

Identification and Expression Analysis of the MADS-box Gene Family in *Camellia sinensis*

Jing Yang

Hainan Normal University

Hanxue Wei

Hainan Normal University

Guangzhen Zhou

gzzhou@hainanu.edu.cn

Hainan Normal University

Research Article

Keywords: *C. sinensis*, MADS-box gene family, Gene expression pattern, Abiotic stress

Posted Date: December 4th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7855824/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

A woody crop that has been important to the economy is *Camellia sinensis*. Tea plants may experience physiological and metabolic changes as a result of drought and high salinity. The *MADS-box* gene family is a significant of transcription factors essential for various facets of plant growth and development. However, comprehensive research on *MADS-box* genes in *C. sinensis* is still sparse. In this study, based on the genomic datas of *C. sinensis*, a total of 70 *MADS-box* genes were identified and phylogenetically categorized into two types Type I (which includes Ma, M β , My, and My, consisting of 26 members) and Type II (comprising MIKCC and MIKC*, with 44 members). Members of the same subfamily have very conserved exon-intron patterns and protein motifs, according to gene structure analysis. The initial detailed study of the *MADS-box* gene family expression profiles and evolutionary features of this gene in *C. sinensis* is presented in this paper. Moreover, qRT-PCR tests showed that these genes are frequently engaged in the reaction to salt stress and drought. These results not only offer valuable insights into the genetic mechanisms that contribute to tea plant diversity but also lay the groundwork for molecular breeding efforts focused on enhancing tea quality and aesthetic characteristics.

1. Introduction

Camellia sinensis belongs to the family Theaceae, and it is a significant economic crop of the genus *Camellia*. This plant is extensively cultivated throughout tropical and subtropical zones and is prized for the intense fragrance of the tea derived from its leaves^[1]. Higher plants are frequently exposed to environmental stressors during growth and development. Drought, low temperature, and high salinity are some of the most prevalent abiotic stresses that can elicit various physiological and metabolic responses in plants, hinder their metabolism and growth, and, in extreme cases, result in irreversible damage or even death.^[2].

The *MADS-box* gene family is widely found in plants, animals, and fungi, and plays a central role in multiple stages of plant development^[3–5]. This gene family includes members that control different parts of the plant growth and development process, such as the stress response, the length of the flowering period, the maintenance of the floral meristem, the formation of the floral organ, and the development of fruit^[6–7]. The *MADS-box* gene family has a core function in plant responses to abiotic stress, highlighting its role as a key integrator of environmental signals and developmental regulation. By modulating processes such as hormone signaling, ROS homeostasis, nutrient response, and flowering time, *MADS-box* genes help plants achieve survival and reproductive success in changing environments^[8]. The name of the *MADS-box* gene family derives from its conserved DNA-binding domain (the MADS domain), formed from the initials of four founder genes: M (MCM1, yeast), A (AGAMOUS, Arabidopsis), D (DEFICIENS, snapdragon), and S (SRF, human)^[9]. The *MADS-box* genes may be separated into the two main classes according to protein domain architecture and phylogenetic relationships: Type I Ma–M β –My and Type II. (C) e II gene are specific to plants and contain, besides the MADS domain, three additional domains: I (intervening), K-box (keratin-like), and a variable C-terminal

domain(C). They are additionally categorized into MIKCC and MIKC*subtypes^[10]. With known expression patterns or mutant characteristics, the majority are made up of the MIKCC kind, which also includes all of the plant *MADS-box* genes. *MIKCC-type* genes can be divided into 12 different clades according to phylogeny: AG-like, AGL12-like, AGL15-like, AGL17-like, AGL6-like, AP1-like, AP3/PI-like, Bsister, FLC-like, SEP-like, SOC1-like, and SVP-like. Type I genes evolve more rapidly and have relatively diverse functions, whereas Type II genes are more conserved in floral organ development and form the genetic basis of the classical ABC(D)E model^[11–12].

Yifei Mou et al. were the first to comprehensively identify the MIKC-type *MADS-box* gene family in peanuts, revealing its phylogeny, structural characteristics, expression patterns, and potential functions in abiotic stress responses. Their results indicated that peanut *MADS-box* genes may be involved in various stress adaptation processes through hormone signaling pathways and stress-responsive elements^[13]. Xue Y et al. first identified a MIKCC-class MADS transcription factor gene, GhFYF, in cotton. They discovered that it is concentrated in the nucleus and is highly expressed in flowers and seeds. This gene is activated by drought and salt stress; its overexpression improves salt tolerance in *Arabidopsis*, boosts seed germination and proline accumulation, and interacts with the stress-responsive protein GhGPP2. These data show that GhFYF may be important in abiotic stress responses, especially salt stress, in cotton^[14]. Lu J et al. systematically identified the *MADS-box* gene family in flax and clarified the key functions of these genes in floral organ development and stress response. Their study further revealed that Type II genes are primarily involved in floral organ development, whereas Type I genes are more associated with embryonic development. In addition, multiple LuMADS genes showed significant expression changes under cold, drought, and salt stress conditions, among which LuMADS27 exhibited regulatory roles under multiple stresses^[15]. Although abiotic stresses significantly impact *C. sinensis*, the role of the *MADS-box* gene family remains poorly characterized. To fill this knowledge gap, this study provides a genome-wide systematic analysis. Firstly, we identified all *MADS-box* genes and analyzed their phylogenetic evolution, gene structure, and conserved motifs. Additionally, RNA-seq and qRT-PCR were employed to determine their expression patterns under different abiotic stresses. The findings will contribute to elucidating the molecular regulatory mechanisms of *C. sinensis* under stress and provide a theoretical basis and candidate genes for molecular breeding and genetic improvement of tea plants.

2. Results

2.1 Identification of *MADS-box* gene family members in *C. sinensis*.

To ensure a comprehensive screening of *MADS-box* gene family members, putative *MADS-box* genes were identified by combining the results of Hidden Markov Model (HMM) and BLAST search. Our genome-wide analysis identified 70 members of the *MADS-box* gene family in *C. sinensis*. These proteins exhibited different lengths, with the longest being 1358 aa (*CaS14G024890.1*) and the shortest being 100 aa (*CaS05G005620.1*) (Fig. 1). Correspondingly, *CaS14G024890.1* had the heaviest molecular weight

(MW) at 154041.95 Da, while *CaS05G005620.1* had the lightest protein at 11721.28 Da. The isoelectric point (PI) of the protein varied from 4.76 (*CaS07G000990.1*) to 10.99 (*CaS05G005610.1*), with 26 members of the *MADS-box* family exhibiting an acidic isoelectric point below 7, while the remaining 44 proteins displayed a PI greater than 7, classifying them as basic (Fig. 1). According to the instability index, only three members, *CaS13G020990.1*, *CaS03G014820.1*, and *CaS05G002930.1*, were considered stable proteins (value less than 40), and the remaining 67 were unstable. The variation in the aliphatic amino acid index (A.I.), from 66.32 in *CaS13G020990.1* to 102.16 in *CaS13G013420.1*, demonstrates significant differences in protein thermal stability (Fig. 1). The total average hydrophobicity scores (GRAVY) of the proteins were all negative, implying that the detected *MADS-box* family members all encode hydrophilic proteins (Fig. 1). Based on the results of subcellular localization prediction, the primary location of the proteins was found to be the nucleus, which was followed by the cytoplasm and chloroplasts. Five members *CaS14G013270.1*, *CaS06G034490.1*, *CaS08G023160.1*, *CaS08G023340.1*, and *CaS05G008260.1* are located in the mitochondria, and two members *CaS06G011830.1* and *CaS06G014230.1* are located in the plasma membrane (Fig. 1).

2.2 Phylogenetic analysis of *MADS-box* genes of *C. sinensis* var. *assamica* var.

A phylogenetic analysis was performed to determine the relationships and subfamilies of the *C. sinensis* *MADS-box* proteins. The tree was constructed in MEGA 11 using the neighbor-joining method (Fig. 2A). The tree was based on the results of the comparison of *MADS-box* gene proteins, i.e., 108 sequences from *Arabidopsis thaliana* and 70 sequences from *C. sinensis*. Based on earlier research on the phylogenetic tree, the *MADS-box* gene family from *C. sinensis* includes 70 members were categorized into two groups, namely: type I, consisting of 26 members (8 Ma, 16 Mb, and 2 My), whereas the remaining 44 members were classified as type II (Fig. 2B). Within the type II group, only one *MADS-box* gene protein belonged to the MIKC* category, whereas the remaining members were categorized as MIKC^C, and the MIKC^C proteins were further classified into 12 branches, in all of which sequences could be identified in both *C. sinensis* and *Arabidopsis thaliana* (Fig. 2C). Phylogenetic analysis revealed that the majority of type II *MADS-box* genes in *C. sinensis* were located within the SOC1-like branch (9 members), while the AGL17-like and AGL15-like branches contained only one member each.

2.3 Examination of the structure, conserved structural domains and cis-acting elements of the *MADS* gene of *C. sinensis* var. *assamica*.

An analysis of conserved motifs and exon-intron structures was conducted to explore the structural diversity and evolutionary relationships of the *MADS-box* genes in *C. sinensis* (Fig. 3). From the results, it can be observed that all 70 members of the *C. sinensis* *MADS-box* gene family contain coding sequences (CDS); however, untranslated regions (UTRs) are not present in all members. Furthermore, an investigation into the conserved motifs was conducted using the MEME search tool. This analysis revealed a total of ten conserved motifs across the 70 *MADS-box* proteins. It was found that members within the same phylogenetic group generally exhibited similar motif structures, a pattern that was

particularly evident in the SOC1-like and B-sister groups. Nevertheless, the number and distribution of these motifs varied significantly among different *C. sinensis MADS-box* proteins, ranging from a maximum of seven to a minimum of only one. In addition to motifs, a total of 15 distinct types of structural domains were predicted among the *MADS-box* family members. Analysis showed that no single domain was universally present in all members. Among these domains, the *MADS_MEF2_like* and the *K-box* were the most abundant in terms of their occurrence, which may suggest that these two domains are more highly conserved within the *C. sinensis MADS-box* gene family.

The cis-element species and distribution may partially imply the transcriptional regulation and expression patterns of the genes involved. To elucidate the biological pathways in which the members of the *C. sinensis MADS-box* gene family may be involved, a prediction and analysis of the promoter cis-acting elements of these gene family members were performed (Fig. 5). A total of 34 cis-elements were identified using the PlantCARE database and classified into four categories (Fig. 6): light-responsive (17), phytohormone-responsive (9), stress-responsive (5), and plant growth-regulated (3). Most of the genes exhibited high proportions of light-responsive progenitors, followed by phytohormone-responsive elements, whereas elements related to plant growth regulation were the most sparse. The distribution of cis-acting elements suggests that the expression of *MADS-box* genes in *C. sinensis* is significantly influenced by light. No single element type was found in all members. The most prevalent light-responsive elements were Box 4 (171), G-box (112), and GT1-motif (74). A substantial fraction of elements were phytohormone-related, with ABRE (abscisic acid; 112), CGTCA-motif (methyl jasmonate; 65), TGACG-motif (methyl jasmonate; 64), and TCA-element (salicylic acid; 52) being the most abundant. Notably, certain elements were concentrated on specific genes; for instance, a high number of ABRE elements in *CaS12G015340.1* and *CaS03G004360.1* implies their potential importance in corresponding hormone regulation. Elements associated with abiotic stress responses, including ARE (anaerobic induction; 166), MBS (drought stress; 51), TC-rich repeats (defense and stress; 28), and LTR (low temperature; 21), were also identified. Among the 70 members, *CaS04G016390.1* possessed the greatest variety and number of elements (18 types, 36 total), whereas *CaS07G002960.1* and *CaS08G023160.1* had the fewest (5 and 6, respectively). Based on these results, it is hypothesized that *C. sinensis MADS-box* genes are involved in light and stress responses, potentially influencing flowering and stress tolerance pathways.

2.4 Chromosomal localization and covariance analysis of the *MADS* gene in *Camellia*

Chromosomal localization and gene covariance analysis of *MADS-box* genes in *C. sinensis* was performed based on the genome annotation files (Fig. 6). The results indicated that the 70 *MADS-box* gene family members were unevenly distributed across 15 chromosomes, with no genes located on chromosome 9. The distribution was concentrated on chromosomes 4, 5, 6, 12, 13, and 14, with chromosome 5 harboring the highest number. In contrast, chromosomes 1, 2, 3, 7, 8, 10, 11, and 15 contained relatively few members. Specifically, only one gene was found on chromosomes 1 and 11, while two genes were located on each of chromosomes 7 and 10. To investigate the evolutionary

relationships, covariance analysis was conducted using MCScanX in TBtools. The analysis identified six tandem duplicated gene pairs, indicated by red curves, on chromosomes 2, 5, 6, 13, and 14. These collinear genes may regulate core biological processes in *C. sinensis*, such as flowering time, floral organ development, or secondary metabolism. Their selective retention during evolution suggests that these genes possess redundant or complementary functions, potentially enhancing the plant's adaptability to environmental stresses. Fewer covariant genes in the *MADS-box* of *C. sinensis* may reflect the tendency of the family to replicate in tandem or to experience stronger selective pressures in *C. sinensis*.

Collinearity analysis between *C. sinensis* and other species revealed 18, 36, 2, and 7 collinear gene pairs with *Arabidopsis thaliana*, grape, maize, and rice, respectively (Fig. 7). These results indicate that *C. sinensis* shares the highest degree of collinearity with grape (36 pairs), substantially more than with *A. thaliana* (18 pairs), rice (7 pairs), or maize (2 pairs). A plausible explanation is that tea plant and grape share a closer phylogenetic relationship as core eudicots. The grape genome, which has experienced fewer ancient polyploidization events, likely retains a gene complement more similar to that of tea plant. In addition to this, both grape and *C. sinensis* are perennial woody plants that may share certain *MADS-box* gene-regulated developmental pathways. As a model plant, *Arabidopsis thaliana*, which has a more well-studied *MADS-box* gene function, has a lower number of covariates than grapes, probably because *Arabidopsis thaliana* has experienced more frequent genome rearrangements or lineage-specific gene loss. The very low number of collinear genes between *C. sinensis* and rice/maize can be attributed to their evolutionary divergence. As monocotyledons, rice and maize diverged from dicotyledonous plants like *C. sinensis*. This separation was followed by independent genome duplication events in their respective lineages, which weakened the detectable signals of collinear genes. Consequently, some *MADS-box* genes in monocots may have undergone functional divergence or been lost.

2.5 Analysis of Quantitative qRT-PCR Results

To elucidate the expression patterns of the *MADS-box* gene family under abiotic stress in tea plants (*C. sinensis*), seven selected genes were analyzed following six-day drought and salt stress treatments. The results showed that under drought stress, the expression levels of *CaS14G031860.1*, *CaS01G012140.1*, *CaS13G024900.1*, and *CaS04G016390.1* changed significantly over time. Among them, *CaS14G031860.1*, *CaS01G012140.1*, and *CaS13G024900.1* were significantly upregulated on day 3, with expression levels 3–4 times higher than those of the control group. In contrast, the expression of *CaS04G016390.1* increased continuously with prolonged drought treatment, reaching a level 4 times higher than the control by day 6 (Fig. 8A).

Under salt stress, *CaS05G010120.1* and *CaS10G020060.1* exhibited the most pronounced upregulation on day 3, with expression levels 3 times that of the control. Meanwhile, the expression of *CaS14G031860.1*, *CaS01G012140.1*, *CaS10G020070.1*, and *CaS04G016390.1* increased steadily over time and peaked on day 6. A notable observation was the significant downregulation of *CaS13G024900.1* in response to salt stress (Fig. 8B).

3. Discussion

In this study, a genome-wide analysis of *C. sinensis* identified 70 *MADS-box* gene family members. Based on the classification system established for *Arabidopsis thaliana*, these members were categorized into 26 type I (including Ma, M β , and My) and 44 type II (including MIKCC and MIKC*) genes. This number is similar to that found in closely related species, such as oil tea, which possesses 68 *MADS-box* genes^[16]. The significant numerical disparity between types likely reflects greater functional complexity and evolutionary constraints in Type II genes. Currently, 103, 87, 117, 81 and 95 members of the *MADS-box* gene family have been identified in *Arabidopsis thaliana*, dove, yellow early four, quinoa, vermilion rhododendron and chilli pepper, respectively, and the differences in the number of *MADS-box* genes among the species may be due to their genome sizes and whole-genome replication levels resulting from the differences in their genome size and genome-wide replication levels^[17].

Among Type II subfamilies, SOC1-like contained the most members (9), consistent with its conserved role in flowering time regulation, suggesting its critical function in tea phenology^[18]. Conversely, AGL17-like and AGL15-like each contained only one member, aligning with *Oryza sativa* AGL17-like involvement in root development, possibly indicating functional specialization within *C. sinensis* ^[19]. Physicochemical property analysis revealed substantial variation in protein sequence length, isoelectric point (pI), and hydrophilicity among members, potentially reflecting functional diversification.

Gene structure analysis revealed that members within the same subfamily exhibit highly conserved exon-intron patterns and protein motifs. Notably, the *MADS_MEF2_like* and *K-box* domains were ubiquitous among the family members, leading to the hypothesis that they are functionally crucial. As one of the most extensively studied plant transcription factor families, the *MADS-box* genes are known to play vital roles in growth, development, stress response, and secondary metabolism. Consistent with this, our analysis of promoter cis-acting elements in *C. sinensis* indicates that the expression of its *MADS-box* genes is likely regulated by light, phytohormones, and abiotic stresses.

The *MADS-box* genes exhibited an uneven dispersion across the 15 chromosomes of *C. sinensis*, mainly concentrated on chromosomes 4, 5, 6, 12, 13 and 14, and the covariance analysis showed that *C. sinensis* had the most covariant gene pairs with grapes (36 gene pairs), which was much higher than that with *Arabidopsis thaliana* (18 gene pairs), which was probably due to the fact that *C. sinensis* and grapes are dicotyledonous woody plants and have a similar retention pattern of the genome.

This work provides the first systematic analysis of the *MADS-box* gene family in *C. sinensis*, elucidating its evolutionary features and expression patterns, and thereby offering a theoretical foundation for understanding the genetic mechanisms underlying the plant's morphological diversity. Compared to model plants, the *MADS-box* family in *C. sinensis* exhibits both functional conservation (e.g., the organ-identity functions of B- and C-class genes) and divergence (e.g., the reduction of AGL17-like members). These evolutionary dynamics may be associated with its perennial woody growth habit and specific environmental adaptations. Future research should employ CRISPR-Cas9 editing or transgenic approaches to functionally validate key genes, providing candidate targets for tea molecular breeding.

Additionally, promoter cis-element predictions suggest *MADS-box* involvement in abiotic stress responses, opening new avenues for studying tea plant resilience mechanisms.

4. materials and methods

4.1 Studying species genomic data

The whole genome sequence, protein sequence, and annotation files of *C. sinensis* were obtained from the Tea Plant Genome Database (<https://eplant.njau.edu.cn/tea/index.html>). The corresponding files for the *Arabidopsis thaliana* *MADS-box* gene family were acquired from the TAIR database (<https://www.arabidopsis.org/>).

4.2 Identification of the *MADS-box* gene family members in *C. sinensis*

To identify the *MADS-box* genes in *C. elegans*, two different strategies were used: BLAST search versus Hidden Markov Model (HMM) search. First, the gene IDs of known *Arabidopsis thaliana* *MADS-box* genes were obtained from previous studies. Subsequently, their corresponding protein sequences were extracted using TBtools as queries for a BLASTP search against the *C. sinensis* genome. Finally, only homologous sequences with E-values less than 1e-5 were retained as candidate genes for subsequent analysis. To identify sequences containing typical *MADS-box* or *K-box* domains, the corresponding Hidden Markov model (HMM) profiles (PF00319, PF01486) were downloaded from the PFAM database (<https://pfam-legacy.xfam.org/>) and employed in a domain search. The final *MADS-box* genes were obtained by merging the sequences identified by the BLAST search with the HMM search and removing redundant members. The subcellular localization of *MADS-box* proteins was predicted using the online tool WoLF PSORT (<https://wolfsort.hgc.jp/>). Meanwhile, the physicochemical properties of these proteins were analyzed with the ProtParam-based “Protein Parameter Calc” function in TBtools.^[20]

4.3 Phylogenetic analysis of the *MADS* gene of *C. sinensis*

A multiple sequence alignment of all identified *MADS-box* protein sequences from *C. sinensis* was conducted using MUSCLE v3.8.^[21] First, the protein sequences of the *C. sinensis* *MADS-box* gene family were merged with those from *Arabidopsis thaliana*. Subsequently, the optimal model for phylogenetic construction was determined using the Model Selection tool built into MEGA11. Finally, a phylogenetic tree was constructed with the Neighbor-Joining (NJ) method, applying the Gamma-distributed rate variation model (+G)^[22]. After the evolutionary tree construction was completed the tree files generated by MEGA were beautified using Evolview v3(<https://www.evolgenius.info/evolview/#/treeview>) by annotating the type I (M α , M β , M γ) and type II (MIKCC, MIKC*, etc.) genes with different colours^[23].

4.4 Analysis of the structure, conserved structural domains and cis-acting elements of the *MADS* gene of *C. sinensis*

Identification of conserved motifs was performed using the MEME online tool (<https://meme-suite.org/meme/index.html>), configured to search for 10 motifs, with subsequent visualization conducted using the relevant function in TBtools^[24]. For structural domain prediction, protein sequences were submitted to NCBI CDD (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>), and the resulting data were visualized^[25]. Investigation of potential regulatory elements involved the extraction of a 2000 bp putative promoter region upstream of the translation initiation codon for each gene from the genomic files using TBtools. Cis-acting elements within these promoter sequences were identified via the PlantCARE database (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) and visualized using TBtools^[26].

4.5 Genome-Wide Identification of *MADS-box* Genes: Chromosome Distribution and Synteny in *C. sinensis*

The chromosomal locations of the *MADS-box* genes were determined from the *C. sinensis* genome annotation file and visualized using the "Gene Location Visualize" function in TBtools. The One Step MCScan X-Super Fast function of the TBtools software was then used to analyse the covariance of the *MADS-box* gene family members within the *C. sinensis* species. A similar procedure was used to demonstrate the distribution and covariance of *MADS-box* homologous genes on different chromosomes by going to the Ensembl Plants database to retrieve the genomic data of grapevine, *Arabidopsis thaliana*, rice, and maize for inter-species covariance with the *MADS-box* genes of *C. sinensis*^[27].

4.6 RNA Extraction, Reverse Transcription, and Quantitative RT-PCR Analysis in Response to Simulated Drought and Salt Stress

Four potted large-leaf tea plants (*C. sinensis*) from Wuzhishan, Hainan, with uniform growth status were selected and divided into two groups. One group was irrigated with 20% PEG6000 to simulate drought stress, while the other was treated with 200 mmol/L NaCl for salt stress. Leaves were collected on day 0, day 3, and day 6 to form the following groups: CK, PEG-3Day, PEG-6Day, NaCl-3Day, and NaCl-6Day. Upon harvest, all leaf samples were promptly wrapped in aluminum foil, flash-frozen in liquid nitrogen, and stored at -80°C.

Extraction of total RNA from all collected samples was performed using an RNA extraction kit (Tiangen Biotech, Beijing, China) in accordance with the manufacturer's instructions. Following an assessment of RNA purity, reverse transcription was conducted using the FastKing One-Step cDNA Synthesis PreMix (KR118). Quantitative real-time PCR amplification was executed on a LightCycler® 480 II system, and the relative expression levels of seven selected Type II *MADS-box* genes were determined by the $2^{-\Delta\Delta C_t}$ method. The corresponding quantitative primers are provided in Supplementary Table 1.

Declarations

Acknowledgements

This work was supported by the Hainan Province Science and Technology Special Fund [ZDYF2024KJTPY026].

Author contributions

J conducted the majority of the experimental work and drafted the manuscript. J and HX contributed to the experimental work. GZ supervised this study.

Data availability

The authors confirm that the data supporting the findings of this study are available within the Supplementary Materials.

Ethics declarations

Competing interests

The authors declare no competing interests.

Ethical approval and consent to participate

Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sánchez M, González-Burgos E, Iglesias I et al (2020) The Pharmacological Activity of *Camellia sinensis* (L.) Kuntze on Metabolic and Endocrine Disorders: A Systematic Review. *Biomolecules* 10(4):603
2. Wan S, Liang B, Yang L et al (2023) The MADS-box family gene *PtrANR1* encodes a transcription activator promoting root growth and enhancing plant tolerance to drought stress. *Plant Cell Rep* 43(1):16
3. Ng M, Yanofsky M (2001) Function and evolution of the plant *MADS*-box gene family. *Nat Rev* 2(3):186–195
4. Saedler H, Becker A, Winter KU, Kirchner C, Theissen G (2001) *MADS*-box genes are involved in floral development and evolution. *Acta Biochim Pol* 48(2):351–358
5. Smaczniak C, Immink RG, Angenent GC, Kaufmann K (2012) Developmental and evolutionary diversity of plant *MADS*-domain factors: insights from recent studies. *Development* 139(17):3081–3098. <https://doi.org/10.1242/dev.074674>
6. Theissen G (2001) Development of floral organ identity: stories from the MADS house. *Curr Opin Plant Biol* 4(1):75–85

7. Becker A, Theissen G (2003) The major clades of *MADS*-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol* 29(3):464–489
8. Castelán-Muñoz N, Herrera J, Cajero-Sánchez W et al (2019) *MADS*-Box Genes Are Key Components of Genetic Regulatory Networks Involved in Abiotic Stress and Plastic Developmental Responses in Plants. *Front Plant Sci* 10:853
9. de PARENICOVÁL S, KIEFFER M et al (2003) Molecular and phylogenetic analyses of the complete *MADS*-box transcription factor family in *Arabidopsis*: new openings to the *MADS* world. *Plant Cell* 15(7):1538–1551
10. RIECHMANN JL, KRIZEK BA, MEYEROWITZ EM (1996) Dimerization specificity of *Arabidopsis* *MADS* domain homeotic proteins *APETALA1*, *APETALA3*, *PISTILLATA*, and *AGAMOUS*. *Proceedings of the National Academy of Sciences of the United States of America* 93(10): 4793–4798
11. Gramzow L, Theissen G (2010) A hitchhiker's guide to the *MADS* world of plants. *Genome Biol* 11(6):1–11
12. Becker A, Theissen G (2003) The major clades of *MADS*-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol* 29(3):464–489
13. Mou Y, Yuan C, Sun Q, Yan C et al (2022) MIKC-type *MADS*-box transcription factor gene family in peanut: Genome-wide characterization and expression analysis under abiotic stress. *Front Plant Sci* 13:980933
14. Xue Y, Ma L, Wang H et al (2022) The *MADS* transcription factor GhFYF is involved in abiotic stress responses in upland cotton (*Gossypium hirsutum* L). *Gene* 815:146138
15. Lu J, Wu H, Pitt DM et al (2024) Identification and characterization of *MADS*-box gene family in flax, *Linum usitatissimum* L. and its role under abiotic stress. *iScience* 27(12):111092
16. Zhang Y, Li X, Yang H et al (2022) Genome-wide identification and expression analysis of *MADS*-box genes in *Camellia oleifera*. *Horticulturae* 8:102
17. Chen J, Yang Y, Li C et al (2023) Genome-Wide Identification of *MADS*-Box Genes in *Taraxacum kok-saghyz* and *Taraxacum mongolicum*: Evolutionary Mechanisms, Conserved Functions and New Functions Related to Natural Rubber Yield Formation. *Int J Mol Sci* 24(13):10997
18. Lee H, Suh SS et al (2000) The *AGAMOUS*-LIKE 20 *MADS* domain protein integrates floral inductive pathways in *Arabidopsis*. *Genes Dev* 14(18):2366–2376
19. Yu LH, Miao et al (2014) *MADS*-box transcription factor *AGL21* regulates lateral root development and responds to multiple external and physiological signals. *Mol Plant* 7(11):1653–1669
20. Chen C, Chen H, Zhang Y et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 13:1194–1202
21. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797
22. Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol Biol Evol* 38(7):3022–3027

23. Subramanian B, Gao S, Lercher et al (2019) Evolview v3: A webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res* 47(W1):W270–W275
24. Bailey TL, Johnson J, Grant CE et al (2015) The MEME Suite. *Nucleic Acids Res* 43:W39–W49
25. Marchler-Bauer A, Lu S, Anderson JB et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39:D225–D229
26. Lescot M, Déhais P, Thijs G et al (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 30:325–327
27. Andrew D, Yates JA et al (2022) Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res* 50(1):D996–D100

Figures

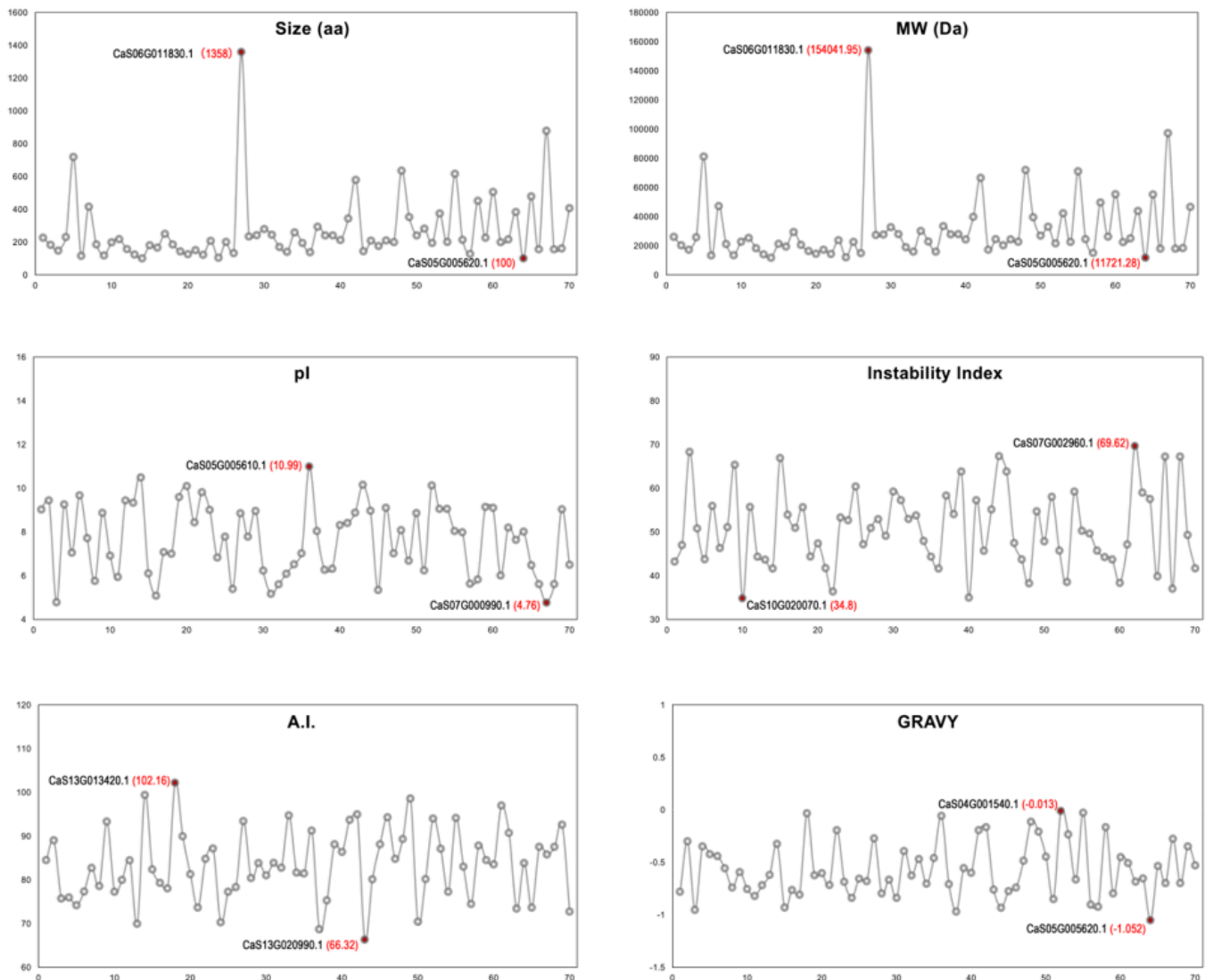


Figure 1

Physicochemical properties of the *MADS-box* gene family members of *C. sinensis* var. *assamica* var.

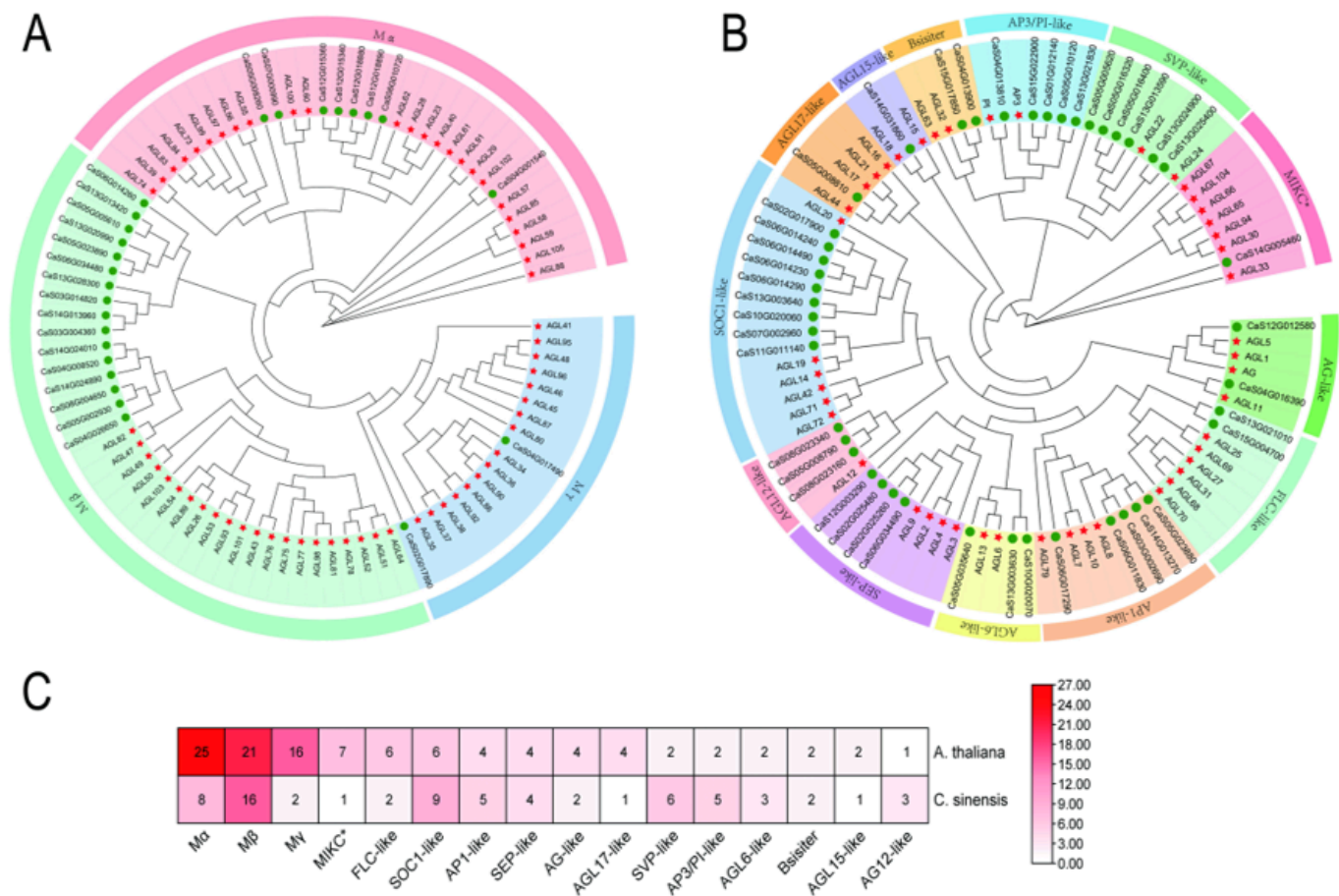


Figure 2

Phylogenetic tree of *MADS-box* genes of *C. sinensis* var. *assamica* and *Arabidopsis thaliana*.
(A) Phylogenetic tree of Type I *MADS-box* genes of *C. sinensis* var. *assamica* and *Arabidopsis thaliana*.
(B) Phylogenetic tree of Type II *MADS-box* genes of *C. sinensis* var. *assamica* and *Arabidopsis thaliana*.
(C) Heatmap of *MADS-box* gene holdings in *C. sinensis* var. *assamica* and *Arabidopsis thaliana*

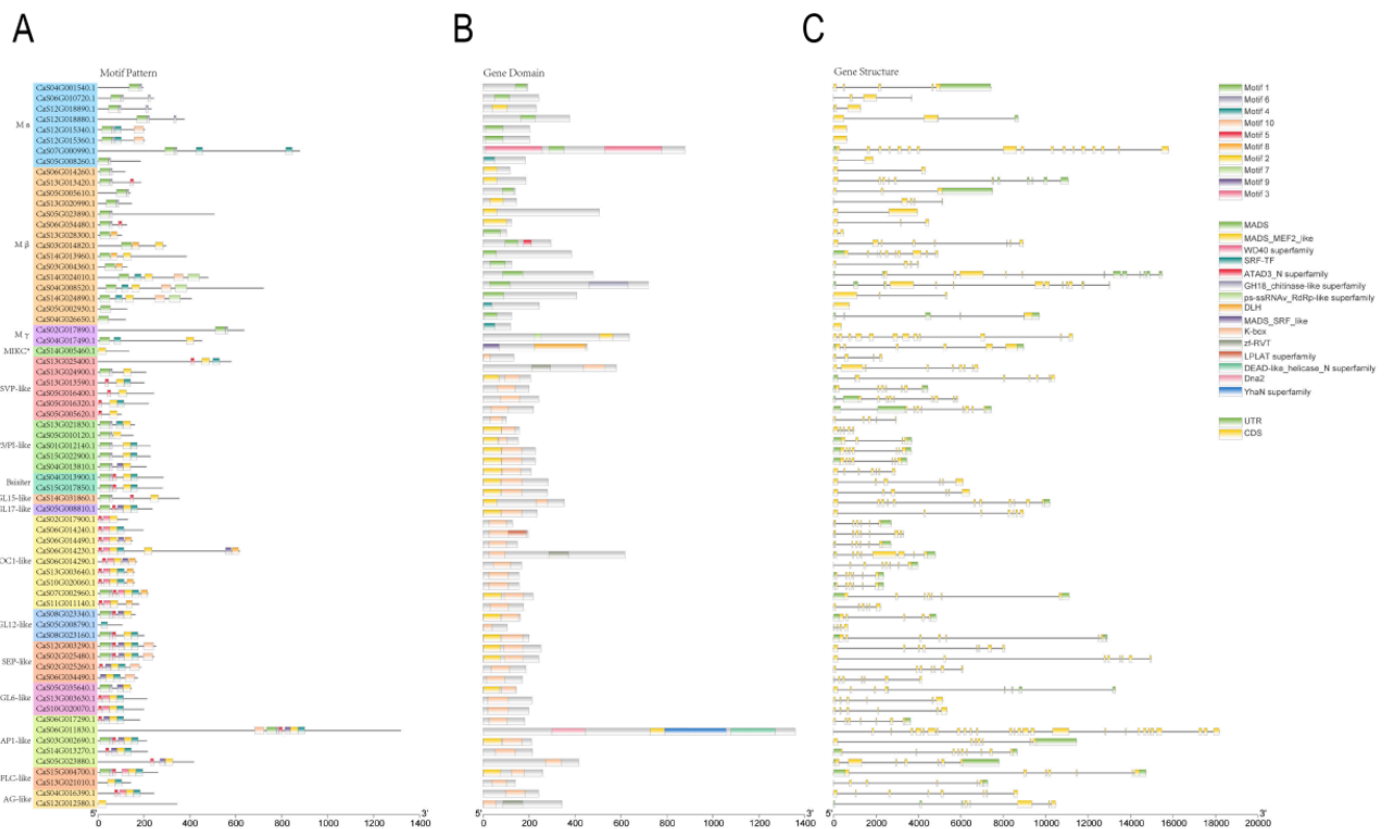


Figure 3

Structure and conserved motif analysis of the *MADS-box* gene of *C. sinensis* var. *Assamica*. (A) Distribution of conserved motifs in *MADS-box* gene proteins. (B) *MADS* domain, K-box domain, and other domains in *MADS-box* gene. (C) The gene structure of *MADS-box* includes untranslated regions (UTRs) and coding sequence (CDS).



Figure 4

Conserved structural domains of the *MADS-box* gene of *C. sinensis* var. *assamica*

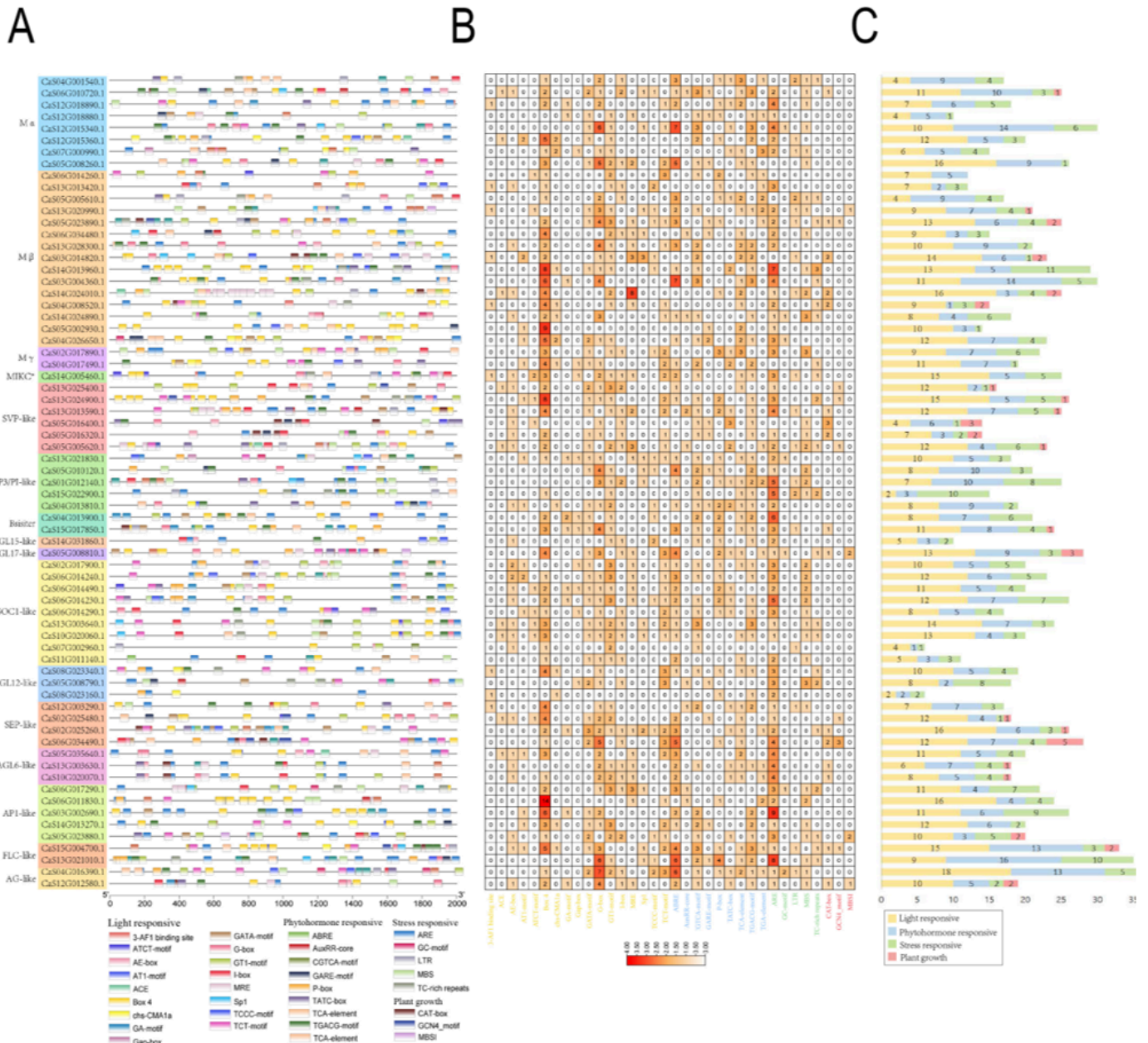


Figure 5

Predictive analysis of cis-acting elements of the promoter of the *MADS-box* gene of *C. sinensis* var. *Assamica*. (A) Analysis of cis-regulatory elements in the promoter regions of *Camellia sinensis* var. *assamica* *MADS-box* genes. (B) Heatmap of cis-regulatory element counts. (C) Bar chart showing aggregate counts of cis-regulatory elements by category

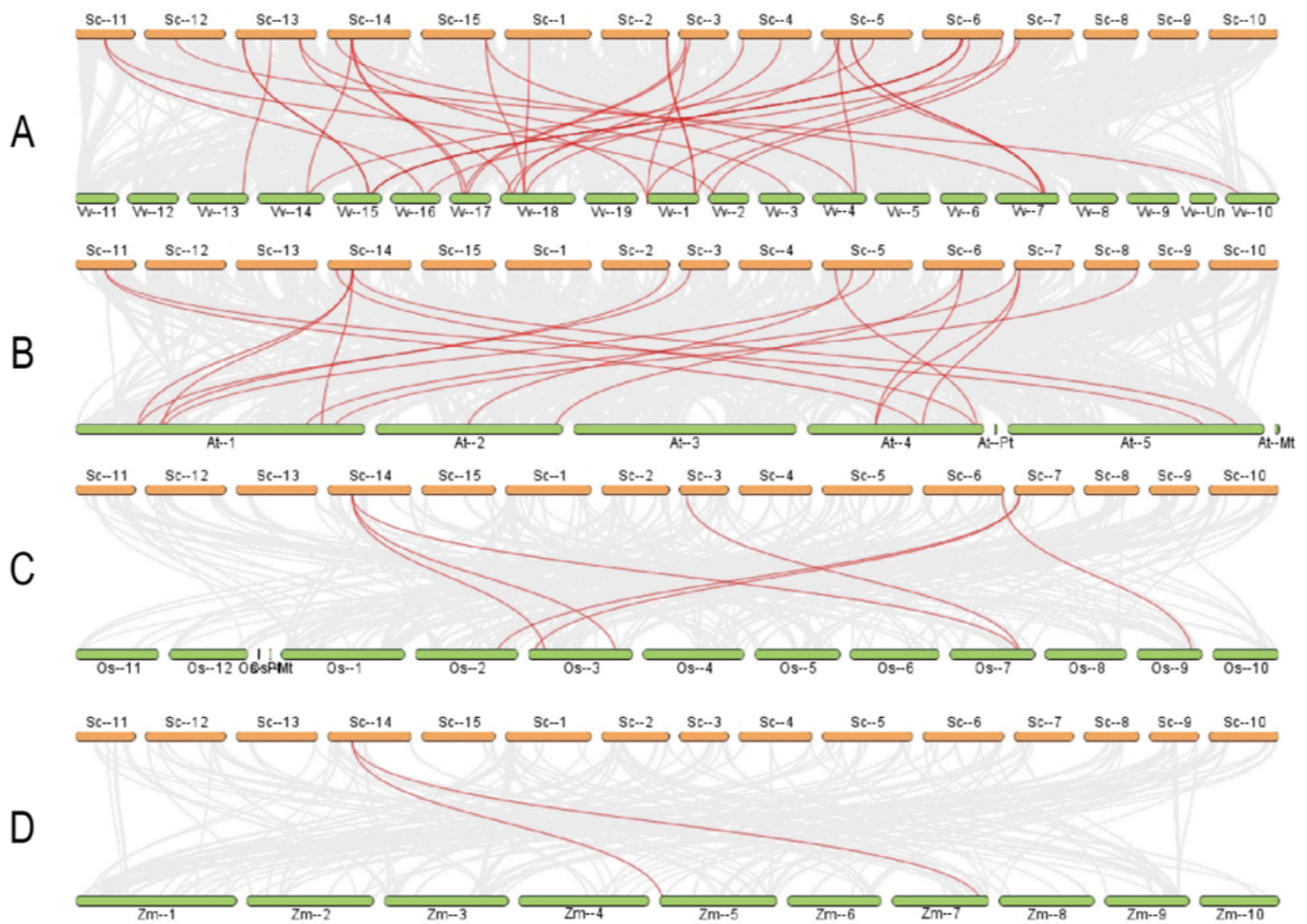


Figure 7

Analysis of covariance between *C. sinensis* var. *assamica* and other species. (A) Analysis of covariance between *C. sinensis* var. *assamica* and Grape. (B) Analysis of covariance between *C. sinensis* var. *assamica* and *Arabidopsis*. (C) Analysis of covariance between *C. sinensis* var. *assamica* and Rice. (D) Analysis of covariance between *C. sinensis* var. *assamica* and Maize

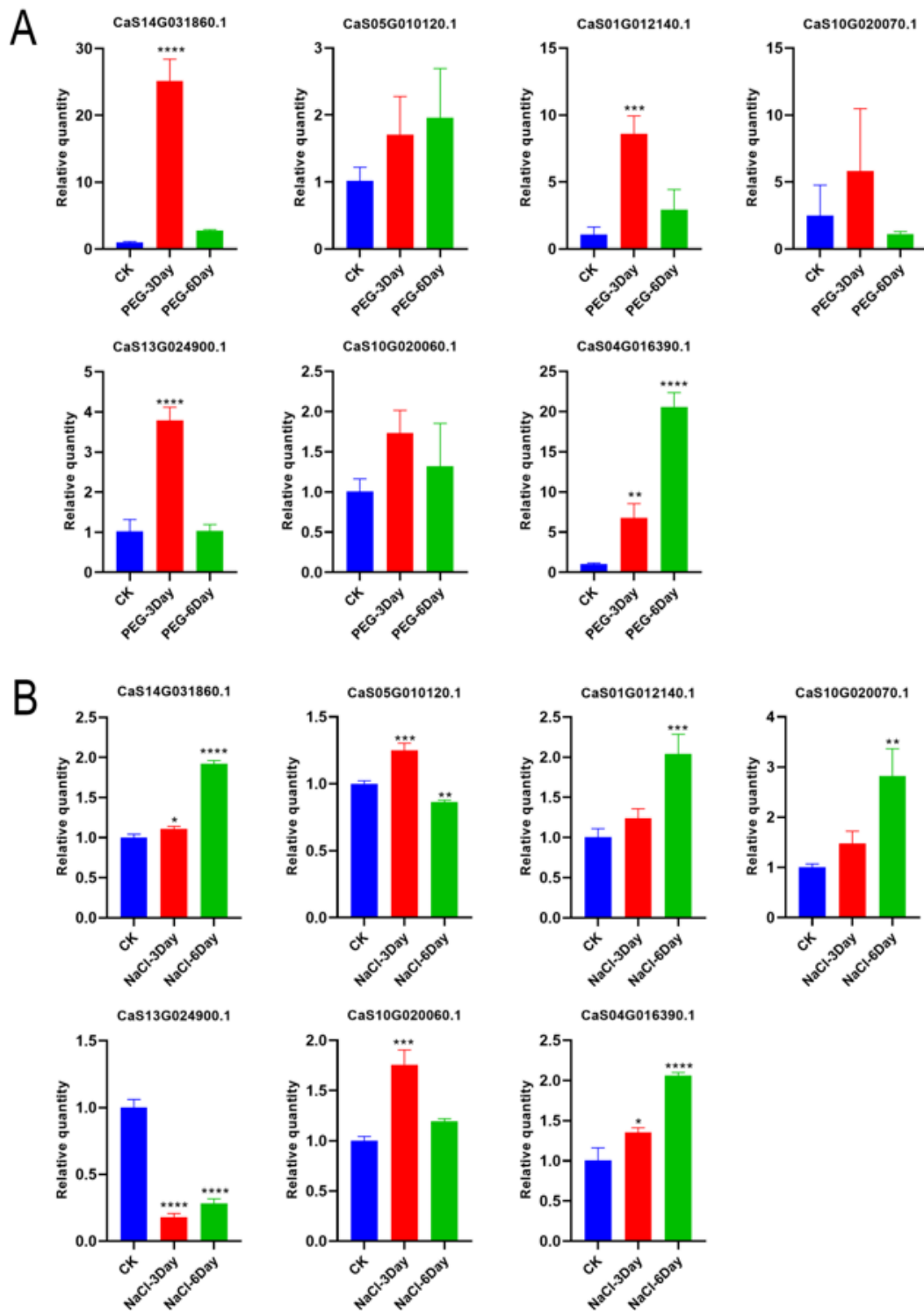


Figure 8

Expression levels of Type II genes at different time points under drought and salt stress treatments. (A) Drought stress. (B) Salt stress.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Schedule.xlsx](#)