



## OPEN ACCESS

## EDITED BY

Ke Qiang Yang,  
Shandong Agricultural University, China

## REVIEWED BY

Xin Li,  
Nanjing Agricultural University, China  
Maria Carmen Valoroso,  
University of Milan, Italy

## \*CORRESPONDENCE

Aizhong Liu  
✉ liuaizhong@mail.kib.ac.cn  
Anmin Yu  
✉ anminyu@swfu.edu.cn

RECEIVED 08 March 2025

ACCEPTED 22 April 2025

PUBLISHED 14 May 2025

## CITATION

Sun J, Zhou Z, Meng F, Wen M, Liu A and Yu A (2025) Characterization analyses of *MADS-box* genes highlighting their functions with seed development in *Ricinus communis*. *Front. Plant Sci.* 16:1589915. doi: 10.3389/fpls.2025.1589915

## COPYRIGHT

© 2025 Sun, Zhou, Meng, Wen, Liu and Yu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Characterization analyses of *MADS-box* genes highlighting their functions with seed development in *Ricinus communis*

Jing Sun<sup>1</sup>, Zekun Zhou<sup>1</sup>, Fanqing Meng<sup>1</sup>, Mengyun Wen<sup>1</sup>, Aizhong Liu<sup>1\*</sup> and Anmin Yu<sup>1,2,3\*</sup>

<sup>1</sup>Key Laboratory for Forest Resources Conservation and Utilization in the Southwest Mountains of China, Ministry of Education, Southwest Forestry University, Kunming, China, <sup>2</sup>Yunnan Provincial Key Laboratory for Conservation and Utilization of In-forest Resource, Southwest Forestry University, Kunming, China, <sup>3</sup>Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

The *MADS-box* gene family plays a pivotal role in regulating floral organ development and various aspects of plant growth. Despite its well-established importance in many species, the function and evolution of *MADS-box* genes in *Ricinus communis* (castor) remain unexplored. This study presents an extensive genome-wide analysis of the *MADS-box* gene family in castor, covering their physicochemical characteristics, phylogenetics, gene architecture, chromosomal distribution, evolutionary dynamics, expression profiles, and co-expression networks. In total, 56 *MADS-box* genes were categorized into two main phylogenetic groups: type-I and type-II, which were further subdivided into three and two subgroups, respectively. Segmental duplication was found to be the primary driver of *MADS-box* gene expansion in castor, while purifying selection was evident across the entire gene family, as indicated by the *Ka/Ks* ratio. In-depth analyses of gene expression, promoter motifs, co-expression networks, and experimental validation (Y1H assays and qRT-PCR) revealed that *RcMADS16* and *RcMADS41* are key regulators of castor seed development, with *RcMADS16* may involve in seed coat formation and *RcMADS41* in oil accumulation. This study not only provides the first detailed insight into the evolutionary and functional landscape of *MADS-box* genes in castor, but also establishes a foundation for future investigations into the role of these genes in seed and organ development, both in castor and other plant species.

## KEYWORDS

**MADS-box, co-expression, seed development, regulators, castor**

## Introduction

*Ricinus communis* L. (castor) is one of the most important non-edible oil crops worldwide, contributing significantly to biodiesel feedstock through its high hydroxy fatty acid content. Originating in Africa or India, castor bean was domesticated approximately 3,200 years ago in East Africa (Xu et al., 2021). Nowadays, castor bean is primarily cultivated in regions, such as South Asia, East Asia, and South America, due to its high seed oil content and significant economic value (Landoni et al., 2023). Castor seeds are considered representative of dicot endospermic seeds, with an embryo consisting of two thin, leaf-like cotyledons sandwiched between large endosperm tissues, all encased in a lignified and reinforced seed coat (Yu et al., 2019). Plant seed development begins with double fertilization and is influenced by communication between the embryo, endosperm, and maternal tissues (Bleckmann et al., 2014). Numerous transcription factors (TFs) from various families have been characterized as part of the complex regulatory network governing seed development, with MADS-box proteins emerging as key players.

MADS-box proteins are homeotic TFs involved in reproductive development across vascular plants, including flower development, inflorescence architecture, pollen development, and seed/fruit development (Dreni et al., 2007; Gramzow and Theissen, 2010; Smaczniak et al., 2012). These TFs are present and conserved across nearly all eukaryotic groups, from *Saccharomyces cerevisiae* to *Homo sapiens* (Shen et al., 2021). The acronym “MADS-box” is derived from the yeast MINI CHROMOSOME MAINTENANCE 1 (MCM1), the *Arabidopsis* AGAMOUS (AG), the Antirrhinum DEFICIENS (DEF), and the mammalian SERUM RESPONSE FACTOR (SRF) (Norman et al., 1988; Passmore et al., 1988; Schwarz-Sommer et al., 1990; Yanofsky et al., 1990). MADS-box genes are among the largest families of transcription factors in plants, with 107 members in *Arabidopsis*, 300 in Wheat, and 79 in peach (*Prunus persica*) (Riechmann et al., 2000; Parenicová et al., 2003; Wells et al., 2015; Schilling et al., 2020). These genes are characterized by a highly conserved MADS-box (M) domain in the N-terminal region, and could specifically bind to the CC(A/T)6GG sequence, known as the CARG-box motifs, found in the regulatory elements of their target genes (Hayes et al., 1988; Parenicová et al., 2003). Based on their C-terminal sequences, MADS-box genes are categorized into two distinct groups: type I and type II (Alvarez-Buylla et al., 2000). Type I genes are further subdivided into M $\alpha$ , M $\beta$ , and M $\gamma$  subgroups, whose function are not fully understood, although emerging studies suggest their involvement in regulating female gametophyte and endosperm development in *Arabidopsis* and grasses (Bemer et al., 2008; Colombo et al., 2008; Steffen et al., 2008; Roszak and Köhler, 2011; Shirzadi et al., 2011; Hehenberger et al., 2012; Chen et al., 2016; Batista et al., 2019; Paul et al., 2022; Zhang et al., 2023). Type II MADS-box genes, also referred to as MIKC-type genes, are distinguished by four typical protein domains: MADS (M), intervening (I), keratin-like (K), and C-terminal (C), and are further classified into MIKC\* and MIKC<sub>C</sub> subgroups (Kaufmann et al., 2005). MADS-box genes are central to the widely accepted ABCDE model of flower development, where

they control floral organ identity and growth by forming multimeric complexes (Cheng et al., 2017). In *Arabidopsis*, class A MADS-box gene *APETALA1* (*API*); class B genes *AP3* and *PISTILLATA* (*PI*); class C gene *AGAMOUS* (*AG*); class D genes *SEEDSTICK* (*STK*), *SHATTERPROOF1* (*SHP1*), and *SHP2*; and class E genes *SEPALLATA* (*SEP*) 1-4, are key regulators (Guo et al., 2015). Class A genes determine sepal identity, class A and B determine petal identity, class B and C together define stamen identity, class C alone specifies carpel identity. Class D genes are involved in ovule development, while class E genes function as “bridging molecules”, facilitating the specification of sepals, petals, stamens, and carpels by promoting protein-protein interactions (Immink et al., 2009). Several studies have shown that MADS-box genes are crucial for various physiological processes, including floral organ identity, ovules, and fruit development (Nam et al., 2003; Becker and Theissen, 2004; Ruelens et al., 2013). Recent researches have highlighted the role of MADS-box TFs in seed development and seed oil accumulation in *Arabidopsis* (Zhang et al., 2024b).

In this work, we identified 56 MADS-box genes in castor genome and performed phylogenetics, gene architecture, chromosomal distribution, evolutionary dynamics analyses to elucidate their evolutionary relationships at the genome level. Furthermore, tissue-specific expression and seed development profiles analysis revealed that *RcMADS16* and *RcMADS41* play essential roles in castor seed coat formation and endosperm development, respectively. Together, our analysis offers fresh perspectives on the possible role of these genes in castor seed development.

## Materials and methods

### Plant materials

Castor seeds from the inbred lines small-seed ZB306 and large-seed ZB107 were germinated in an incubator at 30°C for two weeks. Subsequently, the seedlings were subsequently moved to the greenhouse at Southwest Forestry University, where they were maintained under a 16-hour light/8-hour dark cycle. Seeds were collected at 5, 15, 25, 35, and 55 days after pollination (DAPs), and the seed coat were separated, frozen quickly in liquid nitrogen and kept at -80°C for storage.

### Identification and characterization of MADS-box genes in castor

Two MADS-box domains (PF00319 and PF01486) were used as queries to search the castor genome with HMMER (v3.3.2) (Wells et al., 2015). A local BLASTP search was then performed against all protein sequences of castor, using *Arabidopsis* MADS-box genes as queries with an E-value threshold of < 1e-10. The identified MADS-box genes were further analyzed using the Venny tool (<https://bioinfogp.cnb.csic.es/tools/venny/>). The ExpASY ProParam (<https://web.expasy.org/>) was employed to calculate the predicted

isoelectric point (PI) and molecular weight (MW) of the candidate RcMADS proteins (Gasteiger et al., 2003). Subcellular localization was predicted by Plant-mPLoc v2.0 (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>).

## Phylogenetic analysis of *MADS-box* genes in castor

*MADS-box* protein sequences of Arabidopsis and castor were downloaded from the TAIR (<https://www.arabidopsis.org/>) and Euphorbiaceae Database (<http://eupdb.liu-lab.com/>), respectively. All these amino acid sequences were aligned with ClustalW2 (<http://www.clustal.org/clustal2/>). Phylogenetic trees were constructed using the maximum likelihood (ML) method with 1000 bootstrap replicates, and visualized using iTOL (<https://itol.embl.de/>) (Abdullah-Zawawi et al., 2021).

## Conserved domain and gene structure analyses

Conserved domains of individual *RcMADS-box* proteins were predicted using SMART program (<http://smart.embl-heidelberg.de/>). Gene structures were visualized using the Gene Structure Display Server (version 2.0) (<http://gsds.cbi.pku.edu.cn/>). Secondary structures were predicted using the NetSurfP-3.0 (<https://dtu.biolib.com/NetSurfP-3/>) (Lin et al., 2023), and the three-dimensional structures were predicted using AlphaFold3 (<https://alphafoldserver.com/>). The PDB files of these *MADS-box* proteins were visualized using PyMOL 3.0 software (<https://pymol.org/#download>), and the three-dimensional structures alignment also used this software (Jumper et al., 2021).

## Chromosomal distribution, gene synteny, and evolutionary analyses

Chromosomal locations of *RcMADS-box* genes were obtained from the castor genome and visualized using TBtools (Chen et al., 2023). Gene duplication events were detected using MCScanX (Wang et al., 2012). Tandem and segmental duplication gene pairs were extracted from the respective duplication files of the castor genome. Gene collinearity was visualized using jcv (Wan et al., 2025). The *Ks* value of these collinearity gene pairs were calculated using the -ks, -kp programs of WGDI package (Sun et al., 2022).

## Expression profiling and gene co-expression analysis

RNA-seq raw data were generated from ten different tissues, including root, stem, leaf, pollen, inflorescence, capsule, seedling at three weeks post-planting, germinated seed, endosperm, ovule, and five stages of seed development (5, 15, 25, 35, and 55 DAPs, referred to as

S1-S5). Quality control was conducted using fastp v0.24.0 (Chen et al., 2018). Clean reads were aligned to the newly published castor genome using HISAT2, and the gene expression levels were quantified with the FeatureCounts program in the Subread package (Kim et al., 2015; Liao et al., 2019). Tissues-specific expression profiles of all *RcMADS-box* genes were analyzed using the K-means clustering method and visualized with the pheatmap package in R (Xanthopoulou et al., 2022).

To investigate the role of *RcMADS-box* genes during castor seed development, genes co-expressed across the five seed developmental stages (from 15 to 55 DAPs) were identified using ClusterGVis (<https://github.com/junjunlab/ClusterGVis>). Subsequently, co-expression networks were constructed using the GENIE3 R package, based on the FPKM values of genes specifically expressed during the early and middle-late stages of seed development (Huynh-Thu et al., 2010). Gene interactions within these networks were analyzed and visualized using Cytoscape (v3.9.1) (Shannon et al., 2003). For the prediction of *cis*-acting elements in the promoter regions of genes involved in the networks, the 2000 bp upstream sequences from the translation start codons of each gene were extracted according to their genomic positions. Potential binding sites and their associated functions were identified using the New PLACE website (<https://www.dna.affrc.go.jp/PLACE/?action=newplace>), and the results were displayed using TBtools (Chen et al., 2023).

## Yeast one-hybrid assays

Yeast One-Hybrid (Y1H) assays were employed to determine the binding of *MADS-box* transcription factors (TFs) to the promoters of potential target genes. The promoter sequences were amplified from castor genomic DNA and construct the bait vectors, the promoters fragments of target genes *RcXYL1* (from -260 to -500 bp) and *RcPRX42* (from -700 to -1101 bp) containing the CArG cis-elements were cloned into the pAbAi plasmid. The full-length CDS sequences of hub *MADS-box* TFs, synthesized from the cDNA of castor seed total RNA, were recombined into the pGADT7 vector to generate the Prey-Effector plasmids. Y1H assays were carried out using the Yeast-One-Hybrid Media kit (Coolaber, China, YH1001-10T). Bait-Reporter plasmids were transformed into Y1H Gold yeast strains and selected on minimal synthetic dropout medium lacking uracil (SD/-Ura). Yeast cells grown in SD/-Ura liquid media were diluted to 10-, 100-, and 1000-fold concentrations. The optimal concentration of Aureobasidin A (AbA) for selection was determined by incubating bait strains in SD/-Ura solid medium with 0, 100, 300, 500, or 700 ng/mL AbA. The prey vectors were then transformed into all bait strains and grown on SD/-Ura/-Leu medium. The pGADT7 empty vector was used as negative control. All primers used for Y1H assays are listed in Supplementary Table S1.

## RNA isolation and quantitative real-time PCR

Total RNA from seed coats at different developmental stages was extracted using the RNAPrep Pure Plant Plus kit (Tiangen Co. Ltd,

Beijing, China; Catalog No. DP441), designed for samples rich in polysaccharides and polyphenolics. RNA quality and quantity were determined using a NanoDrop2000 Spectrophotometer (Thermo Scientific, USA). Complementary DNA (cDNA) was synthesized using the All-in-One First-Strand cDNA Synthesis SuperMix for qPCR kit (TransGen Biotech, Beijing, China). Primers were designed using Primer3 (<https://bioinfo.ut.ee/primer3/>) with melting temperatures ( $T_m$ ) ranging between 57°C and 59°C. qRT-PCR was performed on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Berkeley, California), utilizing the PerfectStart Green qPCR SuperMix Kit (TransGen Biotech, Beijing, China). Each gene was tested in triplicate biological samples with three technical replicates per sample. Relative expression levels were calculated using the  $2^{-\Delta\Delta CT}$  method, with *RcActin* serving as the internal control. Primers sequences are provided in [Supplementary Table S2](#).

## Results

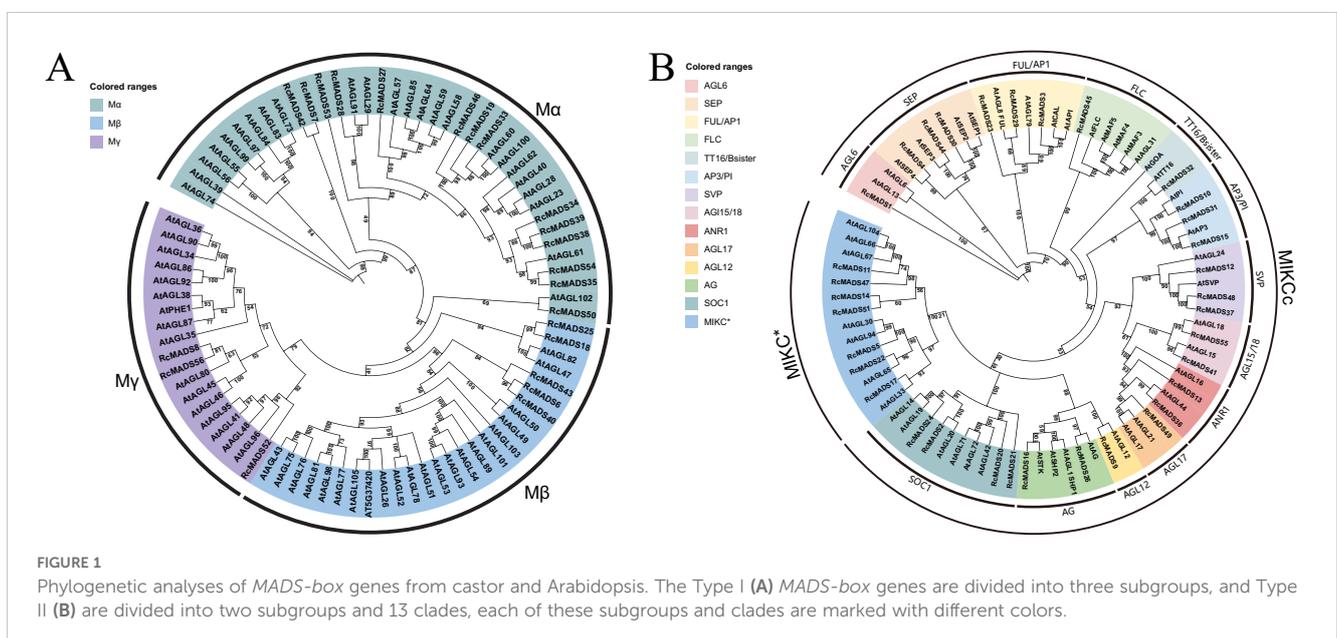
### Identification of *MADS-box* genes in castor

A total of 57 *MADS-box* putative genes were identified in the castor genome via BLAST, while a further 26 genes containing both the MADS domain (PF00319) and K-box domain (PF01486) were detected through an HMMER search. An additional 30 genes containing only the MADS domain were also identified ([Supplementary Figure S1](#)). One gene that contained only the K-box domain was excluded, resulting in 56 *MADS-box* candidates, which were designated as *RcMADS1* to *RcMADS56* based on their chromosomal positions ([Supplementary Table S3](#)). The lengths of these genes varied from 470 to 35,338 bp, with an average length of 5,543.875 bp. Thirty-one genes were longer than 3,000 bp. The amino acid lengths ranged from 119 to 697 aa, with an average length of

280.86 aa; more than 80% of the proteins were between 200 and 400 aa ([Supplementary Table S3](#)). The theoretical molecular weight (Mw) of these proteins ranged from 14.11 to 80.00 kDa, and the isoelectric point (pI) ranged from 4.96 to 10.15 ([Supplementary Table S3](#)). Subcellular localization predictions indicated that approximately 89.29% (50 genes) of these *RcMADS* genes were located in the nucleus, while 5 genes were located to the chloroplasts, and one gene (*RcMADS37*) was predicted to reside in the cytoplasm ([Supplementary Table S3](#)).

### Phylogeny and classification of *RcMADS-box* genes

The identified *RcMADS-box* genes were classified into two major groups: type-I and type-II, containing 22 and 34 genes, based on phylogenetic analysis using Arabidopsis and castor *MADS-box* genes ([Figures 1A, B](#)). Type-I genes were further subdivided into the  $M\alpha$ ,  $M\beta$ , and  $M\gamma$  subgroups, containing 14, 5, and 3 genes, respectively. Phylogenetic analysis suggested that the divergence of type-I *MADS-box* genes between Arabidopsis and castor occurred prior to the separation of these two species ([Figure 1A](#)). Type-II genes were subdivided into the MIKCC and MIKCC\* subgroups, comprising 27 and 7 genes in castor, respectively ([Figure 1B](#)). The MIKCC subgroup could be further divided into 13 evolutionary clades based on the known Arabidopsis *MADS-box* genes, such as AGL6 (*RcMADS1*), SEP (*RcMADS4/30/44*), FUL/AP1 (*RcMADS3/23/29*), FLC (*RcMADS45*), TT16/Bsister (*RcMADS32*), AP3/PI (*RcMADS10/15/31*), SVP (*RcMADS12/37/48*), AGL15/18 (*RcMADS49/55*), ANR1 (*RcMADS13/36*), AGL17 (*RcMADS49*), AGL12 (*RcMADS9*), SOC1 (*RcMADS2/20/21/24*), and AG (*RcMADS16/26*). Genes from castor and Arabidopsis that clustered in the same clade likely share similar functions, providing important insights for the selection of candidate *RcMADS-box* genes in further studies on castor seed development.

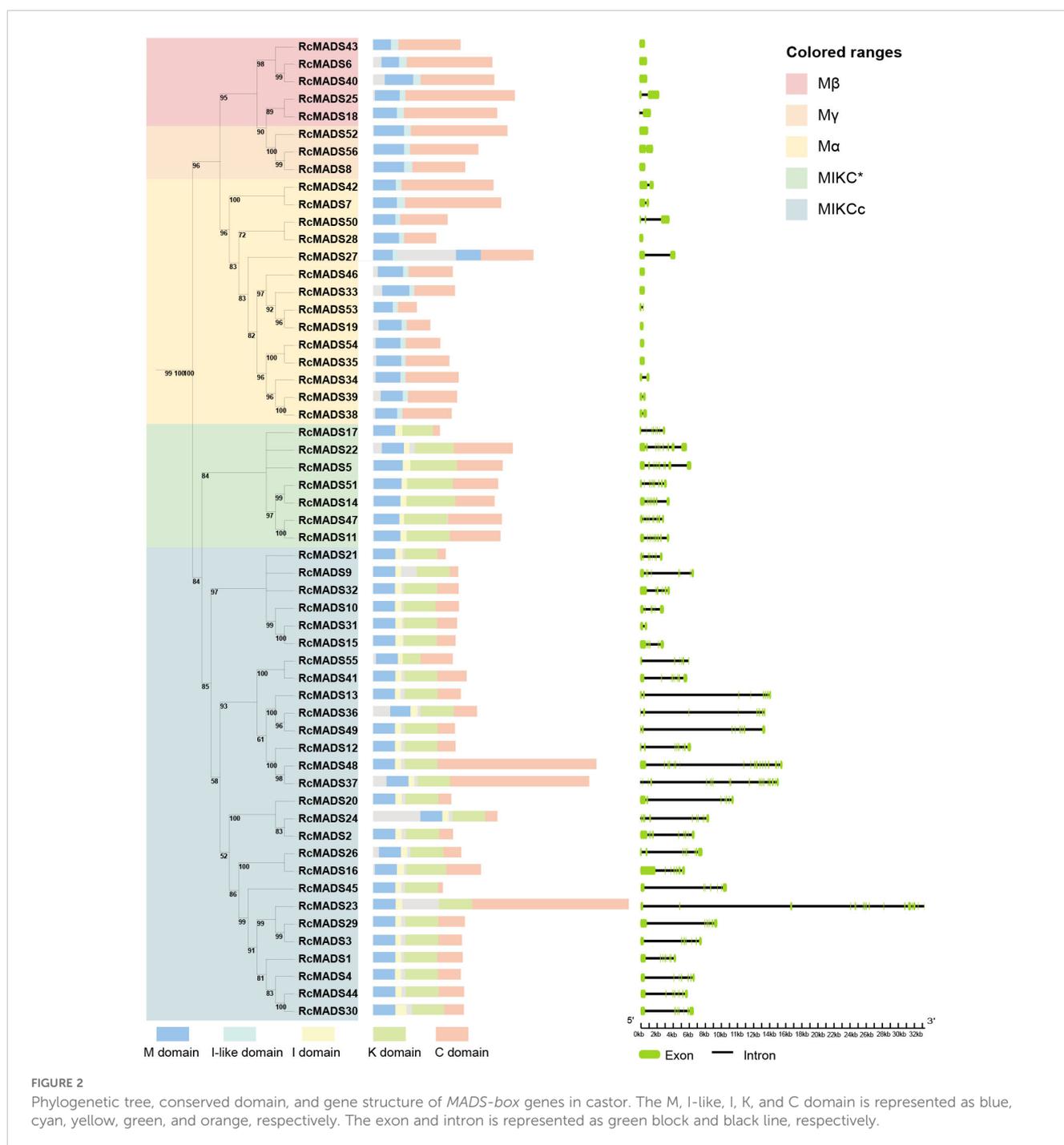


## Conserved domains and structures of *RcMADS-box* genes

Conserved domain analysis revealed differences in the MADS (M) domain sequences between type-I and type-II *RcMADS* proteins in castor (Figure 2). Consistent with the above HMMER search results, the 22 type-I proteins contained only the M, Intervening-like (I-like), and C domains, where the 34 type-II proteins included the M, I, Keratin-like (K), and C domains (Figure 2). Additionally, the intron-exon patterns of type-I genes were simpler than type-II genes. Notably, 12 out of the 22 type-I

*RcMADS-box* genes were intronless, while none of the type-II lacked introns. In contrast, 26 of the 34 type-II *RcMADS-box* genes contained five to ten introns.

The secondary and tertiary structures of *RcMADS-box* proteins were predicted using NetSurfP-3.0 tool and AlphaFold3, respectively (Figure 3). Each *RcMADS-box* gene exhibited similar secondary structures (Supplementary Figures S2–S4). The M domain, located at the N-terminal, consisted of an alpha helix ( $\alpha$ A) and two antiparallel beta strands ( $\beta$ A and  $\beta$ B) (Figure 3). This region is essential for DNA binding, dimer and/or tetramer formation via the canonical CArG-box found in the upstream of

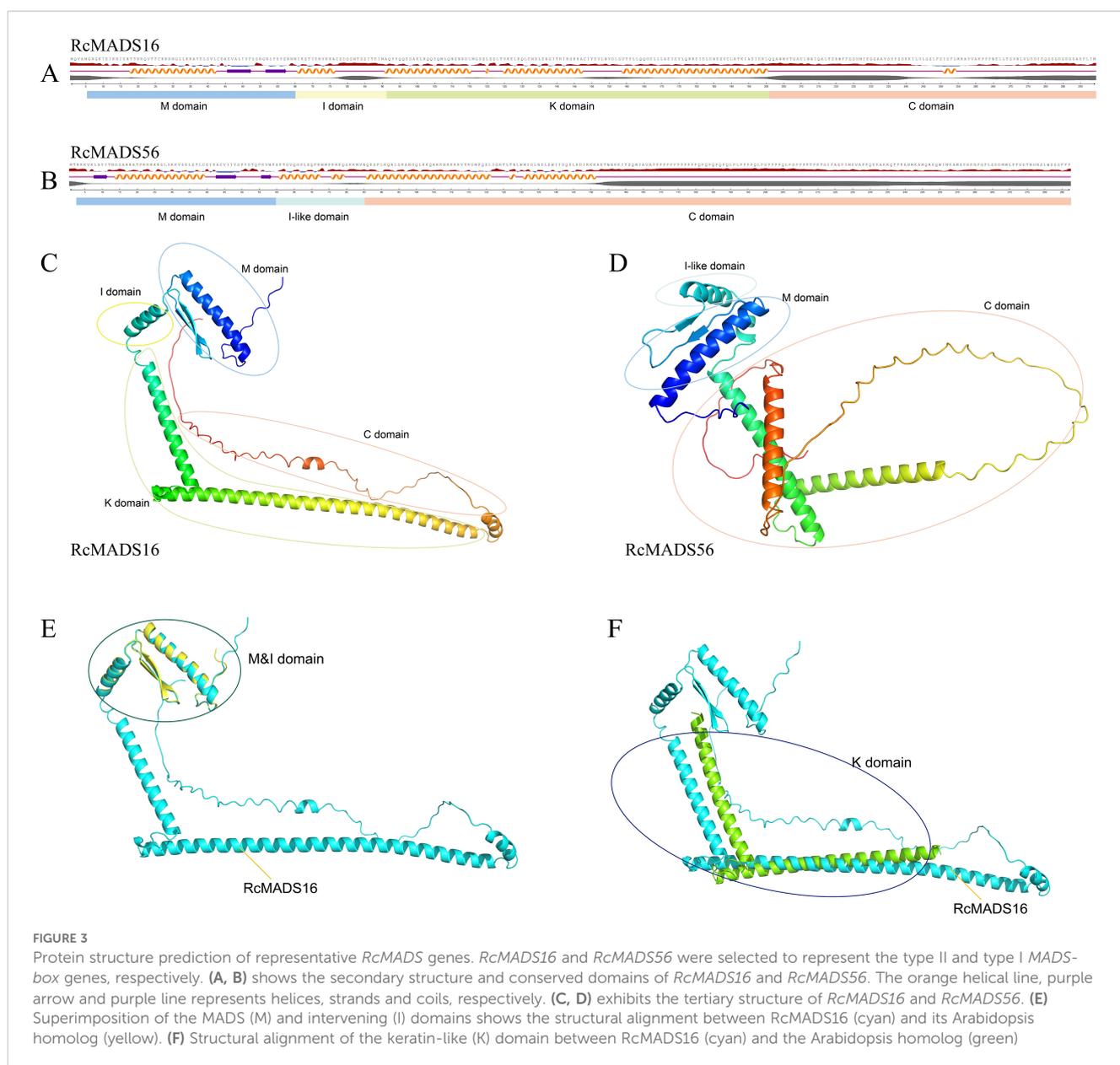


all regulated genes. Several amino acid residues were highly conserved, including Arg (R), Lys (K), Glu (E), and Leu (L) at positions 24, 31, 34, and 38, respectively, in both type I and type II *RcMADS-box* proteins (Supplementary Figure S5). More conserved amino acid residues were detected in type II proteins (Supplementary Figure S6). The I/I-like domain, which contains an  $\alpha$ -helix structure, determines DNA-binding specificity. This domain is weakly conserved, with characteristic residues such as Val (V) 64, Val (V) 67, Leu (L) 68, and Phe (F) 71 in the I domain of type I proteins, and Met (M) 63, Val (V) 66, Ile/Leu (I/L) 67, Arg (R) 69, Tyr (Y) 70 in the I-like domain of type II proteins (Supplementary Figure S7). The K domain, which mediates protein-protein interactions, consists of a series of helices. The C domain, composed of random coils, exhibits sequence variability (Figure 3; Supplementary Figures S2–S4). Homology modeling of the tertiary structures of *MADS-box* proteins from castor and Arabidopsis demonstrated high structural conservation within the

same subfamily but considerable diversity between different subfamilies across different species (Figure 3).

## Chromosomal localization and gene duplications of *RcMADS-box* genes

The 56 *RcMADS-box* genes were unevenly distributed across nine of the ten chromosomes. Notably, there are nine *RcMADS-box* genes on Chr2, Chr8, and Chr9, while only two genes (*RcMADS32/33*) were located on Chr7. Gene duplication analysis using MCScanX revealed that five pairs of *RcMADS-box* genes arose through tandem repeats, including *RcMADS3/RcMADS4*, *RcMADS29/RcMADS30*, *RcMADS34/RcMADS35*, *RcMADS38/RcMADS39*, *RcMADS44/RcMADS45*, which were located on Chr1, Chr5, Chr8, and Chr9, respectively (Supplementary Table S4). Furthermore, a total of 14 gene pairs, including 23 *RcMADS-box*



genes (including 2 type-I and 21 type-II *RcMADS-box* gene), representing 41.07% of the total *RcMADS-box* genes, were located in segmental duplicated regions (Figure 4; Supplementary Table S4). Notably, 10 out of 18 gene pairs belonged to the same subgroups or clades, suggesting that *RcMADS* genes were conserved during gene duplication. Evolutionary analysis revealed that the  $Ka/Ks$  ratios for all *RcMADS-box* gene pairs ranged from 0.09 to 0.32, all below 1, indicating strong purifying selection acting on these genes (Supplementary Table S5). Additionally, 9 out of 18 duplicated gene pairs had higher  $Ks$  values than the genome-wide average, suggesting duplication may accelerate gene evolution.

## Expression profiles of *RcMADS-box* genes in different tissues and stages of castor seed development

To explore the expression profiles of *RcMADS-box* genes across various tissues, RNA-seq data were retrieved from the NCBI under the BioProject accessions PRJNA787114 and Euphorbiaceae database ([http://eupdb.liu-lab.com/R\\_communis\\_browse/](http://eupdb.liu-lab.com/R_communis_browse/)) (Liu et al., 2024), including tissues such as endosperm, ovule, root, pollen, leaf, inflorescence, germinated seed, stem, seedling, and capsule. Based on these expression profiles, a total of 14 *RcMADS* genes with FPKM values below 1 were excluded from the analysis, all of which were type-I, except for *RcMADS55* (Supplementary Table S6). The remaining 42 *RcMADS* genes were selected for k-

means clustering analysis, and grouped into eight clusters (C1-C8) (Figure 5A). Two genes (*RcMADS22/38*) in C1 and four genes (*RcMADS26/30/44/56*) in C2 were highly expressed in both pollen and ovule tissues, while nine genes in C3 were predominantly expressed in pollen (Figure 5A; Supplementary Table S6). Seven genes in C5 were specially expressed in the root, and seven genes in C6 were highly expressed in inflorescence, while only three genes were specially expressed in the leaf. Notably, four genes (*RcMADS14/41/43/52*) in C4 and six genes (*RcMADS1/16/17/28/32/50*) in C7 were highly expressed in the endosperm and ovule, respectively.

We further analyzed gene expression across five stages of seed development (S1 to S5). Genes in C4 and C7 showed higher expression levels during the middle-late (S3-S5) and early-middle (S1-S3) stages during seed development, respectively (Figure 5B; Supplementary Table S7). Additionally, hierarchical clustering for seed-specific genes identified in our previous study (Yu et al., 2023) revealed that 11,212 genes were co-expressed with *RcMADS1/16/17/28/32/50*, which were clustered in C7, and highly expressed in the early stages of castor seed development (Figure 5C; Supplementary Table S8). A total of 1,982 genes co-expressed with *RcMADS14/41/43/52*, clustered in C4, were highly expressed in the middle-late stages of seed development (Figure 5C; Supplementary Table S8). Notably, four of the ten seed-specific genes belonged to type-I, while the remaining genes were type-II, suggesting that both types of *RcMADS* genes play essential roles during castor seed development.

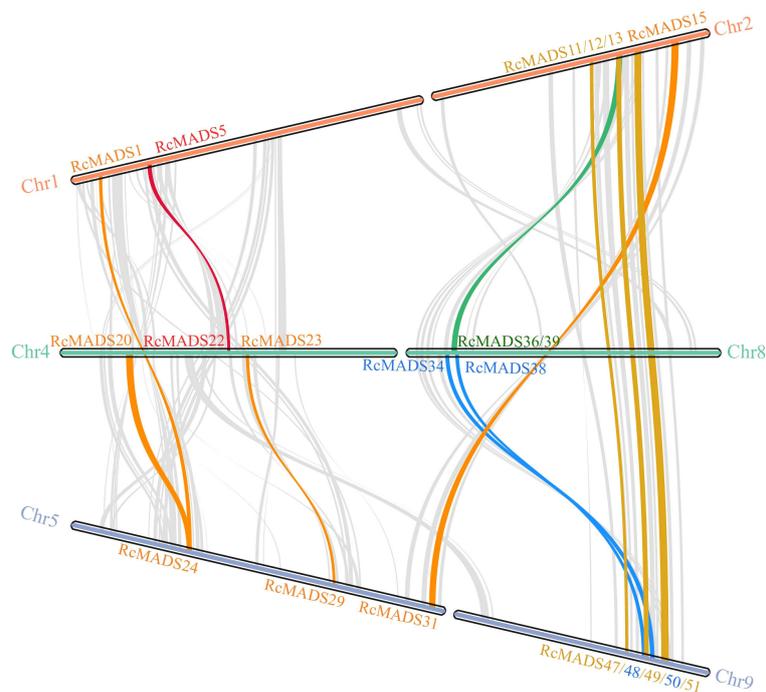


FIGURE 4

Gene duplication relationship of *MADS-box* genes in castor. Chromosomes are represented as horizontal bars, with gene names labeled at their respective positions. Gray curves in the background indicate syntenic relationships across the genome, while colored lines highlight different types of gene duplication events. These duplication patterns provide insights into the evolutionary expansion of the *MADS-box* gene family in castor.

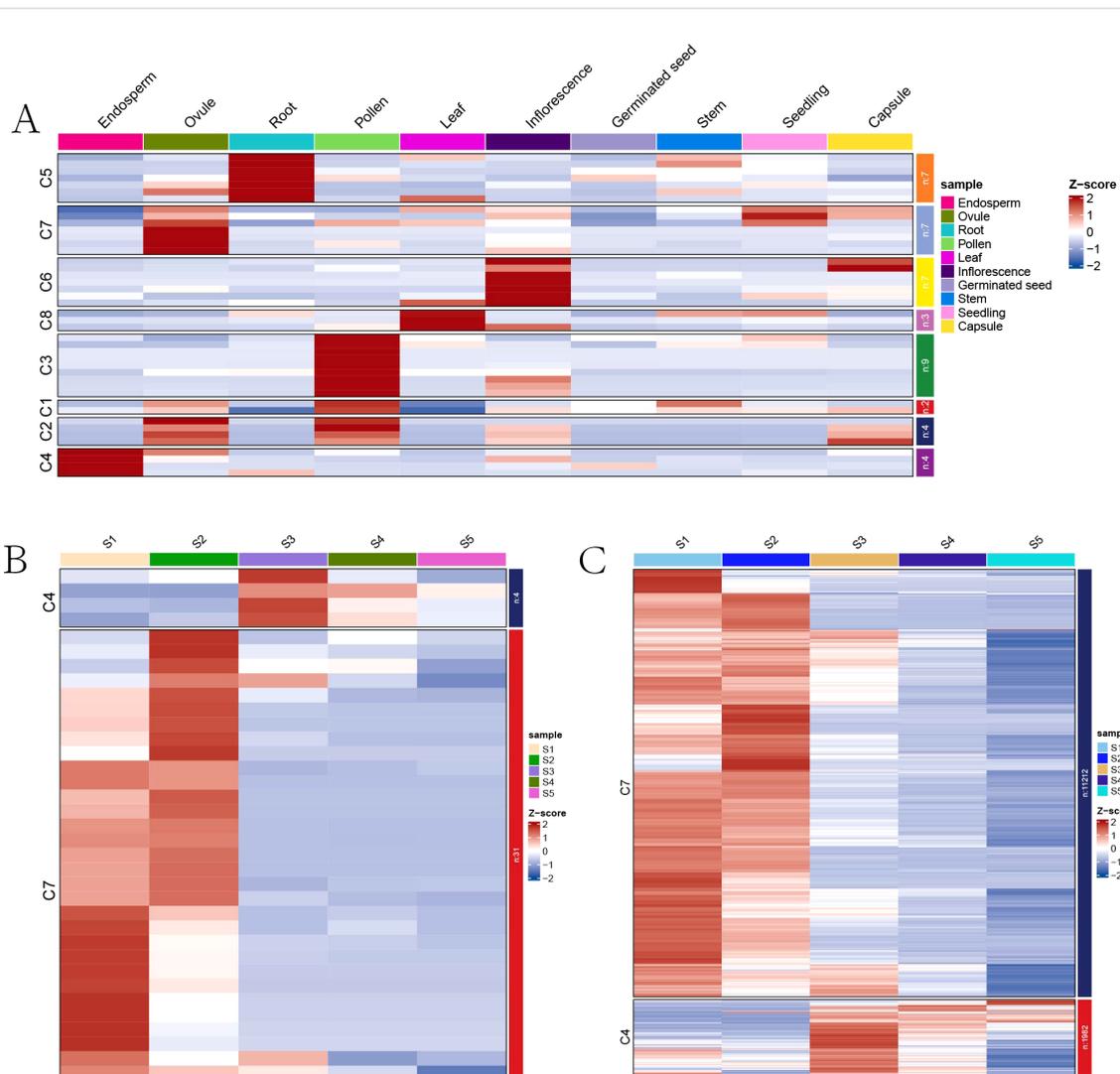


FIGURE 5

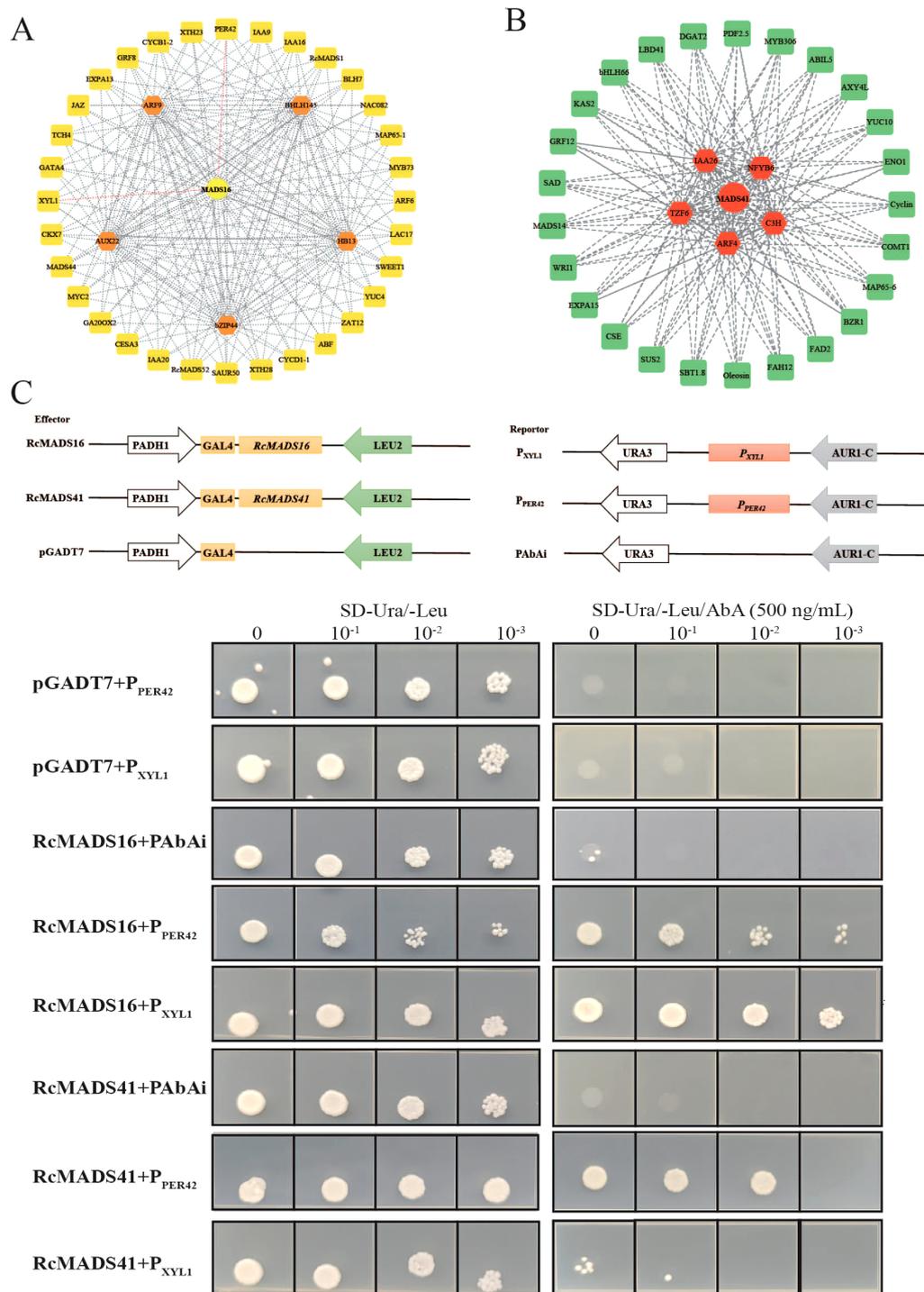
Expression profiles of *RcMADS*-box genes. (A) Expression patterns of *RcMADS*-box genes in different tissues of castor; (B) Expression patterns of *RcMADS*-box genes during five different developmental stages of castor seed; (C) Expression patterns of all seed-specific genes, including the *RcMADS*-box genes during the five seed developmental processes of castor.

## Prediction of gene regulatory networks during castor seed development

We utilized GENIE3 to predict key regulatory factors and their downstream targets, and constructing two gene regulatory networks for early-middle and middle-late stages of castor seed development (Figure 6A). In the early stage, *RcMADS16* (the homolog of *AtSTK*) shared the maximum number of nodes with key TFs in the first layer, including *RcARF9* (*Rc05G010057*), *RcAUX22* (*Rc07G016821*), *RcHB13* (*Rc01G000650*), *RcbZIP44* (*Rc02G004777*), and *RcbHLH145* (*Rc02G004326*) (Figure 6A). GO enrichment analysis of *RcMADS16* target genes indicated that they were primarily involved in plant-type cell wall (GO: 0009505), xyloglucan 1,6-alpha-xylosidase activity (GO: 0080176), and regulation of RNA biosynthetic process (GO: 2001141) (Supplementary Figure S8). The promoter regions of these target genes contained numerous hormone response elements, including Cytokinin, GARE, SAUR,

ERF3, ABA, Auxin/SA, as well as ARF elements (Supplementary Figure S9; Supplementary Table S9). Numerous CARG motifs, and TF binding elements (bZIP, MYB, HD, MYC, E2F, and SOC1) were also identified in their promoters (Supplementary Figure S9). These analyses suggest that *RcMADS16* regulates cell number and growth by modulating the expression of *RcCYCB1-2* (*Rc05G011661*), *RcCYCD1-1* (*Rc08G017624*), and *RcHB13* (*Rc01G000650*), and also influences secondary cell wall (SCW) biosynthesis through these lignin (*RcPER42*, *Rc05G009355*), cellulose (*RcCESA3*, *Rc08G018241*), and hemicellulose (*RcXYL1*, *Rc04G008432*) synthesis-related genes (Figure 6A). This finding suggests that *RcMADS16* and plant hormones, especially auxin, play a central role in regulating seed coat formation in castor.

During the middle-late stage, the *RcMADS41*-mediated regulatory network was explored. *RcIAA26* (*Rc07G016672*), *RcARF4* (*Rc01G001034*), *RcNFYB6* (*Rc10G023274*), *RcTZF6* (*Rc04G007632*), and *RcC3H* (*Rc03G005462*) were identified as the



**FIGURE 6**  
 Prediction and verification of the *RcMADS16* and *RcMADS41* targeted genes during castor seed development. **(A)** Prediction of the *RcMADS16* regulatory network in the early seed development stage. **(B)** Prediction of the *RcMADS41* regulatory network in the middle seed development stage. **(C)** Y1H assays between *RcMADS16* and *RcMADS41* and the promoter fragments of the targeted genes.

first-layer regulators (Figure 6B). KEGG enrichment of their target genes in the second layer revealed that they were predominantly involved in fatty acid biosynthesis (ko00061), plant hormone signal transduction (ko04075), and glycerolipid metabolism (ko00561) (Supplementary Figure S10). The *cis*-elements of all genes in this

network were similar to those described above, except for the 2S (short for 2SSEEDPROTBANAPA) *cis*-element (Supplementary Figure S11, Supplementary Table S10), which is conserved in many storage protein gene promoters. Notably, the CarG motif was detected in nearly all target genes (28 out of 31), including

*RcWRI1* (Rc08G018567), *RcDGAT2* (Rc05G012439), and *RcFAD2* (Rc06G014656), which are key TFs/genes involved in fatty acid accumulation.

To verify the potential interactions between these hub TFs and their target genes, the yeast one hybridization (Y1H) system was employed to test whether *RcMADS16* and *RcMADS41* could directly bind to the promoter of *RcPER42* and *RcXYL1*, two genes related to cell wall formation (Figure 6C). Compared with the negative control, yeast strains transformed with *RcMADS16* in pGADT7 and *P<sub>XYL1</sub>* in pAbAi vectors grew normally on the screening medium. However, when the prey was *P<sub>PER42</sub>*, yeast growth was inhibited, suggesting that *RcMADS16* directly binds to the *RcXYL1* promoter, but with weaker binding strength to the *RcPER42* promoter (Figure 6C). Conversely, when *RcMADS41*-pGADT7 and *P<sub>XYL1</sub>*-pAbAi or *P<sub>PER42</sub>*-pAbAi were transformed into yeast, the strains showed poor growth on medium supplemented with 500 ng/mL AbA, indicating that *RcMADS41* did not activate these promoters. These results suggest that *RcMADS16* is a key TF governing seed development by regulating the lignification of the seed coat, while *RcMADS41* likely plays an essential role in oil accumulation during castor seed development.

## qRT-PCR validation

To verify the reliability of RNA-Seq data and examine the differential expression levels between large-seed ZB107 and small-seed ZB306 during seed coat formation (Figure 7), 16 candidate genes were selected for qRT-PCR analysis. Compared to large-seed ZB107, ten of these genes exhibited higher expression levels in the seed coat of small-seed ZB306 during seed development, excluding *RcMADS16/44*, *RcbZIP44*, and *RcIAA9* (Figure 7). Notably, *RcXYL1* and *RcPER42* showed higher expression levels at the middle stage (25 DAP) of seed coat development, with their relative expression levels being higher in ZB306 than in ZB107. This supports previous findings that the lignin content in the seed coat ZB306 is higher than in ZB107 (Yu et al., 2019). Additionally, we confirmed the co-expression patterns between *RcMADS16* and *RcXYL1*, as well as *RcPER42* (Figure 7). Furthermore, the relative expression levels of selected genes verified by qRT-PCR were consistent with the mRNA abundance profiles from the RNA-Seq analysis.

## Discussion

The model dicotyledonous oil crop, castor, has evolved a unique seed morphology, characterized by a hard seed coat and persistent endosperm (Anjani, 2012). Although secondary cell wall components and important phytochemicals such as flavonoids and lignin have been identified in castor seed coats, little is known about the molecular mechanisms underlying seed coat identity (Tobimatsu et al., 2013; Wang et al., 2020a).

The *MADS-box* gene family is one of the most ancient and extensively studied transcription factor families across various species. Their functions were diverse, including roles in flowers

formation, flowering time control, organ identity, and plant development (Theißen et al., 2016). However, the roles of *MADS-box* proteins in seed development have mainly been explored in Arabidopsis and rice. It's well established that type-I *MADS-box* genes play essential roles in endosperm formation and development in Arabidopsis through imprinted or non-imprinted patterns (Zhang et al., 2018; Wang et al., 2020b; Dziasek et al., 2024). Several endosperm-related type-I *MADS-box* genes have been identified, i.e., *AGL9*, *AGL61*, *AGL62*, *AGL80*, and *PHE1* (Kang et al., 2008; Steffen et al., 2008; Zhang et al., 2024a). In our analysis, type-I genes *RcMADS28/43/50/52* were also highly expressed during the early stage of seed development when endosperm proliferation and cellularization predominantly occur (Figure 5), suggesting their potential functions in castor seed development.

In this study, we conducted a systematic analysis of the *MADS-box* family genes in the castor genome. A total of 56 *MADS-box* genes were identified, a lower number than the 107 and 75 genes found in Arabidopsis and rice, respectively (Parenicová et al., 2003; Arora et al., 2007). Notably, 33 out of the 56 *RcMADS-box* genes expanded through tandem and segmental duplication (Figure 4). The relatively low number of genes in castor can be attributed to the absence of a species-specific whole-genome duplication event in castor, compared to Arabidopsis and rice (Xu et al., 2021). The structures of *RcMADS-box* proteins were characterized by the M, I, K, and C domains. The M domain, in particular, was highly conserved among all *MADS-box* genes in castor and is critical for the proper functioning of *MADS-box* TFs by enabling them to bind specific genes or gene regions (Figure 2). Consistent with previous studies in other species, these *MADS-box* genes in castor could be categorized into two main types: 22 type I (M $\alpha$ , M $\beta$ , M $\gamma$ ) and 34 type II (MIKC<sub>C</sub>, MIKC\*) genes. Among these, the 27 MIKC<sub>C</sub> proteins were further divided into 13 clades based on their phylogenetic relationships (Figure 1A). Genes in the AG and SEP clades are the most extensively studied, corresponding to the D and E function genes in the well known ABC(DE) models of flower development (Dreni and Ferrándiz, 2022). For example, genes in the D class, including *STK*, *SHP1*, and *SHP2*, have been implicated in fruit and seed development (Dreni and Kater, 2014; Fei et al., 2021; Paolo et al., 2021). In Arabidopsis, *STK* has been identified as a master regulator of ovule identity, and *AGL11* in grape was essential for seedless fruit morphogenesis (Ezquer et al., 2016). The variation of the *STK* homolog in the oil palm, *SHELL*, is associated with the presence and thickness of the lignified shell (Malabarba et al., 2017). In Arabidopsis, *STK* regulates seed size by modifying the expression of the cell wall degradation gene  $\alpha$ -XYLOSIDASE1 (*XYL1*) (Di Marzo et al., 2022). It also controls fruit size by regulating cytokinin levels and the expression of *FRUITFULL* (*FUL*) in Arabidopsis (Marzo et al., 2020). In our study, *RcMADS16*, a homolog of *STK*, was highly expressed in the early stage of castor seed development, particularly in the seed coat, suggesting that *RcMADS16* is a master regulator of seed coat formation in castor, controlling processes such as cell cycle progression, plant hormone signal transduction, and secondary cell biosynthesis (Figure 6). Furthermore, *RcMADS16* exhibited significantly higher expression levels in the large-seed genotype

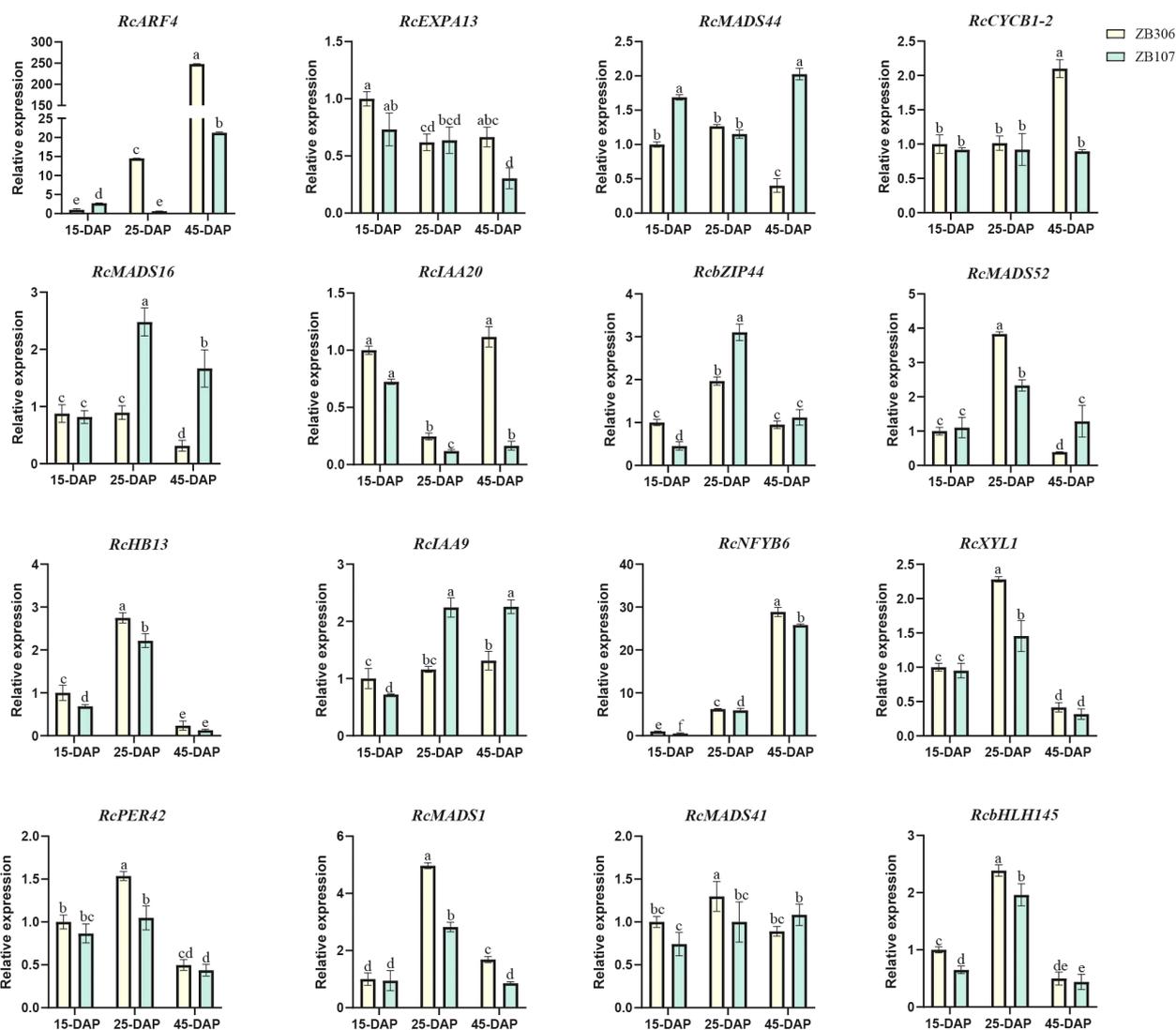


FIGURE 7

Relative expression levels of *RcMADS-box* and their targeted genes during castor seed coat development. The expression of each gene in ZB306 at 15 DAP was set as 1.0. The different letters represent statistical difference based on LSD (least significant difference) analysis with  $p < 0.05$ . The relative expression levels are means  $\pm$  SD ( $n=3$ ).

(ZB107) compared to the small-seed genotype (ZB306) (Figure 7). This expression pattern suggests that *RcMADS16* plays a pivotal role in determining seed size, likely by regulating seed coat development. Given that our Y1H assays demonstrated direct binding of *RcMADS16* to the promoters of *RcXYL1* and *RcPER42*—two genes potentially involved in cell wall remodeling and lignification—*RcMADS16* may influence seed coat thickness and composition (Figure 6). However, the precise molecular mechanism remains to be elucidated. Future studies could incorporate functional validation approaches, such as overexpression or knockout of *RcMADS16* in castor or heterologous systems, to directly assess its impact on seed coat formation, lignin deposition, and possibly oil accumulation. Additionally, further investigation into potential allelic variation in *RcMADS16* between the two genotypes could provide deeper insights into its role in seed size determination at the genetic level.

## Conclusions

In this study, we identified a total of 56 *MADS-box* genes in the castor genome, which can be classified into two main groups, each featuring conserved domains. Gene duplication analysis revealed that segmental duplication was the primary driving force behind the expansion of *RcMADS-box* genes, and the *Ka/Ks* ratio suggested that these genes have undergone purifying selection. Expression profiles showed that *RcMADS-box* genes were widely expressed in different tissues of castor, with 10 genes specifically expressed in castor seeds. Based on the results from co-expression networks, promoter analysis, Y1H, and qRT-PCR, *RcMADS16* and *RcMADS41* were identified as key regulators of castor seed development, with *RcMADS16* associated with seed coat formation and *RcMADS41* linked to oil accumulation. In summary, this study contributes to the understanding of the roles of *MADS-box* family genes in castor

and provides a solid foundation for further investigations into their functions in seed or other organs development.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#). Further inquiries can be directed to the corresponding authors.

## Author contributions

JS: Writing – original draft, Data curation, Software, Visualization. ZZ: Data curation, Writing – original draft, Formal Analysis, Validation. FM: Data curation, Validation, Writing – original draft. MW: Data curation, Writing – original draft, Visualization. AL: Funding acquisition, Writing – original draft, Writing – review & editing. AY: Funding acquisition, Writing – original draft, Writing – review & editing, Supervision.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China (NSFC, 32360475, 32372135, 32261143461), Foundation of Yunnan Agricultural Basic Research (202101BD070001-126), Forestry Innovation Programs of Southwest Forestry University (LXXK-2023Z02), Yunnan Fundamental Research Projects (202201AU070205, 202401AT070269),

## References

- Abdullah-Zawawi, M.-R., Ahmad-Nizamuddin, N.-F., Govender, N., Harun, S., Mohd-Assaad, N., and Mohamed-Hussein, Z.-A. (2021). Comparative genome-wide analysis of WRKY, MADS-box and MYB transcription factor families in Arabidopsis and rice. *Sci. Rep.* 11, 19678. doi: 10.1038/s41598-021-99206-y
- Alvarez-Buylla, E. R., Liljegren, S. J., Pelaz, S., Gold, S. E., Burgeff, C., Ditta, G. S., et al. (2000). MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J.* 24, 457–466. doi: 10.1111/j.1365-313X.2000.00891.x
- Anjani, K. (2012). Castor genetic resources: A primary gene pool for exploitation. *Ind. Crops Prod.* 35, 1–14. doi: 10.1016/j.indcrop.2011.06.011
- Arora, R., Agarwal, P., Ray, S., Singh, A. K., Singh, V. P., Tyagi, A. K., et al. (2007). MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* 8, 242. doi: 10.1186/1471-2164-8-242
- Batista, R. A., Moreno-Romero, J., Qiu, Y., van Boven, J., Santos-González, J., Figueiredo, D. D., et al. (2019). The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *eLife* 8, e50541. doi: 10.7554/eLife.50541
- Becker, A., and Theissen, G. (2004). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0
- Berner, M., Wolters-Arts, M., Grossniklaus, U., and Angenent, G. C. (2008). The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. *Plant Cell* 20, 2088–2101. doi: 10.1105/tpc.108.058958
- Bleckmann, A., Alter, S., and Dresselhaus, T. (2014). The beginning of a seed: regulatory mechanisms of double fertilization. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00452
- Chen, C., Begcy, K., Liu, K., Folsom, J. J., Wang, Z., Zhang, C., et al. (2016). Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. *Plant Physiol.* 171, 606–622. doi: 10.1104/pp.15.01992
- Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Cheng, Z., Ge, W., Li, L., Hou, D., Ma, Y., Liu, J., et al. (2017). Analysis of MADS-box gene family reveals conservation in floral organ ABCDE model of moso bamboo (*Phyllostachys edulis*). *Front. Plant Sci.* 8, 656. doi: 10.3389/fpls.2017.00656
- Colombo, L., Battaglia, R., and Kater, M. M. (2008). Arabidopsis ovule development and its evolutionary conservation. *Trends Plant Sci.* 13, 444–450. doi: 10.1016/j.tplants.2008.04.011
- Di Marzo, M., Viana, V. E., Banfi, C., Cassina, V., Corti, R., Herrera-Ubaldo, H., et al. (2022). Cell wall modifications by  $\alpha$ -XYLOSIDASE1 are required for control of seed and fruit size in Arabidopsis. *J. Exp. Bot.* 73, 1499–1515. doi: 10.1093/jxb/erab514
- Dreni, L., and Ferrándiz, C. (2022). Tracing the evolution of the SEPALLATA subfamily across angiosperms associated with neo- and sub-functionalization for reproductive and agronomically relevant traits. *Plants* 11, 1312–1326. doi: 10.3390/plants11212934

Fund of Yunnan Key Laboratory of Crop Wild Relatives Omics (CWR-2024-05). We thank all the individuals who have helped us in this study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1589915/full#supplementary-material>

- Dreni, L., Jaccia, S., Fornara, F., Fornari, M., Ouwerkerk, P. B. F., An, G., et al. (2007). The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. *Plant J.* 52, 690–699. doi: 10.1111/j.1365-3113.2007.03272.x
- Dreni, L., and Kater, M. M. (2014). MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol.* 201, 717–732. doi: 10.1111/nph.12555
- Dziasek, K., Santos-González, J., Wang, K., Qiu, Y., Zhu, J., Rigola, D., et al. (2024). Dosage-sensitive maternal siRNAs determine hybridization success in *Capsella*. *Nat. Plants* 10, 1969–1983. doi: 10.1038/s41477-024-01844-3
- Ezquer, I., Mizzotti, C., Nguema-Ona, E., Gotté, M., Beauzamy, L., Viana, V. E., et al. (2016). The developmental regulator SEEDSTICK controls structural and mechanical properties of the arabidopsis seed coat. *Plant Cell* 28, 2478–2492. doi: 10.1105/tpc.16.00454
- Fei, X., Shi, Q., Qi, Y., Wang, S., Lei, Y., Hu, H., et al. (2021). ZbAGL11, a class D MADS-box transcription factor of *Zanthoxylum bungeanum*, is involved in sporophytic apomixis. *Hortic. Res.* 8, 23. doi: 10.1038/s41438-020-00459-x
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi: 10.1093/nar/gkg563
- Gramzow, L., and Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biol.* 11, 214. doi: 10.1186/gb-2010-11-6-214
- Guo, S., Sun, B., Looi, L.-S., Xu, Y., Gan, E.-S., Huang, J., et al. (2015). Co-ordination of flower development through epigenetic regulation in two model species: rice and arabidopsis. *Plant Cell Physiol.* 56, 830–842. doi: 10.1093/pcp/pcv037
- Hayes, T. E., Sengupta, P., and Cochran, B. H. (1988). The human c-fos serum response factor and the yeast factors GRM/PRTF have related DNA-binding specificities. *Genes Dev.* 2, 1713–1722. doi: 10.1101/gad.2.12b.1713
- Hehenberger, E., Kradolfer, D., and Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. *Development* 139, 2031–2039. doi: 10.1242/dev.077057
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5, e12776. doi: 10.1371/journal.pone.0012776
- Immink, R. G., Tonaco, I. A., de Folter, S., Shchennikova, A., van Dijk, A. D., Busscher-Lange, J., et al. (2009). SEPALLATA3: the “glue” for MADS box transcription factor complex formation. *Genome Biol.* 10, R24. doi: 10.1186/gb-2009-10-2-r24
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kang, I.-H., Steffen, J. G., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant Cell* 20, 635–647. doi: 10.1105/tpc.107.055137
- Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198. doi: 10.1016/j.gene.2004.12.014
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Landoni, M., Bertagnon, G., Ghidoli, M., Cassani, E., Adani, F., and Pilu, R. (2023). Opportunities and challenges of castor bean (*Ricinus communis* L.) genetic improvement. *Agronomy* 13, 2076. doi: 10.3390/agronomy13082076
- Liao, Y., Smyth, G. K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47, e47. doi: 10.1093/nar/gkz114
- Lin, Y., Qi, X., Wan, Y., Chen, Z., Fang, H., and Liang, C. (2023). Genome-wide analysis of the MADS-box gene family in *Lonicera japonica* and a proposed floral organ identity model. *BMC Genomics* 24, 447. doi: 10.1186/s12864-023-09509-9
- Liu, J., Li, Y., Li, J., Chen, W., Pan, B., Liu, A., et al. (2024). EupDB: An integrative and comprehensive functional genomics data hub for Euphorbiaceae plants. *Plant Commun.* 5, 100683. doi: 10.1016/j.xplc.2023.100683
- Malabarba, J., Buffon, V., Mariath, J. E. A., Gaeta, M. L., Dornelas, M. C., Margis-Pinheiro, M., et al. (2017). The MADS-box gene *Agamous-like 11* is essential for seed morphogenesis in grapevine. *J. Exp. Bot.* 68, 1493–1506. doi: 10.1093/jxb/erx025
- Marzo, M. D., Herrera-Ubaldo, H., Caporali, E., Novák, O., Strnad, M., Balanzá, V., et al. (2020). SEEDSTICK controls arabidopsis fruit size by regulating cytokinin levels and FRUITFULL. *Cell Rep.* 30, 2846–2857.e3. doi: 10.1016/j.celrep.2020.01.101
- Nam, J., dePamphilis, C. W., Ma, H., and Nei, M. (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.* 20, 1435–1447. doi: 10.1093/molbev/msg152
- Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55, 989–1003. doi: 10.1016/0092-8674(88)90244-9
- Paolo, D., Rotasperti, L., Schnittger, A., Masiero, S., Colombo, L., and Mizzotti, C. (2021). The arabidopsis MADS-domain transcription factor SEEDSTICK controls seed size via direct activation of E2Fa. *Plants* 10, 192. doi: 10.3390/plants10020192
- Parenticová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., 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, 1538–1551. doi: 10.1105/tpc.011544
- Passmore, S., Maine, G. T., Elble, R., Christ, C., and Tye, B.-K. (1988). Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT $\alpha$  cells. *J. Mol. Biol.* 204, 593–606. doi: 10.1016/0022-2836(88)90358-0
- Paul, P., Joshi, S., Tian, R., Diogo Junior, R., Chakrabarti, M., and Perry, S. E. (2022). The MADS-domain factor AGAMOUS-Like18 promotes somatic embryogenesis. *Plant Physiol.* 188, 1617–1631. doi: 10.1093/plphys/kiab553
- Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., et al. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110. doi: 10.1126/science.290.5499.2105
- Roszak, P., and Köhler, C. (2011). Polycomb group proteins are required to couple seed coat initiation to fertilization. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20826–20831. doi: 10.1073/pnas.11171111108
- Ruelens, P., de Maagd, R. A., Proost, S., Theissen, G., Geuten, K., and Kaufmann, K. (2013). FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nat. Commun.* 4, 2280. doi: 10.1038/ncomms3280
- Schilling, S., Kennedy, A., Pan, S., Jermini, L. S., and Melzer, R. (2020). Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* 225, 511–529. doi: 10.1111/nph.16122
- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *antirrhinum majus*. *Science* 250, 931–936. doi: 10.1126/science.250.4983.931
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shen, G., Jia, Y., and Wang, W.-L. (2021). Evolutionary divergence of motifs in B-class MADS-box proteins of seed plants. *J. Biol. Res. (Thessalon)* 28, 12. doi: 10.1186/s40709-021-00144-7
- Shirzadi, R., Andersen, E. D., Bjerkan, K. N., Gloeckle, B. M., Heese, M., Ungru, A., et al. (2011). Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of AGAMOUS-LIKE36. *PLoS Genet.* 7, 1–18. doi: 10.1371/journal.pgen.1001303
- Smaczniak, C., Immink, R. G. H., Angenent, G. C., and Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* 139, 3081–3098. doi: 10.1242/dev.074674
- Steffen, J. G., Kang, I.-H., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). AGL61 interacts with AGL80 and is required for central cell development in arabidopsis. *Plant Physiol.* 148, 259–268. doi: 10.1104/pp.108.119404
- Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., et al. (2022). WGD: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* 15, 1841–1851. doi: 10.1016/j.molp.2022.10.018
- Theissen, G., Melzer, R., and Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* 143, 3259–3271. doi: 10.1242/dev.134080
- Tobimatsu, Y., Chen, F., Nakashima, J., Escamilla-Treviño, L. L., Jackson, L., Dixon, R. A., et al. (2013). Coexistence but independent biosynthesis of catechyl and guaiacyl/syringyl lignin polymers in seed coats. *Plant Cell* 25, 2587–2600. doi: 10.1105/tpc.113.113142
- Wan, Z., Qin, B., Lin, M., Chi, X., Shen, P., Wang, X., et al. (2025). Genome-wide analysis of MADS-box genes reveals their roles in floral organ development in Lagerstroemia. *Sci. Hortic.* 339, 113887. doi: 10.1016/j.scienta.2024.113887
- Wang, Y., Jiang, H., and Wang, G. (2020b). PHERES1 controls endosperm gene imprinting and seed development. *Trends Plant Sci.* 25, 517–519. doi: 10.1016/j.tplants.2020.03.004
- Wang, S., Su, S., Xiao, L., Wang, B., Sun, R., and Song, G. (2020a). Catechyl lignin extracted from castor seed coats using deep eutectic solvents: characterization and depolymerization. *ACS Sustain. Chem. Eng.* 8, 7031–7038. doi: 10.1021/acsuschemeng.0c00462
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293
- Wells, C. E., Vendramin, E., Jimenez Tarodo, S., Verde, I., and Bielenberg, D. G. (2015). A genome-wide analysis of MADS-box genes in peach [*Prunus persica* (L.) Batsch]. *BMC Plant Biol.* 15, 41. doi: 10.1186/s12870-015-0436-2
- Xanthopoulou, A., Moysiadi, T., Bazakos, C., Karagiannis, E., Karamichali, I., Stamatakis, G., et al. (2022). The perennial fruit tree proteogenomics atlas: a spatial map of the sweet cherry proteome and transcriptome. *Plant J.* 109, 1319–1336. doi: 10.1111/tpj.15612
- Xu, W., Wu, D., Yang, T., Sun, C., Wang, Z., Han, B., et al. (2021). Genomic insights into the origin, domestication and genetic basis of agronomic traits of castor bean. *Genome Biol.* 22, 113. doi: 10.1186/s13059-021-02333-y
- Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., and Meyerowitz, E. M. (1990). The protein encoded by the Arabidopsis homeotic gene

agamous resembles transcription factors. *Nature* 346, 35–39. doi: 10.1038/346035a0

Yu, A., Wang, Z., Zhang, Y., Li, F., and Liu, A. (2019). Global gene expression of seed coat tissues reveals a potential mechanism of regulating seed size formation in castor bean. *Int. J. Mol. Sci.* 20, 1282. doi: 10.3390/ijms20061282

Yu, A., Zhou, Z., Chen, Y., Sun, J., Li, P., Gu, X., et al. (2023). Functional genome analyses reveal the molecular basis of oil accumulation in developing seeds of castor beans. *Int. J. Mol. Sci.* 25, 92. doi: 10.3390/ijms25010092

Zhang, A., He, H., Li, Y., Wang, L., Liu, Y., Luan, X., et al. (2023). MADS-box subfamily gene gmAP3 from glycine max regulates early flowering and flower development. *Int. J. Mol. Sci.* 24, 2751. doi: 10.3390/ijms24032751

Zhang, X., He, W., Wang, X., Duan, Y., Li, Y., Wang, Y., et al. (2024b). Genome-wide analyses of MADS-box genes reveal their involvement in seed development and oil accumulation of tea-oil tree (*Camellia oleifera*). *Int. J. Genomics* 2024, 3375173. doi: 10.1155/2024/3375173

Zhang, S., Mohanty, D., Muzaffar, A., and Ni, M. (2024a). Two MADS-box proteins, AGL9 and AGL15, recruit the FIS-PRC2 complex to trigger the phase transition from endosperm proliferation to embryo development in *Arabidopsis*. *Mol. Plant* 17, 1110–1128. doi: 10.1016/j.molp.2024.05.011

Zhang, S., Wang, D., Zhang, H., Skaggs, M. I., Lloyd, A., Ran, D., et al. (2018). FERTILIZATION-INDEPENDENT SEED-polycomb repressive complex 2 plays a dual role in regulating type I MADS-box genes in early endosperm development. *Plant Physiol.* 177, 285–299. doi: 10.1104/pp.17.00534