the final selection of each group set. This was necessary because, during the construction of gene trees for the selection pressure analyses, some genes produced phylogenies that lacked biological coherence or displayed alignment issues that compromised the accuracy of their evolutionary relationships. Therefore, genes with poorly resolved or unreliable phylogenetic trees were excluded from the candidate gene selection thus, for the selection pressure analyses to ensure the robustness and interpretability of downstream evolutionary inferences. As a result, we identified 32 candidate genes from the combined selection across all four sets (target search, developmental stage, tissue type, and interspecific comparison). Specifically, 7 genes were selected from the target search, 8 from the developmental stage comparison (4 upregulated, 4 downregulated), 9 from the tissue type comparison (5 upregulated, 4 downregulated), and 8 from the interspecific comparison between *T. alata* and *T. grandiflora* (4 upregulated, 4 downregulated) (Table S9-S11).

Selection pressure analyses

The estimation of the ratio *ω* (*ω* = *Ka*/*Ks* or *dN*/*dS*) of non-synonymous (*Ka* or *dN*) to synonymous (*Ks* or *dS*) substitutions was performed for 32 candidate genes from 21 species (2 from *Thunbergia* genus). Various site models (M0, M1a & M2a, M3, M7, M8 & M8a) were employed to compare model fit using likelihood ratio tests (LRTs). Among these models, only comparisons of M1a vs. M2a and M7 vs. M8 were used to specifically test for the presence of positive selection and amino acids exhibiting positive selected sites. Considering *ω* ratios of Model 0, all the genes were identified under purifying selection ($\omega < 1$), indicating selective pressure to maintain their functions (Table S12 and 13). However, when comparing the models, the results showed 1 gene and one of its copies with sites evolving under positive selection, as evidenced by the log-likelihood values and the likelihood ratio test (LRT) results (Table S14-15). In particular, the *DEGP1* gene and *DEGP1*(2) shows a significant result in their likelihood ratio test statistic of 58.303552 and 30.273948 respectively for the comparison of M1a vs. M2a, and 89.645546 and 168.697148 respectively for the comparison of M7 vs. M8 in *T. alata* and *T. grandiflora*, indicating sites under strong evidence of positive selection (Table 2 and 3). These genes also exhibit statistically significant *p*-values well below the threshold of 0.05. Among these genes, *DEGP1*(2) displayed the highest count of positive amino acid sites (11), trailed by *DEGP1* gene (8). Functionally, these genes under positive selection belong to a category that is crucial for protein modification and degradation processes (Table S11).

Table 2 Likelihood ratio test (LRT) results for positive selection in candidate genes were obtained using the site-specific model from EasyCodeML, assuming the foreground branch in *T. alata*. A significant LRT statistics and *p*-value (≤ 0.05) indicate evidence for positive selection. Only the comparison where genes presented specific codon sites under positive selection are reported here.

Comparison	Gene	M1a vs. M2a 2*|ΔlnL|	M7 vs. M8 2*|ΔlnL|	LRTs (*p*-value) M1a vs. M2a	LRTs (*p*-value) M7 vs. M8	Positive selection in specific sites
Tissue type	*ROPGEF12*	2.899999999	39.67959	1.00E+00	2.00E-09	False
	AED3	0	18.40005	1.00E+00	1.01E-04	False
	DEGP1	**58.303552**	**89.645546**	**0.00E+00**	**0.00E+00**	**True (8 sites)**
	DEGP1(2)	**30.273948**	**168.697148**	**2.67E-07**	**0.00E+00**	**True (11 sites)**
	HEMC	0	44.462646	1.00E+00	0.00E+00	False
	ALDHC4	0	67.560796	1.00E+00	0.00E+00	False
	SEP3	0	7.061692	1.00E+00	2.93E-02	False
	NCAPH2	0	4.995436	1.00E+00	8.23E-02	False
	PAP2	0	2.030334	1.00E+00	3.62E-01	False

Table 3 Likelihood ratio test (LRT) results for positive selection in candidate genes were obtained using the site-specific model from EasyCodeML, assuming the foreground branch in *T. grandiflora*. A significant LRT statistics and *p*-value (≤ 0.05) indicate evidence for positive selection. Only the comparison where genes presented specific codon sites under positive selection are reported here.

Comparison	**Gene**	**M1a vs. M2a**	**M7 vs. M8**	**LRTs (*p*-value)**		**Positive selection in specific sites**
		2*|ΔlnL|	**2*|ΔlnL|**	**M1a vs. M2a**	**M7 vs. M8**	
Tissue type	*ROPGEF12*	0	39.67959	1.00E+00	2.00E-09	False
	AED3	0	18.40005	1.00E+00	1.01E-04	False
	DEGP1	**58.303552**	**89.645042**	**0.00E+00**	**0.00E+00**	**True (8 sites)**
	DEGP1(2)	**30.273948**	**168.697148**	**2.67E-07**	**0.00E+00**	**True (11 sites)**
	HEMC	0	44.462646	1.00E+00	0.00E+00	False
	ALDHC4	2.000000677	67.560796	1.00E+00	0.00E+00	False
	SEP3	0	7.061692	1.00E+00	2.93E-02	False
	NCAPH2	0	4.995436	1.00E+00	8.23E-02	False
	PAP2	0	2.030334	1.00E+00	3.62E-01	False

The likelihood ratio tests (LRTs) for positive selection using the branch-site model with *Thunbergia* species individually as the foreground branch revealed strong evidence of genes presented specific codon sites under positive selection for 4 genes (Table S16 and S17). Specifically, the genes *POK*, *ANJ1*, *ACLB1* and *NAC92* show significant LRT statistics and *p-values* (*p*-value ≤ 0.05), indicating that they have undergone adaptive evolution. Particularly, *POK* and *ANJ1* exhibit a *p*-value of 1.47E-07 and 1.41E-03 respectively, and it's important to note that these two genes exhibit low expression levels at positively selected sites in *T. alata* genes, while for *T. grandiflora* a the *ACLB1* and *NAC92* genes exhibit a *p*-value of 0.00E+00 and 5.71E-07 respectively, also indicate robust evidence for positive selection (Table 4 and 5).

Table 4 Likelihood ratio test (LRT) results for positive selection in candidate genes were obtained using the branch-site model from EasyCodeML, assuming the foreground branch in *T. alata*. A significant LRT statistics and *p*-value (≤ 0.05) indicate evidence for positive selection. Only the comparison where genes presented specific codon sites under positive selection are reported here.

Comparison	**Gene**	**Model A null**	**Model A**	**Model A null vs. Model A**		**Positive selection in specific sites**
		|ΔlnL|	**|ΔlnL|**	**2*|ΔlnL|**	**LRTs (*p*-value)**	
Developmental stage	*MFP2*	-21795.78842	-21798.78929	6.00	1.43E-02	False
	GAPDH	-8057.298279	-8057.366728	0.14	7.11E-01	False
	AGT1	-10894.96631	-10894.96631	0.00	1.00E+00	False
	NAC92	-4738.888507	-4740.869828	3.96	4.65E-02	False
	SR45	-5930.66369	-5930.66369	0.00	1.00E+00	False
	PBL8	-5396.120038	-5397.651782	3.06	8.01E-02	False
	POK	**-6716.300219**	**-6730.114718**	**27.63**	**1.47E-07**	**True (17 sites)**
	ANJ1	**-3853.841361**	**-3858.936887**	**10.19**	**1.41E-03**	**True (7 sites)**

Table 5 Likelihood ratio test (LRT) results for positive selection in candidate genes were obtained using the branch-site model from EasyCodeML, assuming the foreground branch in *T. grandiflora*. A significant LRT statistics and p-value (≤ 0.05) indicate evidence for positive selection. Only the comparison where genes presented specific codon sites under positive selection are reported here.

Comparison	**Gene**	**Model A null |ΔlnL|**	**Model A |ΔlnL|**	**Model A null vs. Model A**		**Positive selection in specific sites**
				2*|ΔlnL|	**LRTs (*p*-value)**	
Target search	*PR3*	-4672.925896	-4672.925896	0.00	1.00E+00	False
	EIN3	-12045.66579	-12047.48235	3.63	5.66E-02	False
	BPM4	-9882.733597	-9882.733523	0.00	9.90E-01	False
	PEPC	-28197.82366	-28197.82366	0.00	9.99E-01	False
	VDAC2	-3172.361146	-3172.361146	0.00	1.00E+00	False
	ACLB1	**-9573.44234**	**-9593.0973**	**39.31**	**0.00E+00**	**True (8 sites)**
	TADA2	-9832.343887	-9832.860738	1.03	1.03E+00	False
Developmental stage	*MFP2*	-21801.12125	-21801.12125	0.00	1.00E+00	False
	GAPDH	-8048.378378	-8049.987442	3.22	7.28E-02	False
	AGT1	-10896.25757	-10896.25757	0.00	1.00E+00	False
	NAC92	**-4707.61022**	**-4720.114415**	**25.01**	**5.71E-07**	**True (6 sites)**
	SR45	-5930.516267	-5930.51491	0.00	9.58E-01	False
	PBL8	-5397.904058	-5397.904058	0.00	1.00E+00	False
	POK	-6778.652096	-6780.435442	3.57	5.89E-02	False
	ANJ1	-3877.747838	-3877.746353	0.00	9.57E-01	False

According to the results from additional evolutionary analyses conducted in MEGA 11 using a pairwise comparison approach, all 32 candidate genes showed ω (*Ka*/*Ks*) values indicative of purifying selection ($\omega < 1$) (Table S18-S50). While slight variation in ω estimates were observed, the results were very consistent with those obtained using EasyCodeML, supporting the reliability of the selection patterns inferred across both methods.

Discussion

This study aimed to analyze the molecular evolution of genes potentially associated with invasiveness in two species belonging to *Thunbergia* genus. Integrative approaches combining genetic, molecular and evolutionary analyses provide comprehensive insights into the factors that contribute to the adaptive and invasive processes to expand and succeed in new environments. For example, multi-omics approaches have been employed to identify genes involved in several biological processes, which are often upregulated in invasive species, suggesting their important role in the successful colonization and competition with native species (Mable 2019; Sherpa & Després 2021; Qi et al. 2023). Additionally, by examining gene co-expression networks we could identify functional modules containing candidate genes that potentially have evolved to confer adaptive advantages (Serin et al. 2016), and through the construction of phylogenetic trees we could elucidate the evolutionary relationships and divergence among species (Smith et al. 2020). Understanding the molecular and evolutionary mechanisms underlying plant invasiveness can provide relevant insights and knowledge for developing effective management strategies to control and mitigate the spread and impact of invasive species (Lee 2002; Le Roux & Wieczorek 2009).

Phylogenetic relationships between *Thunbergia* species and other angiosperms

Previous studies have utilized transcriptomic techniques to conduct phylogenetic analyses from various invasive plant species of interest, with the aim of supporting the development of effective control and management strategies at a global scale (Guo et al. 2018; Ruan et al. 2022; Chuang et al. 2022). In this study, we employed both CDSs and transcriptomic data to reconstruct the phylogenetic relationships among 21 species of angiosperms including the two *Thunbergia* species. Significantly, the phylogenetic tree constructed using these two data types presented high consistency, validating the reliability of utilizing plant transcriptomes for phylogenetic reconstruction.

Understanding phylogenetic relationships is essential in evolutionary studies, as it provides a framework for tracing how traits (morphological, molecular, or ecological) have evolved over time. Modern phylogenetics, driven by sequence data, enables comparative analyses that link trait evolution to lineage diversification and reveal patterns of adaptation and conservation across species (Soltis et al. 2019; Smith et al. 2020). To ensure the robustness of phylogenetic inferences, several studies employ multiple tree reconstruction methods, such as Maximum Likelihood (ML), Neighbor-Joining (NJ), and Bayesian Inference (BI), each with distinct algorithmic approaches and assumptions (Feng et al. 2024; Cauz-Santos et al. 2025). Integrating results from different phylogenetic methods enhances the reliability of evolutionary hypotheses and minimizes biases that may arise from methodological limitations or dataset-specific tools.

The *Thunbergia* genus includes a wide variety of ornamental and horticultural plants with considerable economic importance. However, some species within this family are highly invasive in certain regions, causing significant ecological impacts. A major challenge in both promoting economically valuable crops and controlling and managing invasive species lies in the lack of a well-resolved taxonomy based on phylogenetic relationships (Feng et al. 2024). Therefore, a comprehensive investigation into the evolutionary history of *Thunbergia* is essential to overcome these issues. This study specifically examines two species within the genus *Thunbergia*: *T. alata* and *T. grandiflora*.

Co-expressed network of genes potentially associated with invasiveness

The application of WGCNA for co-expression network analysis constitutes a robust approach for identifying correlated gene modules in transcriptomic studies, facilitating functional inference from a large-scale of differential expression data (Zhang & Horvath 2005; Langfelder & Horvath 2008). In the *T. alata* transcriptome, WGCNA revealed modules enriched in processes potentially associated with invasiveness. This approach allowed the identification of key hub genes with important regulatory roles and facilitated the retention of the most significantly co-expressed genes by their differential expression, linked to the functional dynamics of the co-expressed modules (Serin et al. 2016; Rao & Dixon 2019; Ovens et al. 2021). The use of weighted networks combined with soft-thresholding, improves biological robustness and enhances the cohesion within co-expressed modules, compared to other hard-threshold methods (Zhang & Horvath 2005). Additionally, hierarchical modularity alongside with intramodular connectivity metrics correlate closely with the biological relevance of genes potentially associated with invasiveness (Ruan et al. 2010).

On the other hand, we suggest that the lack of robust co-expression results for the *T. grandiflora* transcriptome may be attributed to differences in the transcriptomic data available for each species. The transcriptome of *T. alata* included a comprehensive set of samples including early and late stages of both reproductive and vegetative tissues, while the data for *T. grandiflora* was limited to vegetative and reproductive tissues, potentially reducing the variability necessary to detect robust co-expression modules. In addition, the consolidation of co-expression modules identified for *T. grandiflora* didn’t provide statistically reliable results, which could lead to the identification of gene clusters that appear to be co-expressed, but they are, in fact, “artifacts” generated by the parameters of the co-expression model employed in the co-expression analysis.

Through the GO terms classification and enrichment analysis conducted on co-expressed hub genes from *T. alata* transcriptome (Table S7), several genes were found to be associated with distinct biological processes across different comparative categories, including target search, developmental stage, tissue type, and interspecific comparison. Many of these genes appear to be involved in cellular functions such as protein folding and stress response, exemplified by *Heat Shock Cognate Proteins 70-1 y 80* (*HSCP70-1* and *HSCP80*) which are linked to cellular response to unfolded proteins, protein folding; as well as *Ras-related protein* (*RAP*) related to stress and defense responses (Fig. 5). Moreover, genes such as *Sugar Will Eventually Exported Transporter 1* (*SWEET1*) associated with carbohydrate metabolism, and *Cytochrome P450* (*CYP450*) are involved in metabolic processes, highlighting important metabolic activities in tissue-specific contexts (Fig. 5). Additionally, genes related to DNA repair and replication, including *Minichromosome Maintenance Complex Component 5* (*MCM5*) and *Plant U-box Protein 35* (*PUB35*), were identified in developmental stage comparisons, suggesting active regulation of genome stability during growth phases (Fig. 5). According to the interspecific comparison, genes like *Ribosomal Protein L21C* (*RPL21C*) y *Expansin A8* (*EXPA8*) were identified being associated with protein activity and cell wall dynamics respectively, indicating a potential difference in protein synthesis and cellular structure between *T. alata* and *T. grandiflora* (Fig. 5). Overall, these findings suggest that co-expressed gene modules capture key biological processes such as development, reproduction, metabolism, and stress response (Table S51), that have been reported in previous studies promoting adaptability, which increase the invasive potential of the species that they are associated with (Guo et al. 2018; Xu & Wang 2019; Parvizi et al. 2023).

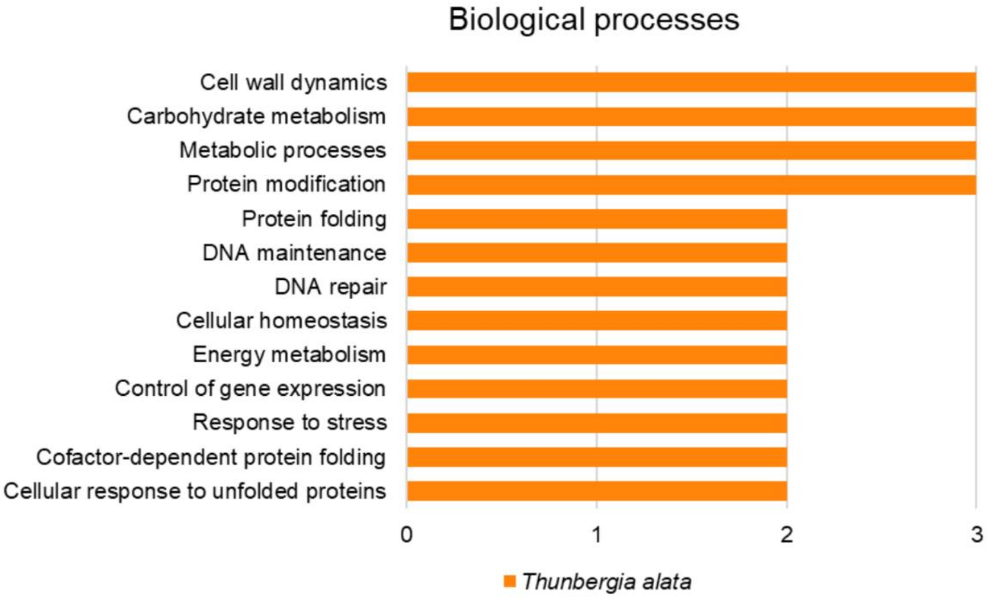

Fig. 5 Gene ontology (GO) classification and enrichment analysis focused on biological processes were conducted for potential candidate genes from co-expressed gene modules of *T. alata* transcriptome. These genes were categorized based on previous comparisons including targeted search, developmental stage, tissue type, and interspecific comparisons. The GO terms presented here correspond to those with more than one associated gene (hit) in the *T. alata* species.

Interestingly, the *SWEET1* gene, belonging to the green module associate with reproductive tissues in *T. alata*, is a key transporter for carbohydrate distribution in plants and is essential for reproductive development such as pollen growth and seed filling, processes that support plant reproduction success, potentially linked for invasive plants (Chen et al. 2012; Eom et al. 2015). Additionally, *SWEET1* contributes to abiotic stress response by regulating plant homeostasis and improving tolerance to conditions like drought and salinity through its role in sugar transport at cellular level (Zhu et al. 2024). Meanwhile, in the magenta module linked to vegetative tissues and interspecific comparisons, *RPL21C*, a ribosomal protein involved in translation and protein synthesis, plays an essential role in the formation and development of functional chloroplasts (Wilson & Cate 2012; Yin et al. 2012). The capacity for efficient protein synthesis is vital under changing environmental conditions faced by invasive plants. Similarly, the *RAP* gene, encoding a *RAS-related GTPase*, is implicated in intracellular signaling pathways regulating stress responses and adaptive mechanisms (Hodge & Ridley 2016). In addition, *RAP* as a RIN4-associated protein localized at plasma membrane, plays a key role in coordinating plant immune signaling and response to biotic stress (Liu et al. 2009). Together, these genes suggest molecular pathways by which *T. alata* may promote reproductive and vegetative adaptability, providing the basis of its ecological success and potential invasiveness.

All results generated through this methodology require support from complementary models that can produce findings consistent with those obtained through WGCNA. Additionally, increasing the number of biological replicates could enhance the identification of hub genes whose co-expression is truly correlated with biological functions relevant to invasiveness in *T. alata* specifically. These hub genes may serve as a foundation for future studies involving validation strategies, such as mutant validation, controlled population experiments, and selection pressure analyses. Furthermore, the study of molecular networks could be fundamental for identifying groups of genes whose expression may shift under specific conditions, such as stress responses or reproductive changes associated with invasiveness of plant species.

Insights into the identification and selection of candidate genes

The identification and selection of candidate genes is fundamental to elucidating the genetic and molecular basis of adaptive traits, including invasiveness associated with biological processes such as metabolism, reproduction, stress response and pathogen defense. Traditionally, the candidate gene identification and selection often depend on existing knowledge of biological traits (such as biological processes and functional assumptions) which can limit discovery, particularly when the genetic and molecular basis of the biological traits are not well characterized (Zhu & Zhao 2007).

To overcome these limitations, novel strategies have emerged. The Digital Candidate Gene Approach (DigiCGA), for example, integrative high-throughput genomic information to identify potential candidate genes without depend on functional assumptions, enhancing the scalability and objectivity of gene discovery (Zhu & Zhao 2007). Similarly, systems genetic approaches, particularly expression quantitative trait loci (eQTL) and co-expression network mapping, offer different ways to identify candidate genes based on their transcriptional activity patterns and regulatory architecture across populations (Feltus 2014). A particularly promising approach is offered by transcriptomics. This omics technologies allows the detection of expression changes in response to environmental conditions and has been significant in uncovering the role of transcriptional plasticity in biological invasions (Vaughan & Dhami 2024). Transcriptomics approaches can reveal genes that are differentially expressed during important phases of invasion or stress response, and may serve as indicators of phenotypic plasticity, a key role to facilitate the rapid adaptation when genomic variation is limited (Qi et al. 2023; Vaughan & Dhami 2024).

In this study, we used transcriptomic methodologies data to identify candidate genes potentially associated with invasiveness while studying their molecular evolutionary mechanisms by selection pressure analyses in *T. alata*, using its congener *T. grandiflora* as a comparative model. This approach, reflects a volunteer integration of target search, transcriptomic profiling, and co-expressed network mapping, enabling the discovery of genes likely involved in the adaptive and invasive potential of *T. alata*. These findings not only contribute to understanding the genetic and molecular basis of its invasiveness but also reinforce the value of methodologies such as co-expressed networks mapping, transcriptomics and multi-omics strategies in revealing complex mechanisms of plant invasiveness. The application of these integrative methodologies in further studies will be essential for understanding invasive species and designing efficient control approaches.

Adaptive evolution in *Thunbergia* species candidate genes

Adaptation to diverse environments is often driven thus, associated with positive (adaptive) selection, which promotes advantageous genetic changes. In contrast, negative (purifying) selection plays a central role in maintaining genomic stability by eliminating deleterious mutations, thereby preserving sequence integrity across long evolutionary timescales (Mosely et al. 2018, Cvijović et al. 2018, cited by Feng et al. 2024). The ratio of non-synonymous to synonymous substitution rates, commonly expressed as ω, is widely used indicator of selective pressure, with values greater than one typically signaling positive selection (Guo et al. 2018; Feng et al. 2024; Cauz-Santos et al. 2025). By examining the ratio ω, we found that all 32 candidate genes exhibited ω values < 1, indicating that purifying selection predominates across these genes. Despite the minor variations in the ω estimates results between MEGA 11 and EasyCodeML, the overall consistency in results supports the reliability of these selection inferences. Notably, MEGA 11 relies on pairwise sequence comparisons and doesn't incorporate phylogenetic relationships visualized in phylogenetic trees inputs. In contrast, EasyCodeML integrates phylogenetic context, enabling more refined detection of selection across specific branches or clades, in particular employing models such as site-specific or branch-site models. Therefore, the consistency between both approaches suggests that the observed purifying selection is robust, reflecting evolutionary pressures that limit nonsynonymous changes. The absence of positive selection across the full length of each gene implies a uniform pattern of functional conservation, likely due to the essential roles these genes play in key biological processes within the *Thunbergia* species. However, this overall pattern doesn't include the possibility that specific codon sites within these genes may be under positive selection, such as signals which may not be detectable when selection pressure is averaged across the entire gene (Yang & Nielsen 2002).

In this study, we identified 5 genes exhibiting positive selection sites. Specifically, we identified 2 genes in *T. alata*, 2 genes in *T. grandiflora*, and 1 gene and one of its copies in both *Thunbergia* species, featuring amino acids positive sites. The site-specific model test for positive selection revealed that *DEGP1* and *DEGP1*(2) exhibited a higher count (8, 11 respectively for *T. alata* and *T. grandiflora*) of positive amino acid sites among the tissue type comparison set of candidate genes in *Thunbergia* species, suggesting a potentially significant role of *DEGP1* gene in the adaptive evolution of *Thunbergia* species. This gene, with specific codon sites under positive selection and upregulated expression, is involved in protein modification and degradation processes. The *DEGP1* gene, encoding a serine protease, is involved in degrading damaged proteins within the chloroplast thylakoid lumen, contributing to photosystem II repair, and overall photosynthetic efficiency and heat response, especially under stress conditions (Chassin et al. 2002). While *DEGP1* hasn't been specifically identified in studies of invasive plant species, research on invasive species such as *Mikania micrantha* revealed that the adaptation to new environments involves changes in gene expression related to photosynthesis and stress response (Guo et al. 2018; Ruan et al. 2022).

Remarkably, this finding highlights the possibility that gene copies may be involved in redundant biological processes or may have acquired novel functions (Lallemand et al. 2020). This result aligns with those reported by Ruiz-Londoño and collaborators (2024), who identified homologous sequences to the candidate genes (for their study) in the transcriptomes of *T. alata* and *T. grandiflora*. They further suggested that these homologous genes could exhibit tissue-specific expression patterns with notably higher expression levels in reproductive tissues.

Even more remarkably, the branch-site model test for positive selection revealed several genes under adaptive evolution in the *Thunbergia* species. The genes identified with specific codon sites under positive selection, include those involved in development and reproduction, and defense and stress response processes (*POK*, *ANJ1* respectively) for *T. alata*, while for *T. grandiflora* are involved in primary and secondary metabolism, and transcription and translation processes (*ACLB1*, *NAC92* respectively) (Table S11). Notably, *T. alata* genes with positive selected sites (*POK, ANJ1*) are both downregulated (Table S9), whereas in *T. grandiflora*, the genes with positive selected sites like *ACLB1* or the upregulated gene *NAC92* (Table S10).

Interestingly in *T. alata*, the *POK* gene, which encodes a protein like yeast *Vps52p*, part of the *GARP/VFT* complex for vesicle trafficking, is crucial for pollen tube tip growth which is a form of cell expansion very important for plant fertilization. The *POK* protein localizes in Golgi compartments, indicating its role in membrane trafficking during pollen tube elongation (Lobstein et al. 2004; Scholz et al. 2020). This is particularly relevant as some invasive plant species key roles are those related to reproductive success and adaptation to new environments. Also, the *DnaJ* homolog protein *ANJ1* gene, is part of the *Hsp40*/*DnaJ* chaperone family, which facilitate the protein folding and stress responses interaction with *Hsp70*, according to evidence from a study in *A. thaliana* (Musskopf et al. 2018; Jia et al. 2021). In addition, *ANJ1* has an important role in modulating plasma membrane H+-ATPase enzyme activity that may improve salinity tolerance, a common challenge for invasive species in disturbed habitats (Yang et al. 2010). This suggests that *ANJ1* is involved in functions related to stress response and protein homeostasis which could contribute to the adaptability to a new environment for invasive plants.

Now for *T. grandiflora*, the gene *ACLB1*, which encodes a subunit of the ATP-citrate lyase (ACL) enzyme in plants, is crucial for generating cytosolic acetyl-CoA which it's a central metabolite involved in various biosynthetic pathways essentials for plant structure like the cell wall and function (Zhong et al. 2020). Additionally, evidence from a study in *A. thaliana* reveals the involvement of *ACLB* genes in the regulation of cell death and disease resistance to pathogens, may enable invasive plants to survive against biotic stresses in new environments, facilitating their establishment, reproduction and spread (Liu et al. 2022). Finally, the *NAC92* gene is a transcription factor that plays a significant role in various plant physiological processes, including stress and developmental regulation such as the activation of senescence-associated genes during several abiotic stress conditions (Nakashima et al. 2012; Podzimska-Sroka et al. 2015). While none of these genes themselves hasn't been directly reported on invasive plants, the involvement of these genes in metabolism, stress and defense response, and reproduction and development process, suggests that they may play a key role in the adaptive evolution of invasive species.

The adaptive evolution of candidate genes in *Thunbergia* species can be understood through complex molecular processes that include gene duplication, subfunctionalization, neofunctionalization, and the action of positive selection on specific amino acid sites. In plants, these mechanisms have demonstrated functional diversification, and specialization of key structures and physiological response (Wang et al. 2016). In our results, two DEGs with signals of positive selection in specific amino acid sites were found to be downregulated (*POK, ANJ1* genes found in *T. alata*), which contrast with the common trend of adaptative genes of being upregulated under positive selection pressures; also, it's important to note the presence of more amino acid sites selected under positive selection in the *DEGP1* gene copy. These findings may be explained by the "subneofunctionalization" model, in which a duplicated copy initially undergoes subfunctionalization, but then turn into a progressive acquisition of new or modified functions (neofunctionalization) under relaxed evolutionary restrictions arising from gene multifunctionality (He & Zhang 2005). This process suggests that the observed low expression may not be indicate a lack of functional relevance, but rather an optimization of expression levels required in specific tissues and development stages, as described in experimental models of protein evolution where adaptive mutations may be associated with structural or regulatory effects (Madhani & Kourki 2024). Likewise, the coevolution between coding regions and non-coding regulatory elements could modulate gene expression without compromising the adaptive functionality of the protein result (Jayaraman et al. 2022). These types of evolutionary dynamics add an additional layer of complexity to the functional interpretation of DEGs under positive selection, highlighting the importance of integrating expression, structure, and function data to infer probable evolutionary trajectories.

Our findings indicate that while purifying selection is the dominant evolutionary pressure acting on most candidate genes in *Thunbergia*, certain genes contain sites that appear to be under positive selection. This adaptive evolution may play a role in shaping the ecological and morphological contrasting differences between *T. alata* and *T. grandiflora*. Moreover, comparative transcriptomic and selection pressure analyses help to elucidate the molecular evolution and genetic mechanisms underlying plant invasiveness, contributing to a deeper understanding of invasion biology and the development of effective management strategies. Despite these advances, further research is needed to clarify the environmental or functional factors driving positive selection observed in these candidate and hub genes. In addition, functional validation of the identified sites, along with the exploration of other molecular pathways, is essential for fully understanding the genetic basis of plant invasiveness and should remain a key focus of future studies.

Conclusions

In this study, we demonstrate that the integration of complementary genomic strategies such as targeted gene identification, differential expression analysis, and co-expression network analysis, provides a robust and multidimensional framework for uncovering the genetic basis of the invasive traits. This multi-dimensional approach allowed for the identification of candidate genes in *T. alata* and its non-invasive congener *T. grandiflora*, strengthening the inference of their potential roles in adaptation and invasiveness.

The detection of positive selection signals in specific genes suggests that adaptive evolution plays an important role in shaping traits associated with invasiveness, including metabolism, development, reproduction, and stress responses. These patterns are implied in the invasiveness processes that may be the outcome of accumulated molecular adaptations that confer ecological advantages in new changing environments. Thus, our results support the broader hypothesis that biological invasions are, at least in part, evolutionary processes driven by natural selection acting on functional gene variants.

Our study suggests the value of other omics approaches such as evolutionary genomics as a key tool in invasion biology. It highlights those molecular signatures, in the context of functional annotation and ecological traits, can reveal underlying mechanisms of ecological divergence and specialization. Therefore, studying genes under positive selection is not only a methodological advance, but also a theoretical contribution to understanding how invasive potential genetically evolve.

Future research should aim to functionally validate the roles of candidate genes identified here, including those candidate genes of selection pressure analyses positively selected at specific sites and those identified in co-expression modules. Doing this will elucidate correlative patterns and mechanistic explanations of invasiveness. Such validation will not only confirm the adaptive importance of these genes but also aid in the potential of identifying genetic markers for the control and management of invasive populations. Finally, our findings reinforce the necessity of integrating molecular evolutionary approaches into the study of biological invasions. By doing so, we improve our capacity to prevent the behavior of invasive species, design more effectively control strategies, and predict ecological outcomes in a rapidly changing world.

Research limitations & future perspectives

Our study aimed to identify candidate genes potentially linked to invasive traits in two *Thunbergia* species, and to study their evolutionary dynamics by identifying signatures of positive selection in genes that may contribute to their adaptation to novel environments. Although this study presents a detailed and academic rigorous comparative molecular evolutionary analysis, it’s important to acknowledge certain limitations that constrain the reach of our findings. All analyses were conducted using previously assembled transcriptomes, which included vegetative and reproductive tissues across developmental stages of *T. alata*, and only tissue-type data for *T. grandiflora*, but lacked biological replicates. This may limit the statistical robustness and the detection of intra and interspecific variability. The unequal coverage of samples between species, particularly regarding developmental stages and analyzed tissues, may have also influenced the detection of co-expression modules in *T. grandiflora*. Moreover, the candidate genes identified through positive selection signals or through functional co-expressed modules haven’t undergone experimental validation, and thus their involvement in adaptive or invasive processes remains hypothetical. Some inconsistencies were also observed between signals of positive selection and gene expression levels, and mechanisms such as subfunctionalization or neofunctionalization might be playing a crucial role. Although, these require further functional validation. Additionally, genetic variability within each species wasn’t studied, which could be relevant for interpreting the evolutionary dynamics of the genes studied. Another important methodological consideration is that phylogenetic inference was conducted only under maximum likelihood frameworks, without exploring Bayesian inference methods. This could have provided a more exhaustive search of tree coherence. In addition, is particularly relevant given that ML methods are known to be susceptible to local optima, especially when initial conditions or search strategies aren’t clearly specified. Looking forward, future perspectives include the experimental validation of candidate genes, the extension of positive selection analyses using EasyCodeML to the 40 hub genes identified though co-expression, and the analyses of new transcriptomes such as those related to seed and fruit, will increase biological resolution to better characterize the molecular processes potentially associated with invasiveness in *Thunbergia* species.

Acknowledgments

I want to express my sincere gratitude to my director Pablo Pérez Mesa and co-director Harold Suárez Baron, for all the support, expert guidance, mentorship, and great academic advice during the realization and completion of this work. I also want to express my gratitude to my friend and project teammate David Ruiz Londoño for the guidance, teachings, and support during the fulfillment of this work and for his fundamental contributions to this project. I wish to acknowledge the Pontificia Universidad Javeriana Cali for providing me with the resources and supportive environment to carry out this research. We thank the Oficina de Investigación y Desarrollo (OID) for funding through the "Convocatoria Nuevos Doctores 2022" (project 020100810). I deeply appreciate the Eco-Evo-Devo research group for their support. Also, I especially thank my family, particularly my parents and my brother, for their love, trust, and encouragement, which have given me the strength to keep moving forward. Finally, but not least, I want to extend my gratitude to my friends and classmates who supported me and shared their time with me during all these years.

References

Adhikari, B., Pendry, C. A., Watson, M. F., & Wood, J. R. I. (2013). An account of *Thunbergia* (Acanthaceae) in Nepal, with a description of the new species *T. nepalensis*. Kew Bulletin, 68, 651-661.

Álvarez-Carretero, S., Kapli, P., & Yang, Z. (2023). Beginner's guide on the use of PAML to detect positive selection. Molecular biology and evolution, 40(4), msad041.

Cárdenas B, Baptiste M, Castaño N (2017) Plantas exóticas con alto potencial de invasión en Colombia. Instituto de Investigación de Recursos Biológicos Alexander Von Humboldt.

Cárdenas, D., Castaño, N., & Cárdenas-Toro, J. (2010). Análisis de riesgo de especies de plantas introducidas para Colombia. En: Baptiste M.P., Castaño N., Cárdenas D., Gutiérrez F. P., Gil D.L. y Lasso C.A. (Ed.), Análisis de riesgo y propuesta de categorización de especies introducidas para Colombia. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt y Agrosavia. (51-72).

Cai, M. L., Zhang, Q. L., Zhang, J. J., Ding, W. Q., Huang, H. Y., & Peng, C. L. (2020). Comparative physiological and transcriptomic analyses of photosynthesis in *Sphagneticola calendulacea* (L.) Pruski and *Sphagneticola trilobata* (L.) Pruski. Scientific Reports, 10(1), 17810.

Castri, D. (1989). History of biological invasions with special emphasis on the Old World. Biological invasions: a global perspective, 1-30.

Cauz-Santos, L. A., da Costa, Z. P., Sader, M. A., van den Berg, C., & Vieira, M. L. C. (2025). Chloroplast genomic insights into adaptive evolution and rapid radiation in the genus Passiflora (Passifloraceae). BMC Plant Biology, 25, 192.

Chassin, Y., Kapri-Pardes, E., Sinvany, G., Arad, T., & Adam, Z. (2002). Expression and characterization of the thylakoid lumen protease *DEGP1* from Arabidopsis. Plant Physiology, 130(2), 857-864.

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Pa, J., ... & Xia, R. (2023). TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Molecular plant, 16(11), 1733-1742.

Chen, L. Q., Qu, X. Q., Hou, B. H., Sosso, D., Osorio, S., Fernie, A. R., & Frommer, W. B. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science, 335(6065), 207-211.

Chown, S. L., Hodgins, K. A., & Griffin, P. C. (2016). Biological invasions, climate change, and genomics. Crop Breeding, 59-114.

Chuang, L., Liu, S., Biedermann, D., & Franke, J. (2022). Identification of early quassinoid biosynthesis in the invasive tree of heaven (Ailanthus altissima) confirms evolutionary origin from protolimonoids. Frontiers in Plant Science, 13, 958138.

Clements, D. R., & Ditommaso, A. (2011). Climate change and weed adaptation: can evolution of invasive plants lead to greater range expansion than forecasted?. Weed research, 51(3), 227-240.

Corporación Autónoma Regional de Cundinamarca (CAR). (2020). Plan de Prevención, Control y Manejo de las poblaciones de *Thunbergia alata* Bojer ex Sims para la jurisdicción de la Corporación Autónoma Regional de Cundinamarca CAR. Dirección de recursos naturales

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. Genetics, 209(4), 1235-1278.

Eom, J. S., Chen, L. Q., Sosso, D., Julius, B. T., Lin, I. W., Qu, X. Q., ... & Frommer, W. B. (2015). SWEETs, transporters for intracellular and intercellular sugar translocation. Current opinion in plant biology, 25, 53-62.

Feltus, F. A. (2014). Systems genetics: a paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. Plant Science, 223, 45-48.

Feng, Z., Zheng, Y., Jiang, Y., Pei, J., & Huang, L. (2024). Phylogenetic relationships, selective pressure and molecular markers development of six species in subfamily Polygonoideae based on complete chloroplast genomes.Scientific Reports, 14(1), 9783.

Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., & Ho, S. Y. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML.Ecology and evolution, 9(7), 3891-3898.

Guo, W., Liu, Y., Ng, W. L., Liao, P. C., Huang, B. H., Li, W., ... & Huang, Y. (2018). Comparative transcriptome analysis of the invasive weed Mikania micrantha with its native congeners provides insights into genetic basis underlying successful invasion. BMC genomics, 19, 1-17.

Hall, T., Biosciences, I., & Carlsbad, C. J. G. B. B. (2011). BioEdit: an important software for molecular biology. GERF bull biosci, 2(1), 60-61.

Hamelin, R. C., & Roe, A. D. (2020). Genomic biosurveillance of forest invasive alien enemies: A story written in code. Evolutionary Applications, 13(1), 95-115.

He, X., & Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics, 169(2), 1157-1164.

Hellmann, J. J., Byers, J. E., Bierwagen, B. G., & Dukes, J. S. (2008). Five potential consequences of climate change for invasive species. Conservation biology, 22(3), 534-543.

Hierro, J. L., Maron, J. L., & Callaway, R. M. (2005). A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. Journal of ecology, 93(1), 5-15.

Hodge, R. G., & Ridley, A. J. (2016). Regulating Rho GTPases and their regulators. Nature reviews Molecular cell biology, 17(8), 496-510.

Jayaraman, V., Toledo-Patiño, S., Noda-García, L., & Laurino, P. (2022). Mechanisms of protein evolution. Protein Science, 31(7), e4362.

Jia, T., Li, F., Liu, S., Dou, J., & Huang, T. (2021). *DnaJ* proteins regulate WUS expression in shoot apical meristem of *Arabidopsis*. Plants, 10(1), 136.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

Lallemand, T., Leduc, M., Landès, C., Rizzon, C., & Lerat, E. (2020). An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. Genes, 11(9), 1046.

Langfelder, P., Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559.

Lee, C. E. (2002). Evolutionary genetics of invasive species. Trends in ecology & evolution, 17(8), 386-391.

Le Roux, J., & Wieczorek, A. M. (2009). Molecular systematics and population genetics of biological invasions:
towards a better understanding of invasive species management. Annals of Applied Biology, 154(1), 1-17.

Liu, F., Ma, Z., Cai, S., Dai, L., Gao, J., & Zhou, B. (2022). *ATP-citrate lyase B* (*ACLB*) negatively regulates cell death and plant resistance to Verticillium wilt.

Lobstein, E., Guyon, A., Férault, M., Twell, D., Pelletier, G., & Bonhomme, S. (2004). The putative Arabidopsis homolog of yeast vps52p is required for pollen tube elongation, localizes to Golgi, and might be involved in vesicle trafficking. Plant physiology, 135(3), 1480-1490.

Mable, B. K. (2019). Conservation of adaptive potential and functional diversity: integrating old and new approaches. Conservation Genetics, 20(1), 89-100.

Marin, P., Genitoni, J., Barloy, D., Maury, S., Gibert, P., Ghalambor, C. K., & Vieira, C. (2020). Biological invasion: The influence of the hidden side of the (epi) genome. Functional Ecology, 34(2), 385-400.

Madhani, H., & Kourki, A. N. (2024). The Evolution of Complex Multicellularity in Land Plants. Genes, 15(11), 1472.

McNeely, J. A. (Ed.). (2001). Global strategy on invasive alien species. IUCN.

Mela, D., Dias, M. G., da Silva, T. I., da Silva Ribeiro, J. E., Martinez, A. C. P., & Zuin, A. H. L. (2022). Estimation of *Thunbergia grandiflora* leaf area from allometric models. Comunicata Scientiae, 13, e3722-e3722.

Meyerson, L. A., & Mooney, H. A. (2007). Invasive alien species in an era of globalization. Frontiers in Ecology and the Environment, 5(4), 199-208.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution, 37(5), 1530-1534.

Moseley, R. C., Mewalal, R., Motta, F., Tuskan, G. A., Haase, S., & Yang, X. (2018). Conservation and diversification of circadian rhythmicity between a model crassulacean acid metabolism plant *Kalanchoë fedtschenkoi* and a model C3 photosynthesis plant Arabidopsis thaliana. Frontiers in Plant Science, 9, 1757.

Musskopf, M. K., de Mattos, E. P., Bergink, S., & Kampinga, H. H. (2018). HSP40/DNAJ Chaperones. *eLS*, 1-11.

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2012). *NAC* transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA)*. Gene Regulatory Mechanisms, 1819(2), 97-103.

Nei M. and Gojobori T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3:418-426.

North, H. L., McGaughran, A., & Jiggins, C. D. (2021). Insights into invasive species from whole‐genome resequencing. Molecular Ecology, 30(23), 6289-6308.

Ovens, K., Eames, B. F., & McQuillan, I. (2021). Comparative analyses of gene co-expression networks: Implementations and applications in the study of evolution. Frontiers in Genetics, 12, 695399.

Parvizi, E., Dhami, M. K., Yan, J., & McGaughran, A. (2023). Population genomic insights into invasion success in a polyphagous agricultural pest, *Halyomorpha halys*. Molecular ecology, 32(1), 138-151.

Podzimska-Sroka, D., O'Shea, C., Gregersen, P. L., & Skriver, K. (2015). *NAC* transcription factors in senescence: from molecular structure to function in crops. Plants, 4(3), 412-448.

Prentis, P. J., Wilson, J. R., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. Trends in plant science, 13(6), 288-294.

Pyšek, P., & Richardson, D. M. (2010). Invasive species, environmental change and management, and health. Annual Review of Environment and Resources, 35, 25-55.

Qi, S., Wang, J., Zhang, Y., Naz, M., Afzal, M. R., Du, D., & Dai, Z. (2023). Omics approaches in invasion biology: understanding mechanisms and impacts on ecological health. Plants, 12(9), 1860.

Quijano-Abril, M. A., Castaño-López, M. D. L. Á., Marín-Henao, D., Sánchez-Gómez, D., Rojas-Villa, J. M., & Sierra-Escobar, J. (2021). Functional traits of invasive species *Thunbergia alata* (Acanthaceae) and its importance in the adaptation to Andean forests. Acta botánica mexicana, (128).

Quijano-Abril, M. A., Sierra Escobar, J. A., Gutierrez, B., Alzate, R., Sánchez-Gómez, D., Marín-Henao, D., & Rojas-Villa, J. (2019). Historia, vida y poderes de una especie invasora: estrategia para su control y manejo. Fondo Editorial Universidad Católica del Oriente, Cornare

Rambaut, A. (2018). FigTree: tree figure drawing tool (Version 1.4.4). Retrieved from.

Rao, X., & Dixon, R. A. (2019). Co-expression networks for plant biology: why and how. Acta biochimica et biophysica Sinica, 51(10), 981-988.

Ruan, J., Dean, A. K., & Zhang, W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. BMC systems biology, 4, 1-21.

Ruan, X., Wang, Z., Su, Y., & Wang, T. (2022). Full-length transcriptome analysis of multiple organs and identification of adaptive genes and pathways in Mikania micrantha. Scientific Reports, 12(1), 3272.

Sahm, A., Bens, M., Platzer, M., & Szafranski, K. (2017). PosiGene: automated and easy-to-use pipeline for genome-wide detection of positively selected genes. Nucleic Acids Research, 45(11), e100-e100.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., ... & Weller, S. G. (2001). The population biology of invasive species. Annual review of ecology and systematics, 32(1), 305-332.

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., ... & Sherry, S. T. (2021). Database resources of the national center for biotechnology information. Nucleic acids research,49(D1), D10-D17.

Schierenbeck, K. A., & Ainouche, M. L. (2006). The role of evolutionary genetics in studies of plant invasions. In Conceptual ecology and invasion biology: reciprocal approaches to nature (pp. 193-221). Dordrecht: Springer Netherlands.

Scholz, P., Anstatt, J., Krawczyk, H. E., & Ischebeck, T. (2020). Signalling pinpointed to the tip: the complex regulatory network that allows pollen tube growth. Plants, 9(9), 1098.

Serin, E. A., Nijveen, H., Hilhorst, H. W., & Ligterink, W. (2016). Learning from co-expression networks: possibilities and challenges. Frontiers in plant science, 7, 444.

Sherpa, S., & Després, L. (2021). The evolutionary dynamics of biological invasions: A multi-approach perspective. Evolutionary Applications, 14(6), 1463-1484.

Simberloff, D., Martin, J.-L., Genovesi, P., Maris, V., Wardle, D. A., Aronson, J., Courchamp, F., Galil, B., García Berthou, E., Pascal, M., Pyšek, P., Sousa, R., Tabacchi, E. & Vilà, M. (2013). Impacts of biological invasions: what's what and the way forward. Trends in Ecology & Evolution, 28(1), 58-66.

Smith, S. D., Pennell, M. W., Dunn, C. W., & Edwards, S. V. (2020). Phylogenetics is the new genetics (for most of biodiversity). Trends in Ecology & Evolution, 35(5), 415-425.

Soltis, P. S., Folk, R. A., & Soltis, D. E. (2019). Darwin review: angiosperm phylogeny and evolutionary radiations. Proceedings of the Royal Society B, 286(1899), 20190099.

Stachowicz, J. J., Terwin, J. R., Whitlatch, R. B., & Osman, R. W. (2002). Linking climate change and biological invasions: ocean warming facilitates nonindigenous species invasions. Proceedings of the National Academy of Sciences, 99(24), 15497-15500.

Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution.

Theoharides, K. A., & Dukes, J. S. (2007). Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. New phytologist, 176(2).

Vaughan, A. L., & Dhami, M. K. (2024). Can Transcriptomics Elucidate the Role of Regulation in Invasion Success? Molecular Ecology, e17583.

Wang, J., Tao, F., Marowsky, N. C., & Fan, C. (2016). Evolutionary fates and dynamic functionalization of young duplicate genes in Arabidopsis genomes. Plant Physiology, 172(1), 427-440.

Wilson, D. N., & Cate, J. H. D. (2012). The structure and function of the eukaryotic ribosome. Cold Spring Harbor perspectives in biology, 4(5), a011536.

Xu, C., Ge, Y., & Wang, J. (2019). Molecular basis underlying the successful invasion of hexaploid cytotypes of Solidago canadensis L.: Insights from integrated gene and miRNA expression profiling. Ecology and evolution, 9(8), 4820-4852.

Yang, Z., & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular biology and evolution, 19(6), 908-917.

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. Molecular biology and evolution, 22(4), 1107-1118.

Yang, Y., Qin, Y., Xie, C., Zhao, F., Zhao, J., Liu, D., ... & Guo, Y. (2010). The *Arabidopsis* chaperone J3 regulates the plasma membrane H+-ATPase through interaction with the PKS5 kinase. The Plant Cell, 22(4), 1313-1332.

Yin, T., Pan, G., Liu, H., Wu, J., Li, Y., Zhao, Z., ... & Zhou, Y. (2012). The chloroplast ribosomal protein *L21* gene is essential for plastid development and embryogenesis in Arabidopsis. Planta, 235, 907-921.

Yuan, J., Zeng, X. Y., Geng, K., Wijayawardene, N. N., Bhat, J. D., Wu, S. P., Wang, W., Yang, Z. F. (2021). *Allophoma* species (Pleosporales: Didymellaceae) associated with *Thunbergia grandiflora* in Guangxi Province, China. Biodiversity Data Journal, 9.

Zhang, Z. (2022). KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. Genomics, Proteomics and Bioinformatics, 20(3), 536-540.

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., & Wang, G. T. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular ecology resources, 20(1), 348-355.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology, 4(1).

Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular biology and evolution, 22(12), 2472-2479.

Zhong, R., Cui, D., Richardson, E. A., Phillips, D. R., Azadi, P., Lu, G., & Ye, Z. H. (2020). Cytosolic acetyl-CoA generated by ATP-citrate lyase is essential for acetylation of cell wall polysaccharides. Plant and Cell Physiology, 61(1), 64-75.

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: progress and challenges. International journal of biological sciences, 3(7), 420.

Zhu, Y., Tian, Y., Han, S., Wang, J., Liu, Y., & Yin, J. (2024). Structure, evolution, and roles of *SWEET* proteins in growth and stress responses in plants. International Journal of Biological Macromolecules, 130441.

Supplementary Material

Supplementary material including tables and figures is available at:
https://mega.nz/folder/FVBFwBYB#JzB0dCQ1mx0VM2HCB9WezQ

List of Supplementary Tables

Table S1 The sampling information in this study for the phylogenetic analysis and selection pressure analyses, using the genomic and transcriptomic information of these species in NCBI (https://www.ncbi.nlm.nih.gov/nucleotide/).

Table S2 Genes used as dataset for the targeted search employed in the candidate genes identification.

Table S3 Functional annotation for top 40 transcripts and DEGs in the target search of *T. alata* employed in the candidate genes identification.

Table S4 Functional annotation for top 40 transcripts and DEGs (20 upregulated 20 downregulated) in the developmental stage comparison (early vs. late) of *T. alata* employed in the candidate genes identification. Green codes represent upregulated DEGs, while red codes represent downregulated DEGs.

Table S5 Functional annotation for top 40 transcripts and DEGs (20 upregulated 20 downregulated) in the tissue type comparison (reproductive vs. vegetative) of *T. alata* employed in the candidate genes identification. Green codes represent upregulated DEGs, while red codes represent downregulated DEGs.

Table S6 Functional annotation for top 40 transcripts and DEGs (20 upregulated 20 downregulated) in the interspecific comparison between *T. alata* and *T. grandiflora* employed in the candidate genes identification. Green codes represent upregulated DEGs, while red codes represent downregulated DEGs.

Table S7 Functional annotation for top 40 transcripts and DEGs in co-expressed gene modules. Genes were categorized by the comparison: target search developmental stage, tissue type, and interspecific comparison of *T. alata* DEGs.

Table S8 Correlation of co-expressed candidate genes modules in *T. alata* transcriptome, based on a WGCNA direct search under the following comparison categories: target search, developmental stage, tissue type, and interspecific comparison between *T. alata* and *T. grandiflora.*

Table S9 Functional annotation for top 40 transcripts and DEGs (20 upregulated 20 downregulated) in the candidate gene identification for *T. alata*. Green codes represent upregulated DEGs, while red codes represent downregulated DEGs.

Table S10 Functional annotation for top 40 transcripts and DEGs (20 upregulated 20 downregulated) in the candidate gene identification for *T. grandiflora*. Green codes represent upregulated DEGs, while red codes represent downregulated DEGs.

Table S11 List of 32 candidate genes analyzed for selection pressure within the transcriptomes of *T. alata* and *T. grandiflora*. Gene classification and grouping were validated based on their protein information retrieved from NCBI (https://www.ncbi.nlm.nih.gov/nucleotide/) and UniProt (https://www.uniprot.org) databases.

Table S12 The ratio of non-synonymous substitution rate (*Ka* or *dN*) to synonymous substitution rate (*Ks* or *dS*), known as $\omega = Ka/Ks$ (or dN/dS) for positive selection in candidate genes using the site-specific model from EasyCodeML (Gao et al. 2019), and assuming the foreground branch in *T. alata*.

Table S13 The ratio of non-synonymous substitution rate (*Ka* or *dN*) to synonymous substitution rate (*Ks* or *dS*), known as $\omega = Ka/Ks$ (or dN/dS) for positive selection in candidate genes using the site-specific model from EasyCodeML (Gao et al. 2019), and assuming the foreground branch in *T. grandiflora*.

Table S14 Likelihood ratio test (LRT) results for positive selection in candidate genes using the site-specific model from EasyCodeML (Gao et al. 2019), and assuming the foreground branch in *T. alata*. A significant LRT statistic and *p*-value (≤ 0.05) indicate evidence for positive selection.

Table S15 Likelihood ratio test (LRT) results for positive selection in candidate genes using the site-specific model from EasyCodeML (Gao et al. 2019), and assuming the foreground branch in *T. grandiflora*. A significant LRT statistic and *p*-value (≤ 0.05) indicate evidence for positive selection.

Table S16 Likelihood ratio test (LRT) results for positive selection in candidate genes using the branch-site model from EasyCodeML (Gao et al. 2019), and assuming the foreground branch in *T. alata*. The null mode assumes no positive selection, whereas the alternative model allows for positive selection on the foreground branch. A significant LRT statistic and *p*-value (≤ 0.05) indicate evidence for positive selection.

Table S17 Likelihood ratio test (LRT) results for positive selection in candidate genes using the branch-site model from EasyCodeML (Gao et al. 2019), and assuming the foreground branch in *T. grandiflora*. The null mode assumes no positive selection, whereas the alternative model allows for positive selection on the foreground branch. A significant LRT statistic and *p*-value (≤ 0.05) indicate evidence for positive selection.

Table S18 Evolutionary and selection analyses of *ACLB1* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S19 Evolutionary and selection analyses of *BPM4* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S20 Evolutionary and selection analyses of *EIN3* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S21 Evolutionary and selection analyses of *PEPC* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S22 Evolutionary and selection analyses of *PR3* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S23 Evolutionary and selection analyses of *TADA2* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S24 Evolutionary and selection analyses of *VDAC2* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S25 Evolutionary and selection analyses of *AGT1* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S26 Evolutionary and selection analyses of *ANJ1* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).
Table S27 Evolutionary and selection analyses of *GAPDH* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S28 Evolutionary and selection analyses of *MFP2* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S29 Evolutionary and selection analyses of *NAC92* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S30 Evolutionary and selection analyses of *PBL8* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S31 Evolutionary and selection analyses of *POK* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S32 Evolutionary and selection analyses of *SR45* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S33 Evolutionary and selection analyses of *AED3* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S34 Evolutionary and selection analyses of *ALDHC4* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S35 Evolutionary and selection analyses of *DEGP1* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S36 Evolutionary and selection analyses of *DEGP1(2)* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S37 Evolutionary and selection analyses of *HEMC* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S38 Evolutionary and selection analyses of *MAH1* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S39 Evolutionary and selection analyses of *NCAPH2* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S40 Evolutionary and selection analyses of *PAP2* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S41 Evolutionary and selection analyses of *ROPGEF12* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S42 Evolutionary and selection analyses of *SEP3* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S43 Evolutionary and selection analyses of *CINV1* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S44 Evolutionary and selection analyses of *HIT4* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S45 Evolutionary and selection analyses of *PEMT* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S46 Evolutionary and selection analyses of *SPL* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S47 Evolutionary and selection analyses of *SYP61* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S48 Evolutionary and selection analyses of *VTC2* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S49 Evolutionary and selection analyses of *WNK1* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S50 Evolutionary and selection analyses of *WRKY* orthologous gene sequences using MEGA 11 (Tamura et al. 2021).

Table S51 Annotation hits by biological processes for *T. alata* co-expressed gene modules.

Table S52 Annotation hits by cellular component for *T. alata* co-expressed gene modules.

Table S53 Annotation hits by molecular function for *T. alata* co-expressed gene modules.

List of Supplementary Figures

Fig. S1 Hierarchical dendrogram clustering genes based on their expression patterns, where branch height indicates similarity. The colored bars at the bottom represent modules of co-expressed genes, aiding the identification of functional gene groups within the network.

Fig. S2 Evaluation of the soft-threshold power for constructing the co-expression network. The left plot shows that a power of around 6 achieves a good fit to scale-free topology (high R2), while the right plot indicates that mean connectivity decreases with higher powers, improving network specificity.

Fig. S3 Gene ontology (GO) classification and enrichment analysis focused on cellular component were conducted for potential candidate genes from co-expressed gene modules of *T. alata* transcriptome. These genes were categorized based on previous comparisons including targeted search, developmental stage, tissue type, and interspecific comparisons. The GO terms presented here correspond to those with more than one associated gene (hit) in the *T. alata* species.

Fig. S4 Gene ontology (GO) classification and enrichment analysis focused on molecular function were conducted for potential candidate genes from co-expressed gene modules of *T. alata* transcriptome. These genes were categorized based on previous comparisons including targeted search, developmental stage, tissue type, and interspecific comparisons. The GO terms presented here correspond to those with more than one associated gene (hit) in the *T. alata* species.