

Genome-wide identification of a MADS-box transcription factor family and their expression during floral development in *Coptis teeta wall*

Item Type	Journal article
Authors	Duan, Shao-Feng;Yu, Ji-Chen;Baldwin, Timothy;Yuan, Yuan;Xiang, Gui-Sheng;Cui, Rui;Zhao, Yan;Mo, Xin-Chun;Lu, Ying-Chun;Liang, Yan-Li
Citation	Duan, SF., Yu, JC., Baldwin, T.C. et al. Genome-wide identification of a MADS-box transcription factor family and their expression during floral development in <i>Coptis teeta wall</i> . BMC Plant Biology 24, 1023 [2024]. https://doi.org/10.1186/s12870-024-05714-0
DOI	10.1186/s12870-024-05714-0
Publisher	BMC
Journal	BMC Plant Biology
Download date	2026-04-14 05:42:40
License	https://creativecommons.org/licenses/by-nc-nd/4.0/
Link to Item	http://hdl.handle.net/2436/625755

RESEARCH

Open Access



Genome-wide identification of a MADS-box transcription factor family and their expression during floral development in *Coptis teeta* wall

Shao-Feng Duan^{1,2,3}, Ji-Chen Yu^{1,2,3}, Timothy Charles Baldwin⁶, Yuan Yuan^{1,2,3}, Gui-Sheng Xiang^{2,3}, Rui Cui⁴, Yan Zhao^{1,2,3}, Xin-Chun Mo^{5,8*}, Ying-Chun Lu^{7,8*} and Yan-Li Liang^{1,2,3,8*}

Abstract

Background MADS-box transcription factors have been shown to be involved in multiple developmental processes, including the regulation of floral organ formation and pollen maturation. However, the role of the MADS-box gene family in floral development of the alpine plant species *Coptis teeta* Wall, which is widely used in Traditional Chinese Medicine (TCM), is unknown.

Results Sixty-six MADS-box genes were identified in the *C. teeta* genome. These genes were shown to be unevenly distributed throughout the genome of *C. teeta*. The majority of which (49) were classified as type I MADS-box genes and were further subdivided into four groups (Ma, M β , My and M δ). The remainder were identified as belonging to the type II MADS-box gene category. It was observed that four pairs of segmental and tandem duplication had occurred in the *C. teeta* MADS-box gene family, and that the ratios of Ka/Ks were less than 1, suggesting that these genes may have experienced purifying selection during evolution. Gene expression profiling analysis revealed that 38 MADS-box genes displayed differential expression patterns between the M and F floral phenotypes. Sixteen of these MADS-box genes were further verified by RT-qPCR. The 3D structure of each subfamily gene was predicted, further indicating that MADS-box genes of the same type possess structural similarities to the known template.

Conclusions These data provide new insights into the molecular mechanism of dichogamy and herkogamy formation in *C. teeta* and establish a solid foundation for future studies of the MADS-box genes family in this medicinal plant species.

Keywords *Coptis teeta*, MADS-box transcription factor, Genome-wide, Floral development, RT-qPCR

*Correspondence:

Xin-Chun Mo
26315125@qq.com
Ying-Chun Lu
351545297@qq.com
Yan-Li Liang
943029567@qq.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Transcription factors (TFs) regulate gene expression and play a variety of roles in higher plant species [1, 2]. They are extensively involved in plant growth and development, organ morphogenesis, stress, and hormone signaling responses [3–5]. The MADS-box gene family is one of the largest eukaryotic transcription factor families, and plays an important role in the regulation of a plethora of developmental processes [6], including flower initiation [7, 8], floral organ development [9], pollen maturation and pollen tube growth [10]. The MADS-box family can be divided into two groups, based on the differences in the highly conserved DNA-binding MADS-domain, type I and type II.

Type I MADS-box transcription factors only possess a simple SRF-like MADS domain, whereas type II has a complex characteristic domain: MADS-box (M) domain, intervening (I) domain, keratin-like (K) domain and a variable C-terminal (C) region [11–15].

Type I MADS-box transcription factors can be further classified into four subfamilies ($M\alpha$, $M\beta$, $M\gamma$ and $M\delta$), by the M domain variance of the encoded protein [7, 13]. Type II MADS-box transcription factors were named as MIKC type, derived from their four plant-specific characteristic domains and have been further divided into thirteen subclasses [16, 17].

MADS-box transcription factors (TFs) are widely believed to serve an important function in flowering and floral organ development. In hermaphrodite flowers, herkogamy and dichogamy are common phenomena during the formation of the floral organs.

Herkogamy is the spatial separation and dichogamy is the temporal separation of the stamens and pistils [18], which result in the reproductive assurance of high stringency outbreeding [19–21]. Although dichogamy and herkogamy have been extensively studied, the molecular mechanism by which these processes are regulated is still unclear, especially in less widely studied, non-model species of angiosperm.

Coptis teeta Wall. (Yunnan goldthread), is a perennial, alpine plant species indigenous to the low-latitude, alpine environment of southwest China (E 95 °48, N 25 °09 to E 98 °52, N 29 °51), which belongs to the genus *Coptis* within the family Ranunculaceae [22]. Interestingly, we discovered two floral phenotypes to be present in the Yunnanese population of *C. teeta*; the first type possessed a long pistil with short stamens (F-type) and the second possessed a short pistil with long stamens (M-type). This indicated that this population of *C. teeta* displayed a ‘typical’ herkogamy and dichogamy phenotype. As such, we believe that *C. teeta* may represent a novel system, for the investigation of floral development in general, and the processes of herkogamy and dichogamy in particular.

MADS-box genes are reported to participate in floral organ specificity during angiosperm development [23]. The gene (*CiMADS43*) was found to regulate flowering time in *Arabidopsis* [24]. Moreover, the MADS domain heterodimers of the $M\delta$ subfamily were proven to play an important role in pollen maturation [25–27]. The MADS-box gene (*AGL6*) was shown to fulfill an important role in floral organ identity [28–30]. However, to date, no such information concerning the MADS-box gene family and its potential role in floral development in *C. teeta* has been reported.

Therefore, in order to gain an insight into the *C. teeta* MADS-box gene family, a genome-wide exploration methodology was employed. Sixty-six MADS-box genes from *C. teeta* were shown to be highly homologous to MADS-box proteins previously identified in other plant species. The genetic and molecular characteristics of these MADS-box genes, including phylogenetic relationships, gene structure, conserved motifs, exon-intron organization, and chromosomal location prediction were then elucidated. Furthermore, in order to investigate the potential function of these genes in floral organ formation in *C. teeta*, the expression pattern of *C. teeta* MADS-box genes related to floral organ identity was studied and verified by real-time quantitative polymerase chain reactions (RT-qPCR) during flower development. These data may assist future elucidation of the molecular mechanism of herkogamy and dichogamy in this species. A flowchart depicting the experimental approach used in the current study is presented in Supplementary Fig. 1.

Results

Genome-wide exploration and identification of MADS-box genes in *C. teeta*

The Hidden Markov Model (HMM) was employed to identify MADS-box genes present in the *C. teeta* genome. A total of 66 sequences were obtained using the MADS domain profile PF00319 as a query against the *C. teeta* protein database (Table 1). Amongst which, 49 genes were identified as belonging to the type I MADS-box gene family, which were further classified into four subfamilies. Over 50% (27/49) of the putative type I MADS-box genes were shown to belong to the $M\alpha$ subfamily, approximately 20% (10/49) were grouped into the $M\beta$ subfamily and a further 20% (10/49) were grouped into the $M\gamma$ subfamily. A very small proportion of the genes, 0.4% (2/49) were identified as belonging to the $M\delta$ subfamily. In addition, over a quarter (25.76%, 17/66) were identified as members of the type II group of MADS-box genes (Table 1). The molecular characteristics of the 66 MADS-box genes were then analyzed in more detail.

The results of this more in-depth analysis, revealed that the amino acid sequences of the 66 predicted MADS-box proteins varied from 100 (*CteMADS48*)

Table 1 Overview of MADS-box genes identified in *C. teeta* genome

Gene ID	Name	Type	Protein		
			Length (aa)	MW(KDa)	pI
Cte01G002407.t1	CteMADS01	TypeI (Ma)	270	30.46	4.35
Cte01G002336.t1	CteMADS02	TypeI (Ma)	175	19.88	6.14
Cte01G001931.t1	CteMADS03	TypeI (Mγ)	230	25.95	9.28
Cte01G001798.t1	CteMADS04	TypeI (Ma)	264	29.86	4.27
Cte01G001690.t1	CteMADS05	TypeI (Ma)	264	29.86	4.27
Cte01G001939.t1	CteMADS06	TypeI (Mγ)	230	25.95	9.28
Cte01G002507.t1	CteMADS07	TypeI (Ma)	257	28.94	4.53
Cte01G002498.t1	CteMADS08	TypeI (Ma)	255	28.91	4.78
Cte02G003740.t1	CteMADS09	TypeI (Mβ)	365	42.14	9.12
Cte02G003752.t1	CteMADS10	TypeI (Mβ)	329	38.28	8.57
Cte02G002649.t1	CteMADS11	TypeI (Mγ)	230	25.95	8.93
Cte02G000124.t1	CteMADS12	TypeI (Mγ)	232	26.77	6.02
Cte02G000126.t1	CteMADS13	TypeI (Mγ)	242	27.84	5.44
Cte02G004466.t1	CteMADS14	TypeI (Ma)	228	25.59	9.18
Cte03G002403.t1	CteMADS15	TypeI (Ma)	200	22.64	9.49
Cte03G003343.t1	CteMADS16	TypeI (Mβ)	122	14.60	6.77
Cte04G005234.t1	CteMADS17	TypeI (Ma)	200	22.67	9.56
Cte04G000799.t1	CteMADS18	TypeI (Mδ)	337	38.41	5.60
Cte04G003145.t1	CteMADS19	TypeI (Mγ)	230	26.19	9.49
Cte04G004575.t1	CteMADS20	TypeI (Ma)	342	39.08	5.59
Cte05G002958.t1	CteMADS21	TypeI (Ma)	249	28.11	9.03
Cte06G000790.t1	CteMADS22	TypeI (Mβ)	1206	137.92	6.48
Cte06G001381.t1	CteMADS23	TypeI (Ma)	301	34.07	4.73
Cte06G001749.t1	CteMADS24	TypeI (Mβ)	227	25.95	5.22
Cte06G001699.t1	CteMADS25	TypeI (Mβ)	227	25.95	5.22
Cte06G001038.t1	CteMADS26	TypeI (Mγ)	169	19.34	9.69
Cte06G001925.t1	CteMADS27	TypeI (Mγ)	1873	208.92	7.51
Cte06G001382.t1	CteMADS28	TypeI (Ma)	232	26.46	6.45
Cte06G001245.t1	CteMADS29	TypeI (Mβ)	743	85.44	7.79
Cte07G002817.t1	CteMADS30	TypeI (Mβ)	199	22.12	4.64
Cte07G004663.t1	CteMADS31	TypeI (Ma)	340	38.43	4.23
Cte07G000391.t1	CteMADS32	TypeI (Ma)	218	24.70	9.13
Cte07G000389.t1	CteMADS33	TypeI (Ma)	218	24.97	9.31
Cte07G000390.t1	CteMADS34	TypeI (Ma)	753	82.52	8.42
Cte07G000384.t1	CteMADS35	TypeI (Ma)	192	21.50	6.85
Cte07G000381.t1	CteMADS36	TypeI (Ma)	191	21.32	6.85
Cte07G000383.t1	CteMADS37	TypeI (Ma)	504	55.98	5.73
Cte07G000392.t1	CteMADS38	TypeI (Ma)	190	21.61	9.17
Cte07G002168.t1	CteMADS39	TypeI (Mβ)	199	22.93	5.06
Cte07G000388.t1	CteMADS40	TypeI (Ma)	670	73.44	5.78
Cte08G001368.t1	CteMADS41	TypeI (Mδ)	351	39.57	6.40
Cte09G001067.t1	CteMADS42	TypeI (Mγ)	246	28.05	9.37
Cte09G003719.t1	CteMADS43	TypeI (Ma)	465	52.25	4.51
Cte09G001151.t1	CteMADS44	TypeI (Mγ)	230	26.73	8.62
Cte09G003853.t1	CteMADS45	TypeI (Ma)	220	24.92	4.68
Cte09G001297.t1	CteMADS46	TypeI (Ma)	227	25.70	8.99
Cte09G002568.t1	CteMADS47	TypeI (Ma)	383	42.78	4.20
Cte09G002067.t1	CteMADS48	TypeI (Mβ)	100	11.23	5.66
Cte47G000003.t1	CteMADS49	TypeI (Ma)	202	22.81	9.36
Cte01G005971.t1	CteMADS50	TypeII (MIKC)	241	28.28	8.90
Cte02G005328.t1	CteMADS51	TypeII (MIKC)	222	25.74	8.33
Cte02G000891.t1	CteMADS52	TypeII (MIKC)	242	27.83	9.15

Table 1 (continued)

Gene ID	Name	Type	Protein		
			Length (aa)	MW(KDa)	pI
Cte02G005330.t1	CteMADS53	Typell (MIKC)	233	26.63	7.71
Cte02G004047.t1	CteMADS54	Typell (MIKC)	257	29.27	9.57
Cte03G004943.t1	CteMADS55	Typell (MIKC)	234	26.75	7.58
Cte03G004941.t1	CteMADS56	Typell (MIKC)	247	28.31	9.11
Cte03G004930.t1	CteMADS57	Typell (MIKC)	254	29.24	8.96
Cte03G004957.t1	CteMADS58	Typell (MIKC)	234	26.75	7.58
Cte05G000515.t1	CteMADS59	Typell (MIKC)	531	60.57	5.61
Cte06G001056.t1	CteMADS60	Typell (MIKC)	612	69.29	9.25
Cte06G001042.t2	CteMADS61	Typell (MIKC)	245	28.09	8.73
Cte06G005318.t1	CteMADS62	Typell (MIKC)	247	28.15	8.92
Cte07G002352.t1	CteMADS63	Typell (MIKC)	219	25.80	8.96
Cte07G001694.t1	CteMADS64	Typell (MIKC)	226	25.69	6.93
Cte08G000912.t1	CteMADS65	Typell (MIKC)	244	27.81	8.28
Cte08G000913.t1	CteMADS66	Typell (MIKC)	247	28.52	9.44

Note: pI, isoelectric point; MW, molecular weight

to 1873 (*CteMADS27*) amino acids (AAs) in length, and that the relative molecular weights ranged from 11.23 kDa (*CteMADS48*) to 208.92 kDa (*CteMADS27*). The predicted isoelectric point (pI) ranged from 4.2 (*CteMADS47*) to 9.69 (*CteMADS26*) (Table 1), indicating that the majority (36/66), encoded proteins with an alkaline (pI value > 7.5), 40% (26/66) were acidic (pI values < 6.5), and only four (6%) were predicted to exhibit a neutral pI, with predicted pI values between 6.5 and 7.5 (Table 1).

Phylogenetic analyses

In order to better understand the phylogenetic relationship of the *C. teeta* MADS-box genes to those present in other plant species, and to group them into the established subfamilies, two phylogenetic trees for the type I (49) and type II (17) MADS-box genes were constructed using the IQ-TREE method (Figs. 1 and 2). According to the phylogenetic trees, the type I MADS-box genes could be subdivided into four subfamilies (M α , M β , M γ and M δ) (Fig. 1). The type II MADS-box genes (MIKC) could be further subdivided into nine subfamilies, namely the AG, AP3/PI, Bs/TT16, API/FUL, AGL6, SEP, AGL15, SVP and SOC1 subfamilies (Fig. 2) AGL6 and SEP were the largest subfamilies, both with three members whereas Bs/TT16, AG, AGL15 and SVP each had only one member. Compared with the number of type II MADS-box genes numbers previously recorded in *Coptis chinensis* and *Aquilegia coerulea*, *C. teeta* had fewer members in the AG, SOC1 and AP3 subfamilies.

To investigate the relationship of the MADS-box genes present in *C. teeta* and other, selected plant species (*Arabidopsis thaliana*, *Coptis chinensis* and *Aquilegia coerulea*), comparative synteny analysis of the entire genome sequences of these species was conducted. The results

demonstrated that the MADS-box genes in *C. teeta* had the highest collinearity with *C. chinensis*. Nearly 60% (38/66) of MADS-box genes in *C. teeta* showed a syntenic relationship to *C. chinensis* (Fig. 3), followed by 25 MADS-box genes with a syntenic relationship to *A. coerulea* and 9 genes to those present in *Arabidopsis thaliana*.

Gene structure and conserved motifs

In order to investigate the structural diversity and evolution of *C. teeta* MADS-box genes, the exon-intron structure was analyzed using TBtools (Fig. 4 and Supplementary Table S1). The results showed that the gene structure of *C. teeta* MADS-box genes was quite complex, with differing exon and intron distributions observed between the type I and type II class genes. The number of introns in *C. teeta* MADS-box genes was shown to range from 0 to 27, and the number of exons ranged from 1 to 28. Most type I genes were quite simple and only contained one or two exons, 68.8% contained no introns. However, five type I genes displayed a complex gene structure, *CteMADS27* (M γ) contained 28 exons and 27 introns, *CteMADS22* (M β) possessed 15 exons and 14 introns, *CteMADS29* (M β) had 13 exons and 13 introns, and *CteMADS41* (M δ) and *CteMADS18* (M δ) contained 11 exons and 10 introns. The type II genes also contained multiple exons and introns 94% (16/17) contained at least 7 exons and 6 introns (Fig. 4).

In order to further investigate the structure of the *C. teeta* MADS-box genes, the conserved motifs of the predicted MADS-box proteins were estimated by using the Multiple EM for Motif Elicitation (MEME) program. A total of 20 conserved motifs were found in the *C. teeta* MADS-box proteins, which we named motif-1 to motif-20 (Fig. 5). Motif-1, motif-2 and motif-5 were conserved

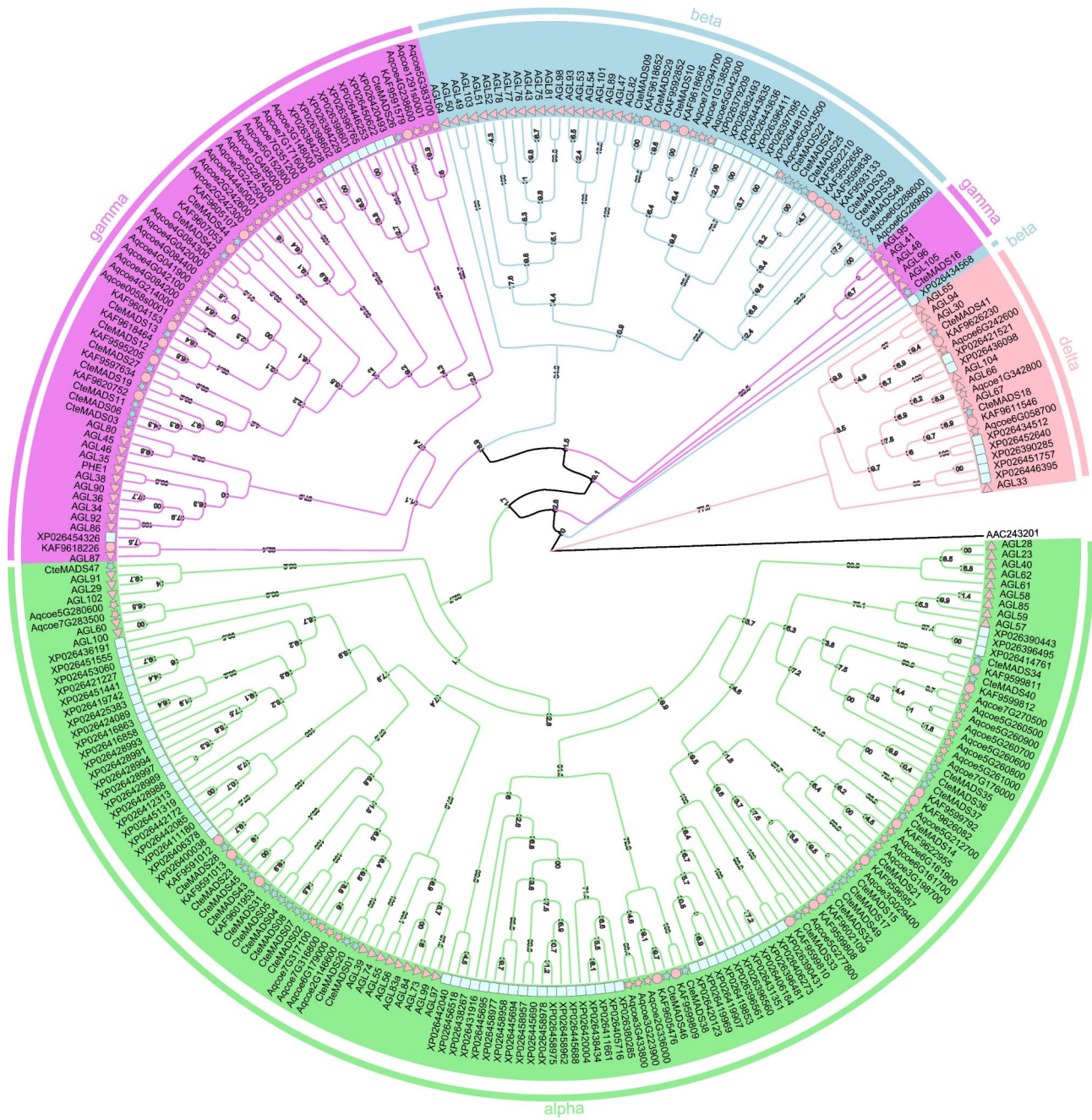


Fig. 1 Phylogenetic tree of type I MADS-box proteins
 Phylogenetic tree of type I MADS-box proteins among *C. teeta* (star, lightblue), *C. chinensis* (circle, pink), *A. coerulea* (star, pink), *P. somniferum* (rectangle, lightcyan) and *A. thaliana* (triangle, pink). The four subdivisions of the type I MADS-box proteins are represented by different colors, Ma (light green), Mβ (light blue), Mγ (violet), and Mδ (pink)

across most *C. teeta* MADS-box proteins, except for the Mγ subfamily, which had no motif-2 domain. Motif-6 and motif-11 were only present in type II MADS-box proteins (Fig. 5). The Mα subfamily contained unique motifs such as motif-4, motif-14, motif-15, motif-18, and motif-20, while the Mγ subfamily possessed the motif-3, motif-7, motif-8, and motif-13. Motif-16 was only present in the Mα and Mβ subfamilies, while the

Mβ subfamily was shown to contain motif-9 and motif-12. Motif-6, motif-10 and motif-17 were also conserved across most of the type II (MIKC) MADS-box proteins, except two proteins encoded by the genes *CteMADS05* and *CteMADS04*. Among the type II MADS-box proteins, motif-19 was only found to be present in three proteins encoded by *CteMADS51*, *CteMADS53* and *CteMADS63*, whereas the motif-10 was only present in

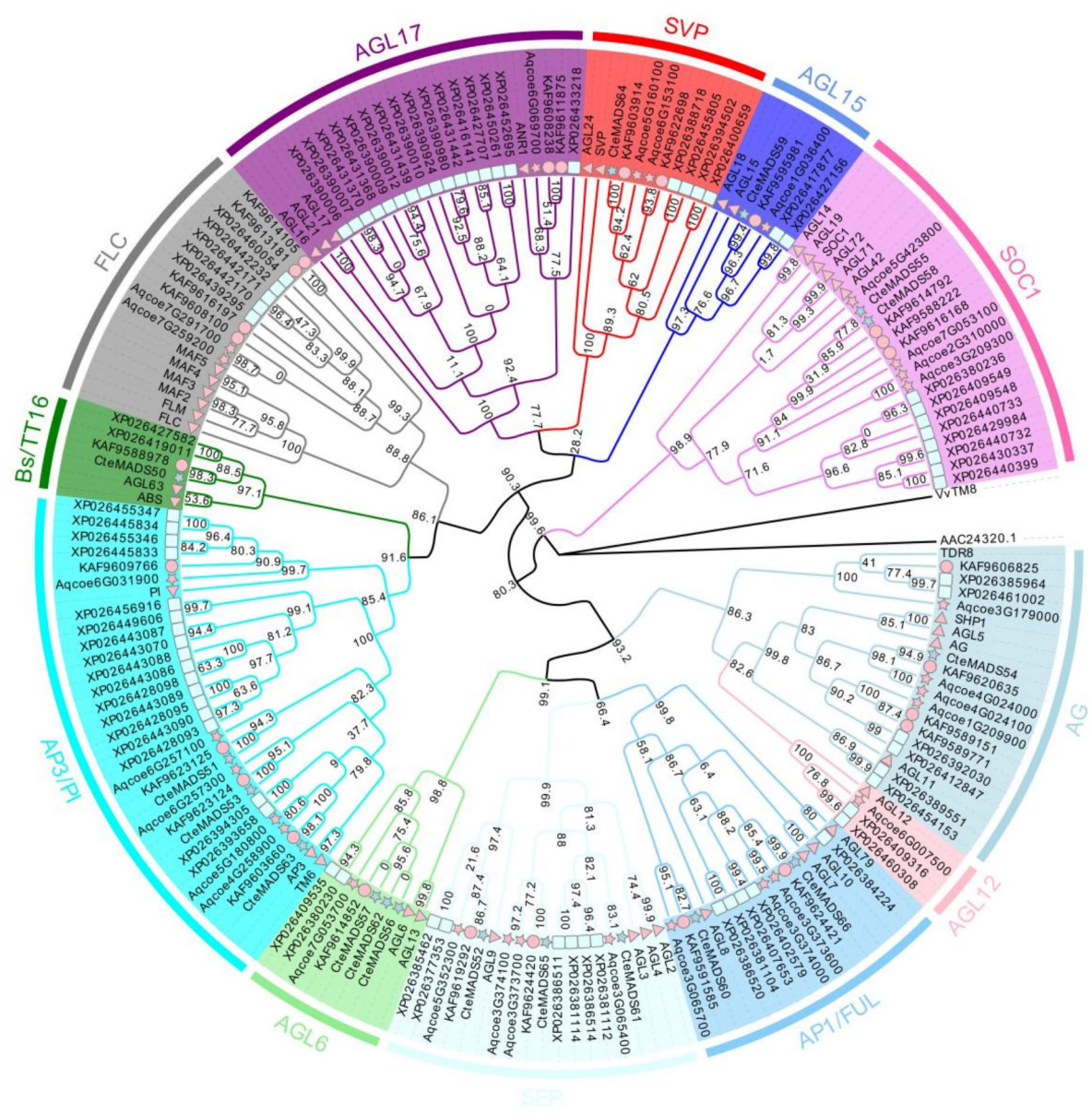


Fig. 2 Phylogenetic trees of type II MADS-box proteins
 Phylogenetic tree of the type II MADS-box proteins between among *C. teeta* (star, light blue), *C. chinensis* (circle, pink), *A. coerulea* (star, pink), *P. somniferum* (rectangle, light cyan) and *A. thaliana* (triangle, pink)

the proteins encoded by *CteMADS56*, *CteMADS57* and *CteMADS62*, respectively.

Chromosomal locations and duplication analysis

Based on the genome database of *C. teeta*, 65 MADS-box genes were located on chromosomes within the genome, and one gene (*CteMADS49*) was found to locate to an unanchored scaffold. The 65 MADS-box genes were shown to be unevenly distributed across the *C. teeta* genome (Fig. 6). The M α and M γ subfamilies of type I genes were widely distributed across the majority of the

C. teeta chromosomes, with the exception of chromosome 8. Similarly, the type II MADS-box genes were distributed on most of the *C. teeta* chromosomes, with the exception of chromosomes 4 and 9. As shown in Fig. 6, chromosome 7 possessed the highest number of MADS-box genes (13), while chromosome 5 contained only two MADS-box genes.

Gene duplication events represent gene family expansion and genome complexity in plants and often result in an increase in numbers of gene families. In the current study, we detected both segmental and tandem

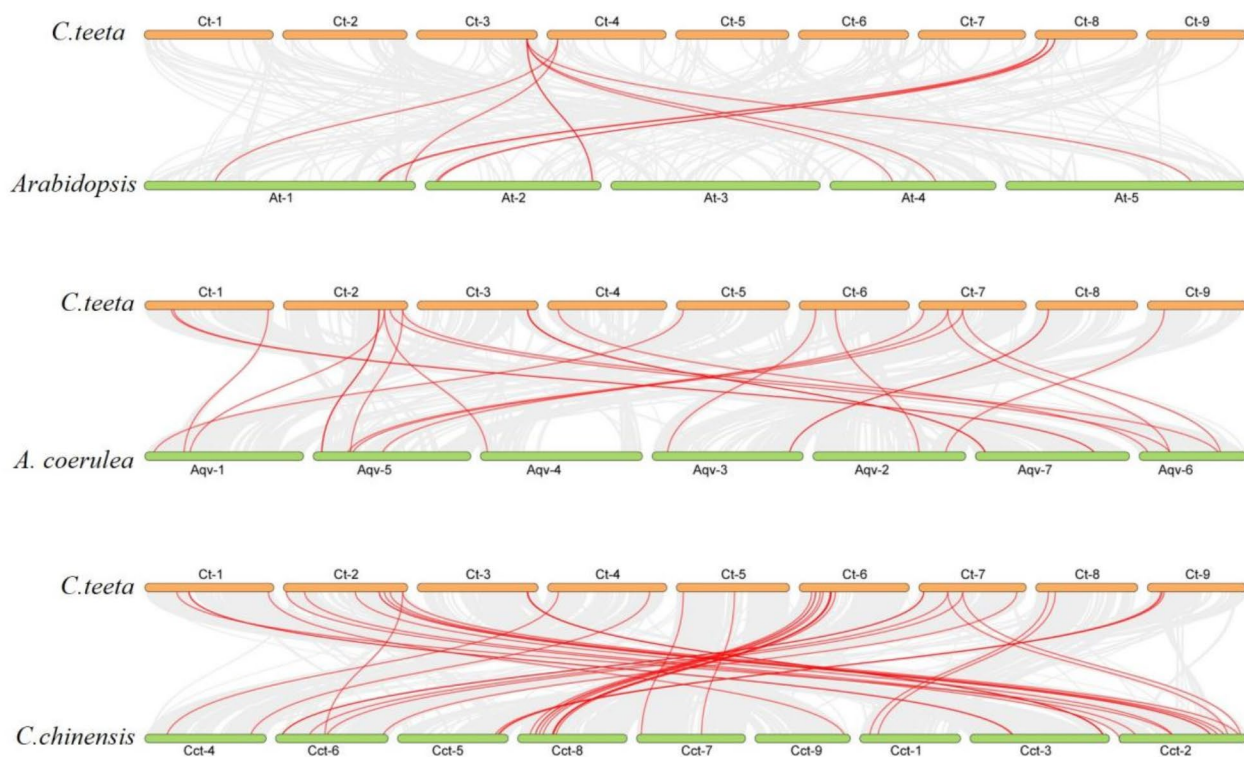


Fig. 3 Synteny analysis of MADS-box genes between *C. teeta* and three selected plant species

The collinearity between *C. teeta* and the other three plant genomes are shown as the gray lines in the background, while the syntenic MADS-box gene pairs are highlighted with red lines

duplication events of the identified MADS-box genes. The results showed that four pairs of genes were tandem duplicates, which were distributed on chromosomes 2 and 7. Four pairs of segmental duplications in the MADS-box genes were found, which were distributed on chromosomes 1, 2, 6, 7 and 9 respectively (Table 2).

Furthermore, to explore the evolutionary selection pressure experienced by MADS-box genes after the duplication, we calculated the substitution ratios of nonsynonymous (K_a), versus synonymous substitution (K_s) for each paralogous pair (Table 2). The K_a/K_s ratio of duplication gene pairs ranged from 0.165 (*CteMADS36* and *CteMADS38*) to 0.709 (*CteMADS37* and *CteMADS35*) with an average of 0.361 (Table 2). The ratios of K_a/K_s of all the MADS-box gene pairs were less than 1, suggesting that these genes may have experienced purifying selection during the process of evolution [31].

Expression profiling of MADS-box genes during floral organ development in *C. teeta*

As mentioned above, two floral phenotypes are observed in the *C. teeta* population indigenous to Yunnan province, China: one with long stamens and a short pistil (M-type) and a second with a long pistil and short stamens (F-type). To understand the potential involvement

of the MADS-box gene family in the development of these two floral phenotypes, we performed *C. teeta* transcriptome sequencing (RNA-seq) of pistils and stamens at a variety of stages of flower development (Fig. 7). A total of 38 MADS-box genes were shown to display differential patterns of expression between the M and F flower types. Thirty MADS-box genes were shown to be highly expressed in the M-type flowers, eight were observed to be highly expressed in the F-type and two (*CteMADS12*, *CteMADS48*) were not expressed in either. In the M-type plants, eight genes were more highly expressed in stamens than in pistils, and *CteMADS56* and *CteMADS21* showed highest expression at the mature stage of flower development, whilst *CteMADS40*, *CteMADS03*, *CteMADS06*, *CteMADS25*, *CteMADS57* and *CteMADS24* were shown to be highly expressed at the initial stage of stamen development (Fig. 7). The results suggest that these 8 genes may play a role in the formation of the M-type floral phenotype by controlling the development of the stamens. Moreover, four MADS-box genes (*CteMADS35*, *CteMADS26*, *CteMADS30* and *CteMADS62*) were found to be highly expressed in the pistils of the F-type and lowly expressed in the stamens of the M-type (Fig. 7). This suggests that these 4 genes may be involved in the formation of the phenotypic

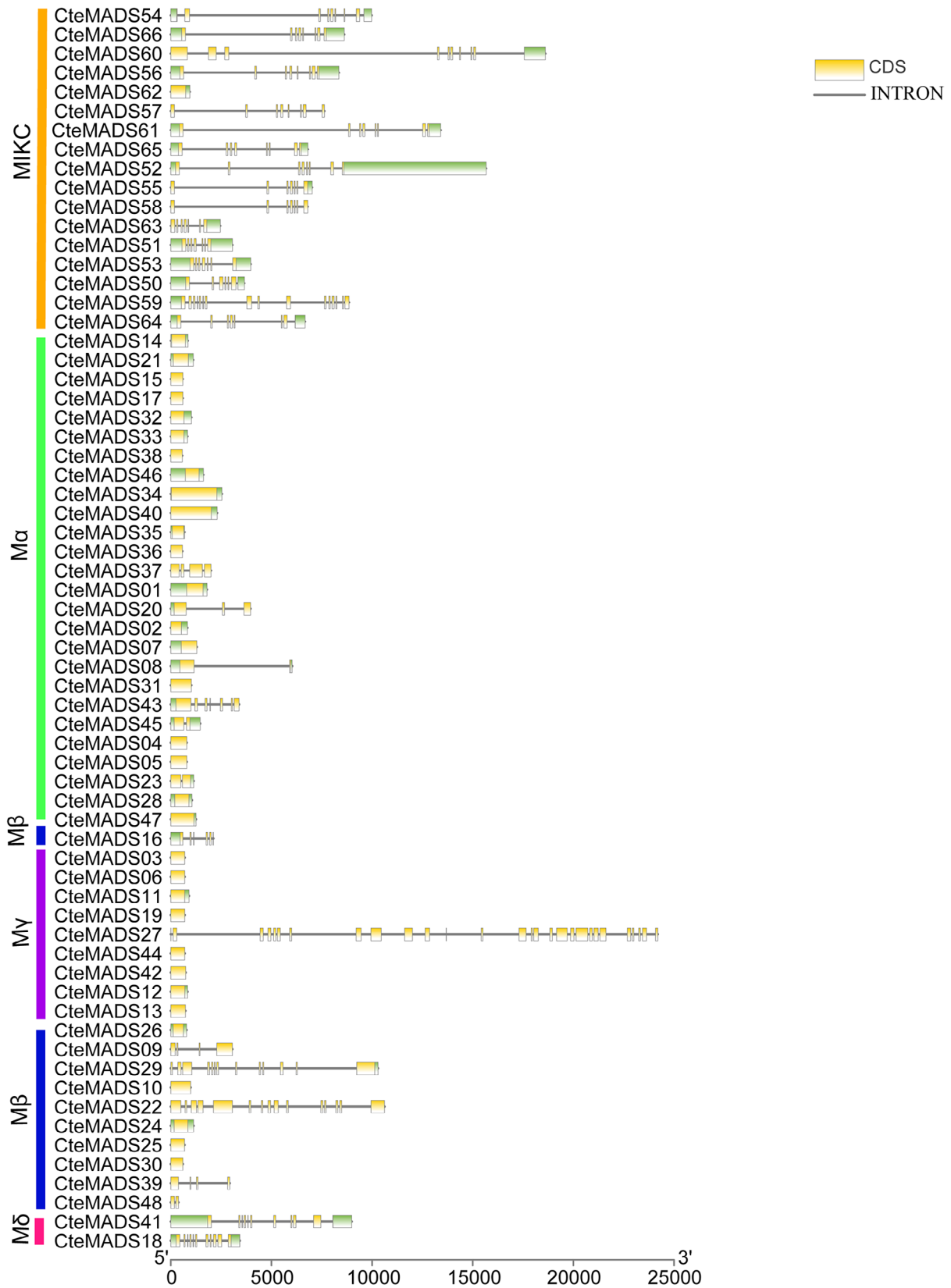


Fig. 4 Gene structural analysis of MADS-box genes
The yellow-white boxes represent CDS, and the black lines represent introns

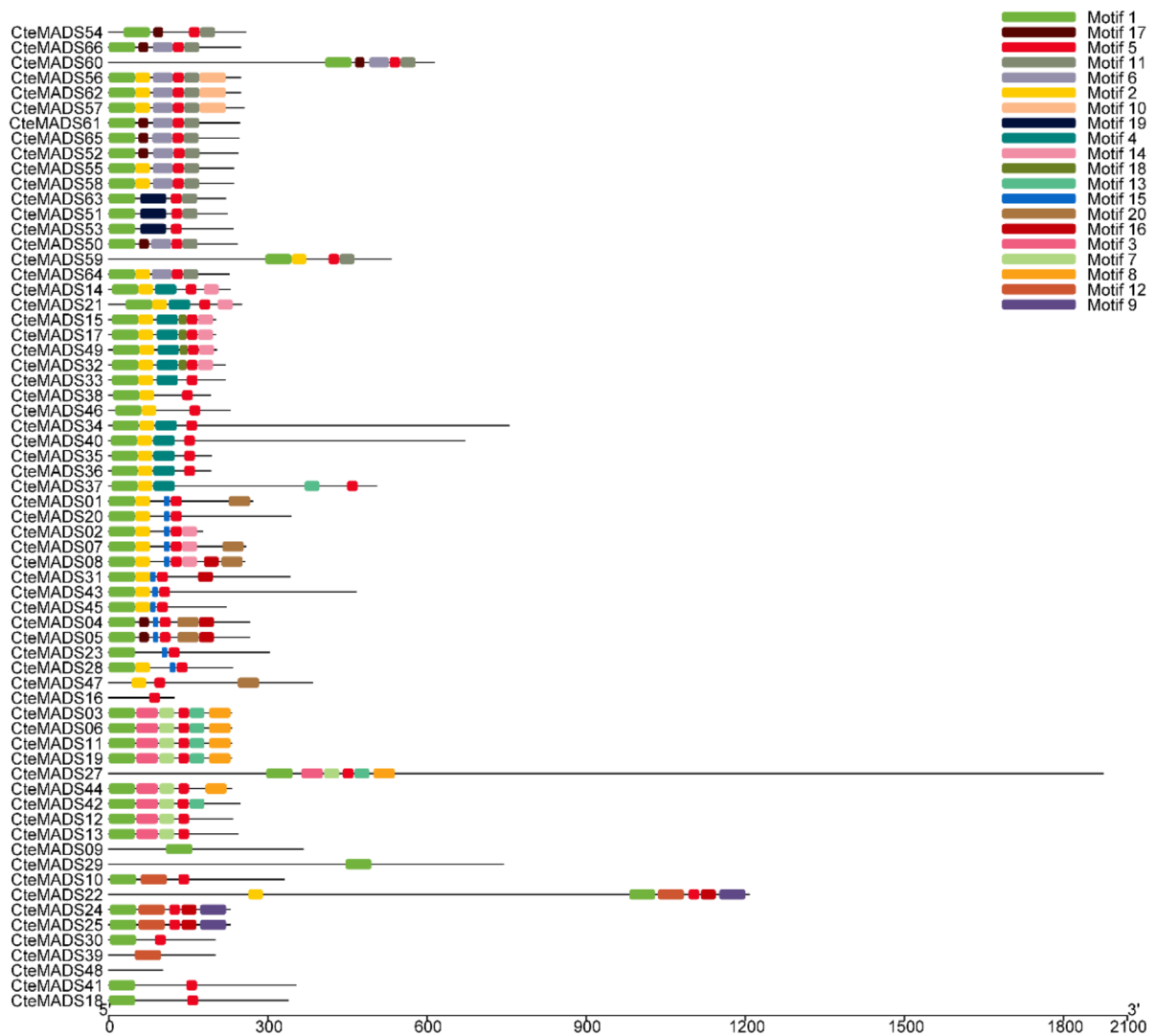


Fig. 5 Conserved motif analyses of MADS-box genes
 All the conserved motifs were predicted by the Multiple EM for Motif Elicitation (MEME) database, using the complete amino acid sequences of the *C. teeta* MADS-box genes. The twenty conserved motifs are indicated by different colors

differences between F and M type by controlling pistil development.

Expression analysis of *C. teeta* MADS-box genes at different stages of flower development

According to the expression profile predictions, sixteen *C. teeta* MADS-box genes were selected, and their expression patterns analyzed by RT-qPCR, at different stages of flower development, using both the M and F type flowers. The selected genes included thirteen type I genes and three type II genes. Amongst which, nine genes belong to the My subfamily (Table 1) and four genes to the Mα subfamily. The results indicated that most of the selected genes showed similar expression patterns to the RNA-seq expression profiles. Fifteen of the selected genes displayed differential patterns of expression between

the M and F flower types. The *CteMADS30(AGL80)*, *CteMADS45(AGL16)*, *CteMADS39*, *CteMADS56(RSB1)*, *CteMADS62(RSB1)*, *CteMADS24(AGL90)* and *CteMADS25(AGL90)* genes were most highly expressed in the M-type plants at full anthesis, with the expression level at this developmental stage being on average 55.6% higher in M-type flowers, than that observed in F-type flowers. These data suggest that these genes may be involved in the formation of the M-type floral phenotype (Fig. 8). In addition, *CteMADS26(AGL80)*, *CteMADS10(AGL47)*, *CteMADS01(AG)*, *CteMADS66(FUL)*, *CteMADS11(AGL6)*, *CteMADS44(AGL80)* showed higher levels of expression in M-type flowers at the initial stage of flower development, but were all highly expressed in the F-type plants in mature flowers. This suggests that these genes play

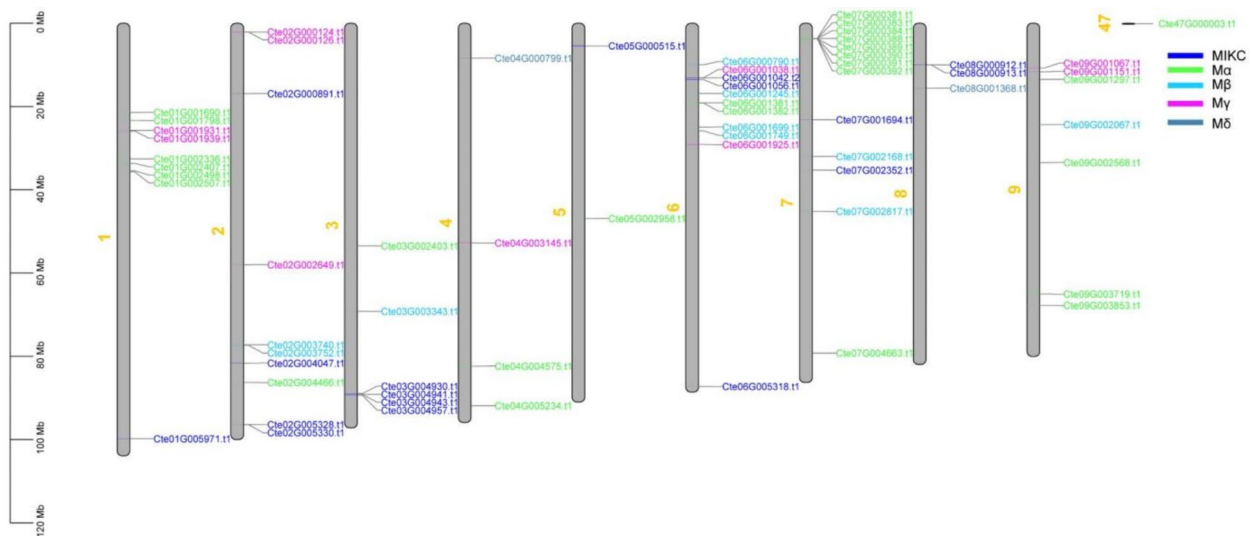


Fig. 6 Chromosomal location of *C. teeta* MADS-box gens

The position of the MADS-box genes on the nine chromosomes of *C. teeta* was based upon the annotation of the *C. teeta* genome database. The different subfamilies are labeled with different colors, MIKC (dark blue), Ma (green), Mβ (light blue), Mγ (rose red) and Mδ (blackish green)

Table 2 Analysis of duplication events of MADS-box gene pairs in *C. teeta*

Gene pairs	Ka value	Ks value	Ka/Ks	Duplicat-ed type
<i>CteMADS12</i> & <i>CteMADS13</i>	0.059	0.095	0.622	tandem
<i>CteMADS36</i> & <i>CteMADS37</i>	0.153	0.247	0.618	tandem
<i>CteMADS36</i> & <i>CteMADS35</i>	0.009	0.054	0.170	tandem
<i>CteMADS37</i> & <i>CteMADS35</i>	0.176	0.248	0.709	tandem
<i>CteMADS06</i> & <i>CteMADS11</i>	0.013	0.035	0.377	segment
<i>CteMADS51</i> & <i>CteMADS63</i>	0.321	1.006	0.319	segment
<i>CteMADS22</i> & <i>CteMADS25</i>	0.440	0.847	0.520	segment
<i>CteMADS40</i> & <i>CteMADS43</i>	0.489	0.940	0.520	segment

different roles at different developmental stages and were involved in the formation of both the F-type and M-type floral phenotype. It is also of interest to note that *CteMADS19*(*AGL80*), which was highly expressed at the mature stage of flower development (P4) displayed a higher level of expression in the F-type, conversely, *CteMADS21*(*AGL62*) which was also highly expressed at the mature stage of floral development showed higher expression in the M-type, both genes may therefore play a role in the maturation of pistils and stamens (Fig. 8).

To further demonstrate the significant role of *CteMADS26*, *CteMADS30*, and *CteMADS62* in the phenotypic differences between F and M types by controlling pistil development, we conducted fluorescence quantitative analysis in different types of pistils and stamens. The results indicated that the genes *CteMADS26* and *CteMADS30* were likely to play a crucial role in the phenotypic differences between F and M types by controlling pistil development (Fig. 8).

Structure prediction of *C. teeta* MADS-box protein

To further understand whether there were structural similarities among the subfamilies, we conducted a 3D structural analysis. Thirteen MADS-box proteins were selected to represent each subfamily. The secondary structure prediction and a 3D structure simulation were conducted, based on homology modelling. The predicted secondary structure of all the selected *C. teeta* MADS proteins was dominated by alpha helices and random coil regions (Table 3). The 3D structural stimulation data showed that type I protein Mδ (*CteMADS18*) and Mα (*CteMADS21*) were structurally similar to the template 7nb0.1 (Lai, X. et al.,2021), with similarities of 51.71% and 62.5%, respectively. The structure of type II proteins AP3 (*CteMADS51*), AG (*CteMADS54*), SOC1 (*CteMADS55*), AGL6 (*CteMADS56*), API (*CteMADS66*) was similar to that of the template 6byy.2 (Lei, X. et al.,2018), and their similarities were 44%, 51.32%, 43.02%, 52% and 51.32%, respectively (Fig. 9; Table 3).

Materials and methods

Plant material, DNA extraction and genome sequencing

Plant material was collected from the low-latitude, alpine environment of Yunnan Province, located in southwest China (E 95 °48, N 25 °09 to E 98 °52, N 29 °51). High-quality genomic DNA was prepared from fresh leaves of a single *C. teeta* plant, using a modified CTAB method that was subsequently used for Illumina, PacBio, and Hi-C library construction. Genomic data was acquired but is yet to be published. These materials were formally identified by Researcher Wenguang Yang (Kunming Institute of Botany). The plant specimens of the materials used in

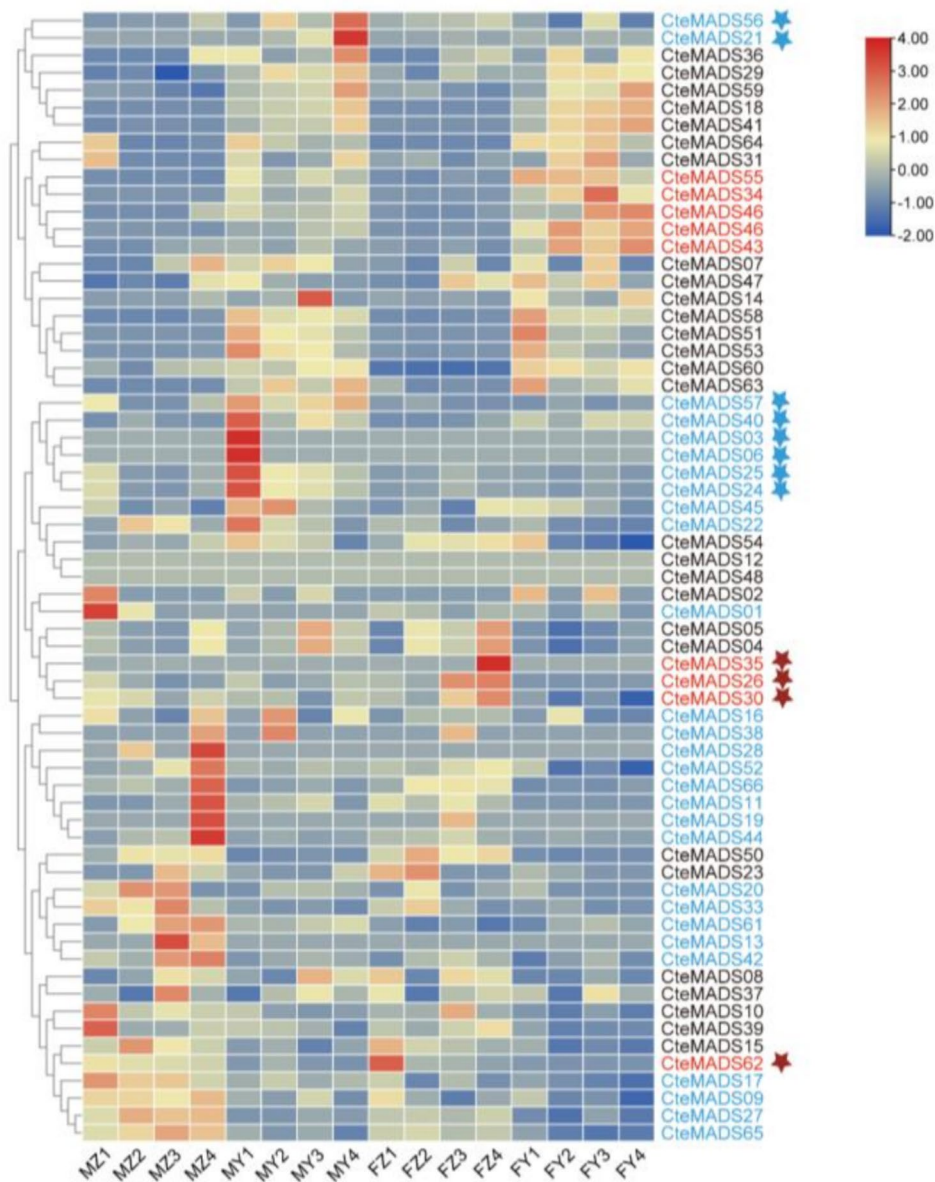


Fig. 7 Heat map representation of MADS-box genes within the two floral phenotypes of *C. teeta* at different stages of floral development M: short pistil and long stamen genotype of *C. teeta*; F: long pistil and short stamen genotype; Z: pistil; Y: stamen; stage 1: The bracts were dehiscent and florets unexposed; stage 2: The florets were just exposed, and sepals were unexpanded; stage 3: The sepals had fully expanded; stage 4: The anthers had dehisced and had begun to disperse pollen. Blue font: expressed highly in the M type; Red Font: expressed highly in the F type; The blue five-pointed star: indicates higher levels of gene expression in the stamens than in pistils of M type; The red five-pointed star: indicates higher levels of gene expression in the pistils of the F type and low gene expression in the stamens of the M type

the experiment are stored in the Southwest Biodiversity Laboratory (No: XINAN201901706).

The test materials used for total RNA extraction for the transcriptome sequencing and RT-qPCR were taken at four stages of floral development: stage 1. In which the bracts were dehiscent and florets unexposed; stage 2: In which the florets were just exposed, and the sepals were unexpanded; stage 3: In which the sepals had fully expanded; stage 4: In which the anthers had dehisced and begun to disperse pollen (Fig. 10). The pistils and

stamens of M and F type flowers were used for transcriptome sequencing and whole, intact flowers of M and F types were used for the RT-qPCR.

Identification and classification of MADS-box genes in *C. teeta*

The hidden markov model (HMM) of the MADS-box domains was retrieved from the Pfam database (PF00319) [32], and the MADS-box genes were identified using TBtools software [33]. The ProtParam tool (<https://web>.

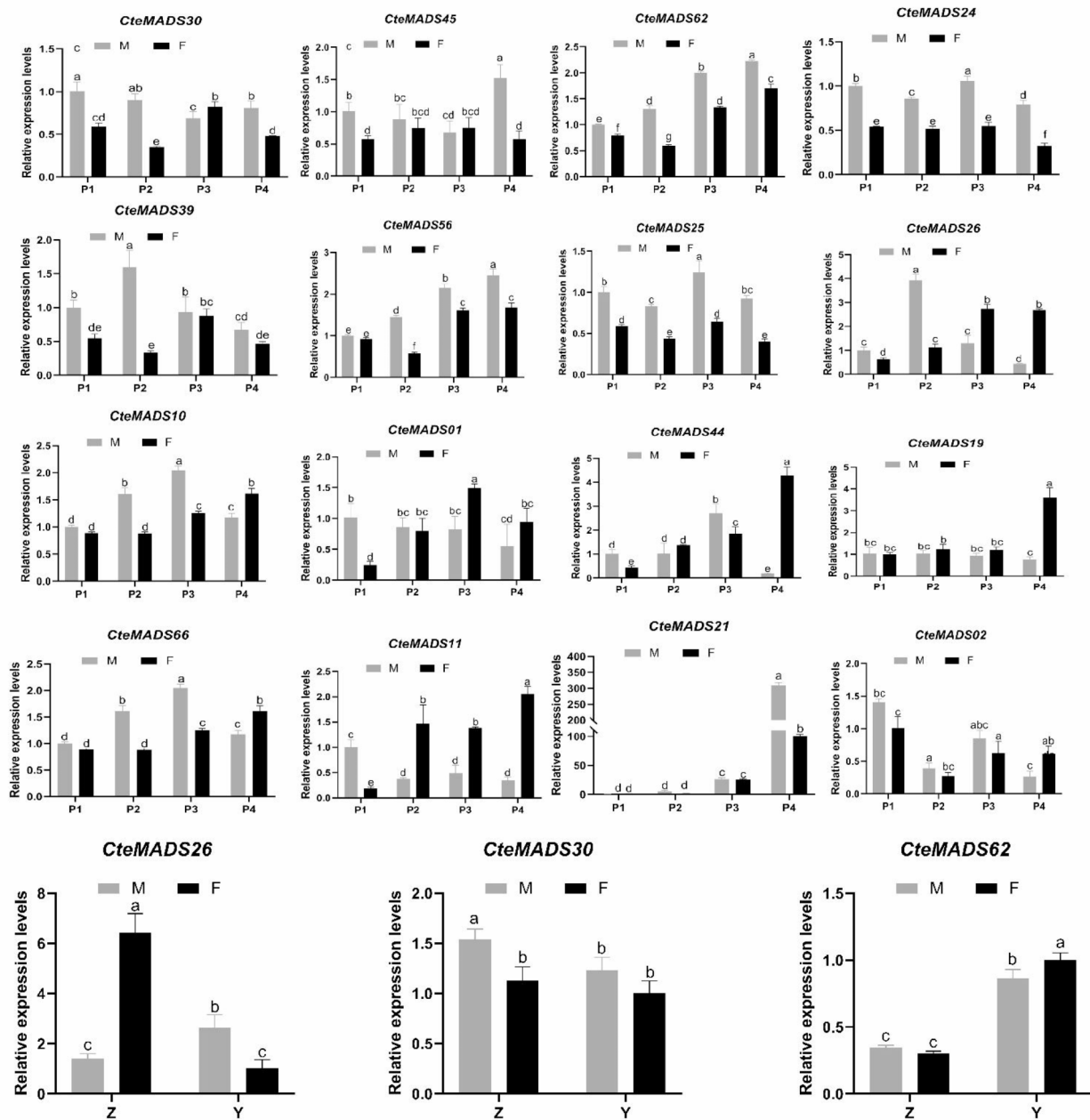


Fig. 8 Expression profiles of selected MADS-box genes analysed by RT-qPCR analysis in different tissues of *C. teeta*. M: short pistil and long stamen genotype of *C. teeta*; F: long pistil and short stamen genotype; Z: pistil; Y: stamen; The error bars indicate the standard deviation (SD). The lowercase letters (a-e) represent the significant differences among the three samples at $p < 0.05$

expasy.org/protparam/) [34] was employed to estimate the physicochemical properties of each *C. teeta* MADS-box predicted protein, including the molecular weight (MW) and theoretical isoelectric point (pI).

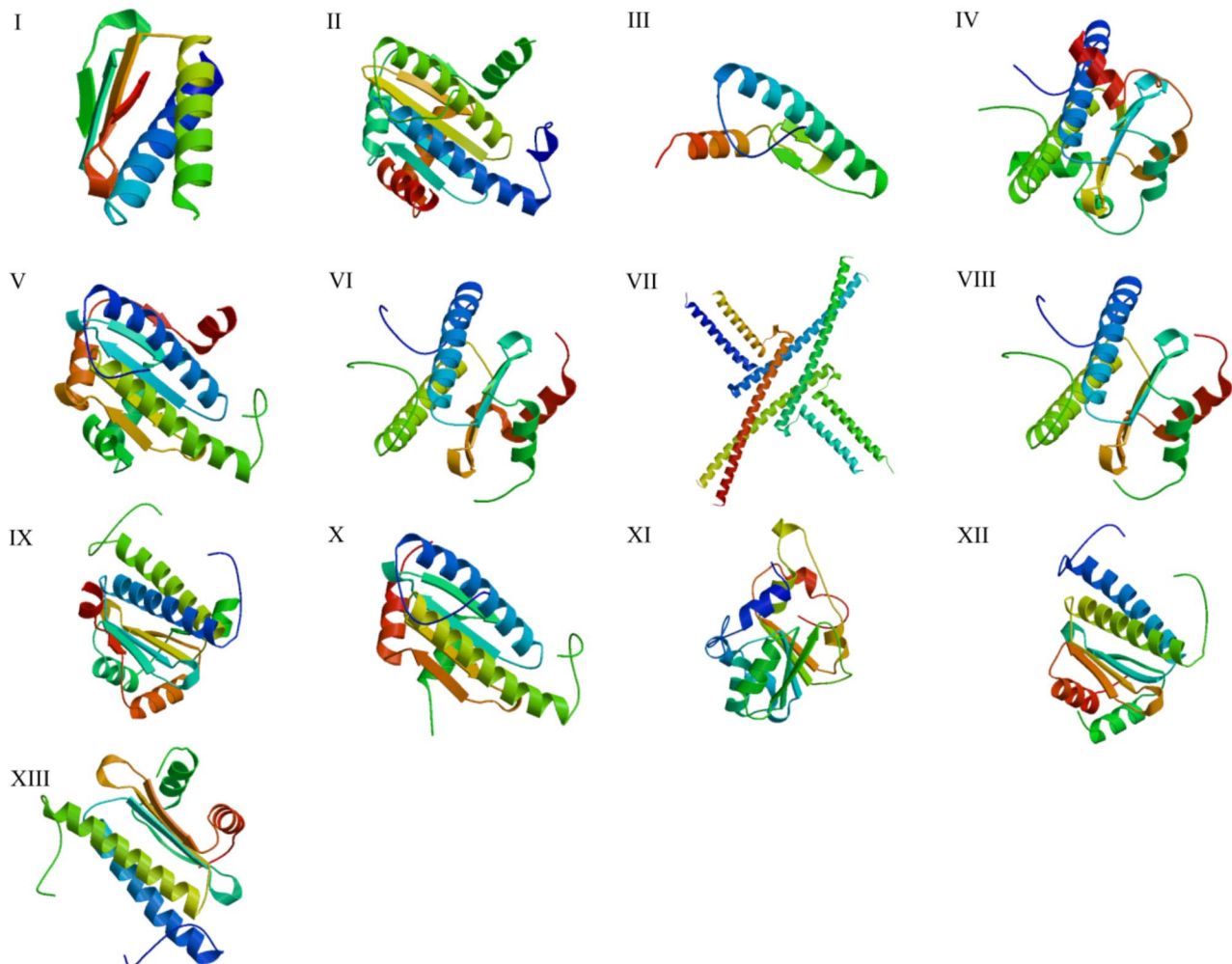
Phylogenetic analysis of MADS-box genes

The MADS-box genes of *Arabidopsis thaliana* were retrieved from the TAIR website (<https://www.arabidopsis.org/browse/genefamily/index.jsp>) [35]. By using the MADS-box protein sequences of *Arabidopsis thaliana* as

templates, the MADS-box protein sequences of *C. teeta* were identified using the multiple sequence alignment by employment of MAFFT v7.490 software. A phylogenetic tree was then constructed using the maximum likelihood (ML) method, using IQ-TREE v2.2.0-beta software, the model was JTT+F+R5, and the parameter was set to-B 2000-rlrt 2000 [36].

Table 3 Secondary structure and 3D structure simulation of *C. teeta* MADS-box protein

Protein Name	Subfamily	AH% (aa)	RC% (aa)	ES% (aa)	BT% (aa)	homologous template	Seq identity%
CteMADS18	M δ	35.31	49.85	11.28	3.56	7nb0.1	51.72
CteMADS24	M β	54.19	28.63	12.33	4.85	6byy.1	32.98
CteMADS21	Ma	40.56	31.33	22.49	5.62	7nb0.1	62.50
CteMADS44	My	54.35	34.78	6.96	3.91	1mnm.1	36.84
CteMADS50	Bs	57.68	27.80	8.71	8.71	7x1n.2	50
CteMADS51	AP3	57.66	18.92	15.77	7.66	6byy.2	44
CteMADS52	SEP	56.20	30.99	8.68	4.13	4ox0.2	73.53
CteMADS54	AG	52.92	31.91	12.06	3.11	6byy.2	51.32
CteMADS55	SOC1	56.84	32.05	8.12	2.99	6byy.2	43.02
CteMADS56	AGL6	53.85	34.82	7.69	3.64	6byy.2	52
CteMADS59	AGL15	38.04	32.58	20.53	8.85	2pxx.1	38.32
CteMADS64	SVP	57.96	26.11	12.39	12.39	6byy.1	48
CteMADS66	API	55.47	32.79	32.79	2.83	6byy.2	51.32

**Fig. 9** 3D structural analysis of *C. teeta* MADS-box predicted proteins

Simulated three-dimensional structures with ligand-binding regions (I–XIII). M δ : CteMADS18 (I), My: CteMADS24 (II), Ma: CteMADS21 (III), My: CteMADS44 (IV), Bs: CteMADS50 (V), AP3: CteMADS51 (VI), SEP: CteMADS52 (VIII), AG: CteMADS54 (VII), SOC1: CteMADS55 (IX), AGL6: CteMADS56 (X), AGL15: CteMADS59 (XI), SVP: CteMADS64 (XII), API: CteMADS66 (XIII)

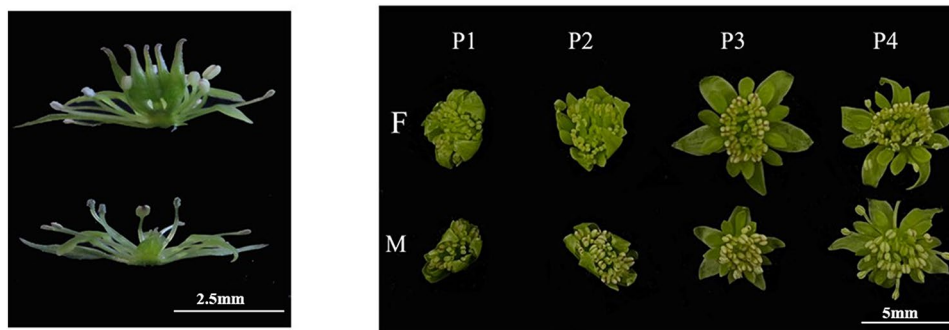


Fig. 10 Selected stages of flower development

The four selected stages of flower development, as shown in this figure, in plants exhibiting the M and F type floral phenotypes were: stage 1 (P1). The bracts were dehiscent and florets unexposed; stage 2 (P2): The florets were exposed, and sepals were unexpanded; stage 3 (P3): The sepals had fully expanded; stage 4 (P4): The anthers had dehiscent and had begun to disperse pollen

Conserved motif and gene structure analysis of MADS-box genes

The conserved motifs of the MADS-box proteins of *C. teeta* were estimated by using MEME (<http://meme.nbcr.net/meme/intro.html>) [37] online tools with the following parameters: maximum number of motifs=20. The number of introns and exons was analyzed and visualized by TBtools [33].

Chromosomal location and gene duplication analysis

The location of the individual MADS-box genes was predicted according to the genome annotation of *C. teeta*. To know the genome collinearity between *C. teeta* and other species (*Arabidopsis thaliana*, *Coptis chinensis* and *Aquilegia coerulea*), the genomic data of each species were downloaded from the NCBI database (<https://www.ncbi.nlm.nih.gov>). The collinearity, chromosome localization, the non-synonymous (Ka) and synonymous (Ks) values of *C. teeta* MADS-box genes were then estimated using TBtools software [33].

Expression analysis

Total RNA was extracted from fresh M and F type flowers according to the manufacturer of the RNA extraction kit (Magen, Guangzhou, China). The concentration of the extracted total RNA was measured using the Nano-Drop technique (Thermo Fisher Scientific, USA), the equivalent amount RNA was reverse transcribed into cDNA by using a reverse transcription kit (TAKARA, Beijing, China). Sixteen candidate genes were chosen to verify the RNA-seq data. The quantitative real-time PCR (RT-qPCR) was performed using an Applied Biosystems QuantStudio 5 system (Thermo Fisher Scientific, USA) with the ChamQ Universal SYBR qPCR Master Mix. The reaction procedure was carried out as follows: pre-denaturation 95°C for 30s, denaturation at 95°C for 30 s, annealing at 58°C for 30s, for 40 cycles, the fluorescence signal was then detected. Subsequently, an extra procedure was conducted as follows: denaturation at 95°C for

15s, annealing at 60°C for 1 min, extension at 72°C for 15s. The fluorescence signal was then detected, and the dissolution curve analyzed. The primer sequences of the candidate genes are listed in Table 4. The relative expression patterns of the target genes were calculated using the $2^{-\Delta\Delta CT}$ method and repeated three times [38]. The selected gene expression patterns were visualized using TBtools [33] (Supplementary Table S3).

Protein structure prediction

The MADS-box domain protein secondary structure (alpha helices, random coils, beta turns and extended strands) was predicted using the self-optimized prediction method with alignment (SOPMA) [39]. The MADS-box domain protein tertiary structure modelling prediction was performed using the SWISS-MODEL [40].

Discussion

MADS-box genes have previously been shown to play an important role in the initiation of floral organ formation and floral morphogenesis in a number of plant species. In light of which, in order to investigate the potential function of MADS-box genes in the formation of herkogamy and dichogamy in developing flowers of *C. teeta*, a genome-wide exploration and identification of the MADS-box genes present in this species was performed.

A total of 66 MADS-box genes were identified in *C. teeta* genome compared to the 0.107 found present in the *Arabidopsis thaliana* genome [41]. In other widely studied species in which MADS-box genes have been identified 101 have been shown to be present in *Litchi chinensis*, 78 in *Callicarpa americana*, 75 in *Oryza sativa* [7], 72 in *Panicum italicum* [42], 69 in *Theobroma cacao* [27] and 45 in *Bletilla striata*, respectively [20]. The unequal distribution of MADS-box genes between different plant species might be due to gene segmental

Table 4 Primer information of selected MADS-box genes in verification of gene expression

Gene	Primer sequence (5'to3')	Anneal temperature (°C)	GC Content (%)
<i>CteMADS30</i>	5'F: TTTGTGTGACGTGGAAGCAT	60.16	45
	3'R: TTCCGATCCAACGAATTAGC	60.04	50
<i>CteMADS26</i>	5'F: CAGCCAAAATGAGGGAGCAA	58.74	50
	3'R: GCCTGAACCACCATAGTCCT	59.09	55
<i>CteMADS21</i>	5'F: TTGGTCAATGCTTCCATGCC	59.10	50
	3'R: AACCACGTCCGAAACCAAAG	59.98	50
<i>CteMADS56</i>	5'F: AGGCTCATGGGAATCAACTG	60.07	50
	3'R: TTAGTCTCCCCAGCAGCACT	60.01	55
<i>CteMADS45</i>	5'F: TGTTTCGAGACGTTGGAGA	58.97	50
	3'R: CCTCTGATGATGACGACGA	58.97	55
<i>CteMADS10</i>	5'F: CATGTGGAAGGACAACATGC	59.97	50
	3'R: CCATCGGGCCTATACTGAAA	59.92	50
<i>CteMADS39</i>	5'F: AAGTCATGGTTGGACGGTTC	59.83	50
	3'R: CCACTTCTTCATGCAGTCCA	59.83	50
<i>CteMADS01</i>	5'F: TCATCATCGTCGAGTTTTGC	59.81	45
	3'R: TTGCTGGCAATCTCATCAAG	59.95	45
<i>CteMADS02</i>	5'F: AGATCGTCGTTCTTGCGTTT	59.88	45
	3'R: TTCTCCCACCAGAACTCAC	60.09	55
<i>CteMADS66</i>	5'F: ACCAACTCGTCCATGAGTCC	59.97	55
	3'R: GGGCCACCAATGTTTAGAGA	59.93	50
<i>CteMADS44</i>	5'F: GGGTAATGCGGGAATGTATG	60.04	50
	3'R: CCTCTTACCACCACAGTT	60.00	55
<i>CteMADS19</i>	5'F: AGGTCAAGCTTGATGGATT	59.70	45
	3'R: GAGATGGCCAAACATCTGGT	59.93	50
<i>CteMADS11</i>	5'F: TCATCGTGTGGTCATGAGGT	59.96	50
	3'R: ATGCCACAGCATTAGTCC	59.96	50
<i>CteMADS62</i>	5'F: AGGCTCATGGGAATCAACTG	60.07	50
	3'R: TTAGTCTCCCCAGCAGCACT	60.01	55
<i>CteMADS24</i>	5'F: TGAGGGAAAGGGTTCATCAG	60.04	50
	3'R: CGTAGCACTCTGCCATCGTA	60.03	55
<i>CteMADS25</i>	5'F: TGAGGGAAAGGGTTCATCAG	60.04	50
	3'R: CGTAGCACTCTGCCATCGTA	60.03	55
<i>EF1a</i>	5'F: CTCCAGGCCATCGTATTTC	60.13	55
	3'R: AGTGAAGGCAAGTAGAGCGT	59.97	50

duplication, tandem duplication, or whole genome duplication [43].

Forty-nine of the *C. teeta* MADS-box genes were categorized as type I genes, including M α subfamily (27 genes), M β subfamily (10 gene), M γ subfamily (10 genes), and M δ subfamily (2 genes). Seventeen *C. teeta* MADS-box genes were categorized as type II genes, including *SOC1* (2), *SVP* (1), *AGL15* (1), *SEP* (3), *AGL6* (3), *API/FUL* (2), *BS/TT16* (1), *AP3/PI* (3), *AG* (1).

The fact that the number of type I genes in *C. teeta* was found to be larger than the number of type II genes is similar to the distribution observed in *Theobroma cacao*, *L. chinensis* and *Rhododendron hainanense* (Supplementary Table S2). These results suggest that the genome evolution of these species may have led to the different distribution patterns of MADS-box genes observed in different species.

Type II MADS-box genes (*SEP*) have been shown to be necessary for the specification of petals, stamens, and carpels in *Arabidopsis thaliana* [44]. *AP3* and *PI* are also involved in petal and stamen development in *Arabidopsis thaliana* [45]. The *AGL6-like* gene *OsMADS6* and the *SVP-like* gene *OsMADS22* have been reported to be differentially expressed in flower organs such as stamen and carpel in rice, during the flowering stage [41] and the gene *SOC1* has been shown to effect flower initiation and to facilitate adaptation to different latitudes in soybean [46]. The homologs of these genes were identified in *C. teeta* and may fulfill similar roles in this species.

The results of the phylogenetic tree analysis showed that there were great differences in the number of MADS-box genes among the selected species which shows that replication events or the expansion and diversification of subfamilies might have occurred during the process of evolution.

MADS-box gene structure and gene duplication events in *C. teeta*

Previous studies found that the intron-exon organization of MADS-box genes from different plant species varied. Thus, the exon-intron organization of the identified *C. teeta* MADS-box genes were analyzed to reveal their structural diversity. Major differences in gene structure were observed between the type I and type II MADS-box genes. The type I genes were shown to either lack introns or only possess a single intron. Conversely, the type II genes were shown to contain multiple introns, with an average number of 8 per gene. This exon-intron arrangement of the type I and II MADS genes also observed in other plant taxa, such as *Vitis vinifera* L [47], and *Lactuca sativa* L [48]. This suggests that the MADS-box transcription factors are highly conserved in plants.

Conserved motif analysis revealed that the same sub-family contained most conserved motifs (Fig. 5), which suggests that these conserved motifs play important roles in family-specific functions.

The location analysis showed that the *C. teeta* MADS-box genes were unevenly distributed throughout the genome similar to data reported for *Medicago sativa* L [5, 23]. It is well known that gene duplication events play an important role in the expansion and evolution of gene families [49]. Duplications of MADS-box genes have been found in many species, including *Medicago sativa* [5] and *Solanum lycopersicum* [50]. Four pairs of tandemly duplicated genes and four pairs of segmentally duplicated MADS-box genes were found in *C. teeta*.

The Ka/Ks ratios of all gene pairs of *C. teeta* were less than 1, indicating that this MADS-box gene family had undergone purification selection pressure, similar to the MADS-box gene family in *Solanum tuberosum* [51].

Potential function of MADS-box genes on the development of floral dichogamy and herkogamy in *C. teeta*

The results of the current study demonstrated that some of the MADS-box genes identified in *C. teeta*, displayed differential patterns of expression within the two flower types. *CteMADS24(AGL90)*, *CteMADS25(AGL90)*, *CteMADS45(AGL16)* and *CteMADS30(AGL80)* were more highly expressed in the M-type than in the F-type plants at full anthesis.

Previous studies have shown that *AGL16* regulates flowering by interacting with other MADS-box genes. It targets approximately seventy flowering related genes involved in multiple developmental pathways [5]. In our study we found that *CteMADS26(AGL80)*, *CteMADS11(AGL6)*, *CteMADS44(AGL80)* showed higher expression in the M-type plants at the initial stage of flower development, and expressed highly in the F-type plants in mature flowers.

Two of the identified genes were highly expressed only at the P4 (mature) stage of flower development. *CteMADS19(AGL80)* only displayed high levels of expression in mature flowers of the F-type plant and *CteMADS21(AGL62)* was only expressed at high levels at the P4 stage of flower development in the M-type plants.

AGL62 is a key transcription factor required for activation of auxin synthesis [52]. Auxin plays an essential role in regulating plant reproduction. Auxin deficient mutants are often associated with impaired reproductive traits such as anther filament elongation [53]. In *A. thaliana*, expression of a fusion of the *AGL6* genomic sequence (*gAGL6::VP16*) led to the formation of ectopic organs. As the plants grew these ectopic organs became increasingly carpel and stamen-like in later developing flowers, and gradually the flowers were completely replaced by staminode bracts, suggesting an important role for *AGL6* in floral organ identity [54]. It has also been demonstrated that *TaAGL6* regulates stamen development in *Triticum aestivum* [31].

To summarise, the majority of the differentially expressed MADS-box genes between the M and F type plants, belong to *AGL* subfamily. The functions of the *AGL* subfamily in flower development and floral organogenesis are well established in the literature [55]. Therefore, it is logical to hypothesize that the differentially expressed genes (*CteMADS24*, *CteMADS25*, *CteMADS21*, *CteMADS45* and *CteMADS56*) may play important roles in the formation of herkogamy and dichogamy in *C. teeta*.

The 3D structure prediction of CteMADS proteins suggested that the MADS-box gene family in *C. teeta* is structurally conserved. The results are similar to earlier reports on *Malus sylvestris* and *Vanilla planifolia* [56, 57].

Conclusions

In this study, a total of 66 MADS-box genes were identified to be present in the *C. teeta* genome, which were subsequently divided into type I and type II. The physicochemical properties, phylogenetic relationships, conserved motifs, chromosomal location, gene duplication, exon-intron structure, and expression were analyzed. These data provide a strong foundation for future studies of the role of MADS-box gene expression in this species, particularly in relation to floral development.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05714-0>.

Supplementary Material 1

Author contributions

YLL, YCL and XCM directed the whole process of the experiment. They also made suggestions for the writing of the manuscript. SFD participated in the whole experiment, analyzed the relevant experimental data, and wrote the paper. JCY, YY, GSX, RC, YZ and Timothy Charles Baldwin provided help with the experiments and data analysis.

Funding

This research was supported by the National Natural Science Foundation of China (project grant no: 31971543), Major Special Projects of Yunnan Province (Biomedicine) (project numbers: 202102AA310045), Yunnan Provincial Science and Technology Department Agriculture Joint Special Project (project numbers: 202101BD070001-008).

Data availability

All data supporting the findings were contained in the manuscript and its supplementary files except the RNA-seq raw data. Data supporting the findings of this research are available from the corresponding author upon reasonable request. And all original transcriptome data were submitted to NCBI (Accession Number: PRJNA973134).

Declarations

Ethics approval and consent to participate

All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Agronomy & Biotechnology, Yunnan Agricultural University, Kunming, Yunnan 650201, China

²Key Laboratory of Medicinal Plant Biology of Yunnan Province, Yunnan Agricultural University, Kunming, Yunnan 650201, China

³National & Local Joint Engineering Research Center on Germplasm Innovation & Utilization of Chinese Medicinal Materials in Southwestern China, Yunnan Agricultural University, Kunming, Yunnan 650201, China

⁴Yunnan Land and Resources Vocational College, Kunming, Yunnan 650201, China

⁵Department of Applied Technology, Lijiang Normal University, Lijiang, Yunnan 674100, China

⁶Faculty of Science and Engineering, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK

⁷Yunnan Agricultural University College of Education and Vocational Education, Yunnan Agricultural University, Kunming, Yunnan 650201, China

⁸Fengyuan Road, Panlong District, Kunming 650201, China

Received: 18 August 2023 / Accepted: 15 October 2024

Published online: 29 October 2024

References

- Riechmann JL, Ratcliffe OJ. A genomic perspective on plant transcription factors. *Curr Opin Plant Biol.* 2000;3(5):423–34.
- Singh KB, Foley RC, Oñate-Sánchez L. Transcription factors in plant defense and stress responses. *Curr Opin Plant Biol.* 2002;5(5):430–6.
- Messenguy F, Dubois E. Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene.* 2003;316:1–21.
- Gahlaut V, Jaiswal V, Kumar A, Gupta PK. Transcription factors involved in drought tolerance and their possible role in developing drought tolerant cultivars with emphasis on wheat (*Triticum aestivum* L.). *Theor Appl Genet.* 2016;129:2019–42.
- Dong X, Deng H, Ma W, Zhou Q, Liu Z. Genome-wide identification of the MADS-box transcription factor family in autotetraploid cultivated alfalfa (*Medicago sativa* L.) and expression analysis under abiotic stress. *BMC Genomics.* 2021;22(1):1–16.
- Becker A, Winter K-U, Meyer B, Saedler H, Theißen G. MADS-box gene diversity in seed plants 300 million years ago. *Mol Biol Evol.* 2000;17(10):1425–34.
- Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics.* 2007;8(1):1–21.
- Hepworth SR, Valverde F, Ravenscroft D, Mouradov A, Coupland G. Antagonistic regulation of flowering-time gene *SOC1* by *CONSTANS* and *FLC* via separate promoter motifs. *EMBO J.* 2002;21(16):4327–37.
- Wang L, Fan S, Song M. Advances in the research of MADS-box gene in plant. *Biotechnol Bull.* 2010;8:12–9.
- Adamczyk BJ, Fernandez DE. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. *Plant Physiol.* 2009;149(4):1713–23.
- Li C, Wang Y, Xu L, Nie S, Chen Y, Liang D, et al. Genome-wide characterization of the MADS-box gene family in radish (*Raphanus sativus* L.) and assessment of its roles in flowering and floral organogenesis. *Front Plant Sci.* 2016;7:1390.
- Gramzow L, Theissen G. A hitchhiker's guide to the MADS world of plants. *Genome Biol.* 2010;11:1–11.
- De Bodt S, Raes J, Van de Peer Y, Theißen G. And then there were many: MADS goes genomic. *Trends Plant Sci.* 2003;8(10):475–83.
- Norman C, Runswick M, Pollock R, Treisman R. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell.* 1988;55(6):989–1003.
- Smaczniak C, Immink RG, Angenent GC, Kaufmann K. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development.* 2012;139(17):3081–98.
- Becker A, Theißen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol.* 2003;29(3):464–89.
- Heijmans K, Morel P, Vandenbussche M. MADS-box genes and floral development: the dark side. *J Exp Bot.* 2012;63(15):5397–404.
- Lloyd DG, Webb C. The avoidance of interference between the presentation of pollen and stigmas in angiosperms I. Dichogamy. *N Z J Bot.* 1986;24(1):135–62.
- Brunet J, Eckert C. Effects of floral morphology and display on outcrossing in blue columbine, *Aquilegia caerulea* (Ranunculaceae). *Funct Ecol.* 1998;12(4):596–606.
- Motten AF, Stone JL. Heritability of stigma position and the effect of stigma-anther separation on outcrossing in a predominantly self-fertilizing weed, *Datura stramonium* (Solanaceae). *Am J Bot.* 2000;87(3):339–47.
- Takebayashi N, Wolf D, Delph L. Effect of variation in herkogamy on outcrossing within a population of *Gilia achilleifolia*. *Heredity.* 2006;96(2):159–65.
- Liu T-h, Zhang X-m, Tian S-z, Chen L-g, Yuan J-h. Bioinformatics analysis of endophytic bacteria related to berberine in the Chinese medicinal plant *Coptis teeta* wall. *3 Biotech.* 2020;10:96.
- Xu Z, Zhang Q, Sun L, Du D, Cheng T, Pan H, et al. Genome-wide identification, characterisation and expression analysis of the MADS-box gene family in *Prunus mume*. *Mol Genet Genomics.* 2014;289:903–20.
- Ye L-X, Zhang J-X, Hou X-J, Qiu M-Q, Wang W-F, Zhang J-X, et al. A MADS-box gene *CiMADS43* is involved in citrus flowering and leaf development through interaction with *CiAGL9*. *Int J Mol Sci.* 2021;22(10):5205.
- Mi Z-Y, Zhao Q, Lu C, Zhang Q, Li L, Liu S, et al. Genome-wide analysis and the expression pattern of the MADS-box gene family in *Bletilla striata*. *Plants.* 2021;10(10):2184.
- Won SY, Jung J-A, Kim JS. Genome-wide analysis of the MADS-Box gene family in *Chrysanthemum*. *Comput Biol Chem.* 2021;90:107424.
- Zhang Q, Hou S, Sun Z, Chen J, Meng J, Liang D, et al. Genome-wide identification and analysis of the MADS-Box gene family in *Theobroma cacao*. *Genes.* 2021;12(11):1799.
- Li H, Liang W, Jia R, Yin C, Zong J, Kong H, et al. The *AGL6*-like gene *OsMADS6* regulates floral organ and meristem identities in rice. *Cell Res.* 2010;20(3):299–313.
- Kong X, Wang F, Geng S, Guan J, Tao S, Jia M, et al. The wheat *AGL6*-like MADS-box gene is a master regulator for floral organ identity and a target for spikelet meristem development manipulation. *Plant Biotechnol J.* 2022;20(1):75–88.
- Wang B-G, Zhang Q, Wang L-G, Duan K, Pan A-H, Tang X-M, et al. The *AGL6*-like gene *CpAGL6*, a potential regulator of floral time and organ identity in wintersweet (*Chimonanthus praecox*). *J Plant Growth Regul.* 2011;30:343–52.

31. Su Y, Liu J, Liang W, Dou Y, Fu R, Li W, et al. Wheat AGAMOUS LIKE 6 transcription factors function in stamen development by regulating the expression of *Ta APETALA3*. *Development*. 2019;146(20):dev177527.
32. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A. The pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427–32.
33. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020;13(8):1194–202.
34. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. Springer; 2005.
35. Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res*. 2003;31(1):224–8.
36. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res*. 2019;47(W1):W5–10.
37. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37(suppl2):W202–8.
38. Livak KJ. TD Schmittgen 2001 Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25 4 402–8.
39. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*. 1995;11(6):681–4.
40. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296–303.
41. Pelucchi N, Fornara F, Favalli C, Masiero S, Lago C, Pè EM, et al. Comparative analysis of rice MADS-box genes expressed during flower development. *Sex Plant Reprod*. 2002;15:113–22.
42. Zhao W, Zhang L-L, Xu Z-S, Fu L, Pang H-X, Ma Y-Z, et al. Genome-wide analysis of MADS-Box genes in foxtail millet (*Setaria italica* L.) and functional assessment of the role of *SiMADS51* in the drought stress response. *Front Plant Sci*. 2021;12:659474.
43. Veron AS, Kaufmann K, Bornberg-Bauer E. Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. *Mol Biol Evol*. 2007;24(3):670–8.
44. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*. 2000;405(6783):200–3.
45. Bowman JL, Smyth DR, Meyerowitz EM. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development*. 1991;112(1):1–20.
46. Kou K, Yang H, Li H, Fang C, Chen L, Yue L, et al. A functionally divergent *SOC1* homolog improves soybean yield and latitudinal adaptation. *Curr Biol*. 2022;32(8):1728–42. e1726.
47. Grimplet J, Martínez-Zapater JM, Carmona MJ. Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genomics*. 2016;17(1):1–23.
48. Ning K, Han Y, Chen Z, Luo C, Wang S, Zhang W, et al. Genome-wide analysis of MADS-box family genes during flower development in lettuce. *Plant Cell Environ*. 2019;42(6):1868–81.
49. Wang R, Ming M, Li J, Shi D, Qiao X, Li L, et al. Genome-wide identification of the MADS-box transcription factor family in pear (*Pyrus bretschneideri*) reveals evolution and functional divergence. *PeerJ*. 2017;5:e3776.
50. Wang Y, Zhang J, Hu Z, Guo X, Tian S, Chen G. Genome-wide analysis of the MADS-box transcription factor family in *Solanum lycopersicum*. *Int J Mol Sci*. 2019;20(12):2961.
51. Vatanever R, Koc I, Ozyigit II, Sen U, Uras ME, Anjum NA, et al. Genome-wide identification and expression analysis of sulfate transporter (SULTR) genes in potato (*Solanum tuberosum* L.). *Planta*. 2016;244:1167–83.
52. Guo L, Luo X, Li M, Joldersma D, Plunkert M, Liu Z. Mechanism of fertilization-induced auxin synthesis in the endosperm for seed and fruit development. *Nat Commun*. 2022;13(1):3985.
53. Wu C, Paciorek M, Liu K, LeClere S, Perez-Jones A, Westra P, et al. Investigating the presence of compensatory evolution in dicamba resistant IAA16 mutated kochia (*Bassia scoparia*). *Pest Manag Sci*. 2021;77(4):1775–85.
54. Koo SC, Bracko O, Park MS, Schwab R, Chun HJ, Park KM, et al. Control of lateral organ development and flowering time by the *Arabidopsis thaliana* MADS-box gene *AGAMOUS-LIKE6*. *Plant J*. 2010;62(5):807–16.
55. Dreni L, Zhang D. Flower development: the evolutionary history and functions of the AGL6 subfamily MADS-box genes. *J Exp Bot*. 2016;67(6):1625–38.
56. Tian Y, Dong Q, Ji Z, Chi F, Cong P, Zhou Z. Genome-wide identification and analysis of the MADS-box gene family in apple. *Gene*. 2015;555(2):277–90.
57. Himani, Sharma A, Ramkumar TR, Sembi JK. Regulatory mechanisms underlying florigenesis in *Vanilla planifolia* Andrews: a study of MADS-box gene family. *J Horticult Sci Biotechnol*. 2021;96(4):428–43.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.