Plant Biotechnology 35, 313–324 (2018)
DOI: 10.5511/plantbiotechnology.18.0621a

Original Paper

Transcriptome-wide analysis of MADS-box family genes involved in aluminum and fluoride assimilation in *Camellia sinensis*

Junting Pan, Pinpin Chang, Xiaoli Ye, Jiaojiao Zhu, Dongqin Li, Chuanlei Cui, Bo Wen, Yuanchun Ma, Xujun Zhu, Wanping Fang, Yuhua Wang*

College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China
* E-mail: wangyuhua@njau.edu.cn Tel & Fax: +86-25-84395182

Received March 9, 2018; accepted June 21, 2018 (Edited by Y. Nagano)

Abstract MADS-box transcription factors (TFs) are involved in a variety of processes in flowering plants ranging from root growth to flower and fruit development. However, studies of the tolerance-related functions of MADS-box genes are very limited, and to date no such studies have been conducted on *Camellia sinensis*. To gain insight into the functions of genes of this family and to elucidate the role they may play in tissue development and Al and F response, we identified 45 MADS-box genes through transcriptomic analysis of *C. sinensis*. Phylogenetic analysis of these CsMADS-box genes, along with their homologues in *Arabidopsis thaliana*, enabled us to classify them into distinct groups, including: M-type (Mα), MIKC* and MIKC^c (which contains the SOC1, AGL12, AGL32, SEP, ANR1, SVP, and FLC subgroups). Conserved motif analysis of the CsMADS-box proteins revealed diverse motif compositions indicating a complex evolutionary relationship. Finally, we examined the expression patterns of CsMADS-box genes in various tissues and under different Al and F concentration treatments. Our qPCR results showed that these CsMADS-box genes were involved in Al and F accumulation and root growth in *C. sinensis*. These findings lay the foundation for future research on the function of CsMADS-box genes and their role in response to Al and F accumulation in root tissues.

Key words: aluminum, *Camellia sinensis*, fluoride, MADS-box, root growth.

Introduction

Tea [*Camellia sinensis* (L.) O. Kuntze], as one of the most popular beverages in the world, consumption its thought to be beneficial to human health (Forester and Lambert 2011). However, the tea plant is known to be an aluminum (Al) and fluoride (F) hyperaccumulator (Wong et al. 2003), and tea leaves can accumulate higher the levels of F and Al than is found in other edible plants (Xie et al. 2001). Tea plants accumulate more than 90% of this F and Al in leaf tissues, especially in old leaves (Simpson et al. 2001). F levels in common tea products range from 200 to 400 mg kg^{-1}, which is considered to be safe for human health (Fung et al. 1999). While some brick tea, processed from older tea leaves, contain much higher levels of F and consumption of brick tea may pose a risk of fluorosis (Wong et al. 2003). In addition, there is a general consensus that overconsumption of Al is harmful to health, which has been linked to kidney weakness and other diseases (Gauthier et al. 2000). Thus, the effect of intake of F and Al by drinking tea is a health question of considerable importance (Cao et al. 2003; Erdemoglu et al. 2000).

Tea plants are absorbed F and Al from soil solutions via roots. However, unlike other plants, tea plants can accumulate high levels of F^- from soils without showing symptoms of toxicity (Ruan and Wong 2001). Previous study has showed that F^- is readily assimilated by tea roots and that most of this is transported to leaves (Ruan et al. 2004). The reasons for high F accumulation in tea plant are not clear, although a recent work demonstrated that F is taken up by tea roots through active rather than passive processes (Zhang et al. 2013). *C. sinensis* shows a similar tolerance for Al: in one study, on a dry-weight (DW) basis, tea dry matter production was significantly higher in the presence of Al and resulted in the accumulation of up to 6.18 and 1.21 mg Al/g DW in roots and leaves, respectively (Hajiboland and Poschenrieder 2015). Tea leaves did not exhibit Al toxicity symptoms high Al concentrations. Furthermore, Li et al. (2011) reported that, in *C. sinensis*, Al^{3+} was able to enhance leaf photosynthetic activity and antioxidant defence, and induced reactive oxygen species, resulting in delayed lignification and enhanced membrane integrity. Further

This article can be found at http://www.jspcmb.jp/
Published online October 25, 2018

study showed that Al^{3+} stimulated the tea roots growth through enhancing nutrient uptake (Fung et al. 2008). Notably, Al and F can form a certain number of stable complexes in solution. AlF_x complexes such as AlF^{2+}, AlF_2^+, AlF_4^- and AlF_3 conduce to total soluble Al and F levels in acidic soil solutions with pH values of <5.5 (Wenzel and Blum 1992). One study showed that Al can form both F complexes and catechin in integrated tea leaves, and the forms of F and Al in tea leaves are consistent with those in soil solutions (Nagata et al. 1993). Moreover, a number of studies have identified the physiological mechanisms of Al uptake that alleviate F toxicity in tea plants (Yang et al. 2016; Zhang et al. 2015b). However, the detailed of molecular mechanism responsible for the hyperaccumulation of Al and F in tea plants remains unknown.

Plant MADS-box genes encode a transcription factors (TFs) family and are known best for their roles in floral organogenesis. MADS-box genes are distinguished by existing a highly-conserved DNA-binding motif of approximately 58 amino acids (*CArG* boxes) (Messenguy and Dubois 2003). Based on phylogenetic analysis, plant MADS-box genes can be subdivided into two lineages, M-type (type I) and MIKC (type II) (Alvarez-Buylla et al. 2000b). M-type genes are further grouped into Mα, Mβ, and Mγ subgroups based on sequence similarity between the MADS-box regions (Parenicova et al. 2003). MIKC genes have been designated MIKC due to the presence of four distinct domains: MADS, Intervening (I), Keratin-like (K), and C-terminal (C) (Cho et al. 1999). The I domain (ca. 30 aa) contributes to the formation of specific DNA-binding dimers, the highly conserved K domain (ca. 70 aa) intercedes dimerization, and the C domain plays a role in transcriptional activation and the formation of higher order protein complexes. Based on sequence divergence in the I domain, MIKC-type genes can be further divided into $MIKC^c$ and MIKC* subgroups; genes of the $MIKC^c$ have contain a shorter I domains and more conserved K domain than MIKC*-type genes (Becker and Theissen 2003). $MIKC^c$-type genes, as the best-studied in plant MADS-box genes, can be further classified into 14 subfamilies according to sequence similarity (Chen et al. 2017).

Plant MADS-box genes are involved in regulating numerous developmental processes, and were initially identified as regulators of flower development. However, later work has shown that they control many major aspects of land plant life histories, including meristem identity, endodormancy transitions, formation of the dehiscence zone, vernalization, fruit ripening, and the growth of vegetative organs such as leaves and roots (Messenguy and Dubois 2003; Saedler et al. 2001). Indeed, in *Arabidopsis*, there is strong evidence that MADS-box genes are expressed in the root (Dolan et al. 1993); moreover, other studies have demonstrated that the MADS-box genes *AGAMOUS-LIKE 21* (*AGL21*), *AGL17*, and *AGL12* play distinct modulator roles during *Arabidopsis* root growth and development (Burgeff et al. 2002). Gan et al. (2005) identified the MADS-box genes (*ANR1*, *SOC1*, *AGL14*, *AGL16*, *AGL19*, *AGL21*, *AGL26*, and *AGL56*) participated in the nutritional mediation of lateral root growth in *Arabidopsis*. However, to date there is no evidence that, in *C. sinensis*, MADS-box genes are implicated in root growth in response to Al and F accumulation.

In this study, we identified 45 MADS-box TFs using RNA-seq data from the 'Longjingchangye' tea plant. We performed a bioinformatics analysis of MADS-box genes found in *C. sinensis*, including phylogenetic comparisons with *MADS-box* genes in *Arabidopsis*, as well as structural analyses of protein motifs. Moreover, we investigated the expression of 14 CsMADS-box genes in different tissues and in response to treatments simulating the accumulation of Al and F. Our results provide a molecular basis for future functional identification and comparative analysis of CsMADS-box family genes.

Materials and methods

Plant material and Al and F treatments

Two-year-old tea seedlings [*C. sinensis* (L.) O. Kuntze cv. Longjingchangye] were pre-incubated in a climate chamber under a 16-h photoperiod regimen with day/night temperatures of $25\pm1/23\pm1$°C with a standardized nutrient solution (Wan et al. 2012) for 2 weeks. Then these pre-incubated tea seedlings were cultured in nutrient solutions containing a gradient of different concentrations of Al^{3+} (0, 0.4, 2 mmol/l) and F^- (0, 8, 16 mg/l). The concentrations of Al and F were selected based on the results of a preliminary experiment (data not shown). The roots of control and treated tea seedlings were collected at various intervals after treatment (0, 3, 6, 12 and 24 h). In addition, the roots, stems, leaves (including first leaf and fourth leaf), buds, flowers, fruits, pollen and pollen tube from 'Longjingchangye' tea plants for qRT-PCR analysis of CsMADS-box genes were separately harvested, washed with deionized water, weighed, and collected. All these samples were frozen in liquid nitrogen, and then stored at −80°C for RNA extraction and cDNA reverse transcription.

Database search and identification of MADS-box TFs

Putative MADS-box proteins were retrieved both from published *C. sinensis* pollen tube transcriptome data (Pan et al. 2016) and root transcriptome data (accession: SRP149468). The de novo assembly and gene annotation methods of RNA-Seq followed our previous report (Pan et al. 2016). *Arabidopsis* MADS-box protein sequences were downloaded from the Database of *Arabidopsis* TFs (DATF) (Guo et al. 2005). The *Arabidopsis* MADS-box domain sequences were used as query sequences to evaluate CsMADS-box proteins with a complete

MADS-box domain. Several quality-control measures followed the identification of putative MADS-box genes. First, all genes were rechecked to avoid duplication. Secondly, short CsMADS-box sequences with an incomplete MADS-box domain were removed. Thirdly, those genes with complete MADS-box domains were assessed by BLASTing the predicted MADS-box sequences against the DATF. Finally, multiple alignments among those CsMADS-box sequences were performed to avoid repetition.

Phylogenetic and conserved motif analysis of the CsMADS-box proteins

To classify CsMADS-box proteins and determine their phylogenetic relationships, 36 *Arabidopsis* MADS-box genes were compared to the 45 identified CsMADS-box genes. A phylogenetic tree was built with MEGA 6.0 by using the neighbor-joining method (Tamura et al. 2013). Conserved motifs in the CsMADS-box factors were identified using the motif finding program MEME (Multiple Em for Motif Elucidation) 4.11.2 (http://meme-suite.org/tools/meme). The motif was searched for using CsMADS-box protein sequencing data according to the following parameters: (1) the critical E value for motifs was set to $1.0\,e^{-10}$; (2) the width of optimum motif length was set to $\geq$10 and $\leq$200; (3) a single motif was distributed among the sequences basing on the pattern: zero or one per sequence (-mod zoops).

qRT-PCR analysis and statistical analyses

Specific primers for qRT-PCR were designed by using Primer Premier 5.0. The *β-actin* gene of the tea plant was selected as an internal reference. The primers of selected CsMADS-box genes are listed in Supplementary Table S1. Cycling profiles were based on the recommended profile for the SYBR Premix Ex-Taq kit® (TaKaRa, Dalian, China). Each reaction contained a 20 μl reaction volume, which contained: 0.4 μl (10 mM) of each primer, 7.2 μl of ddH_2O, 2 μl of the cDNA template, and 10 μl of premixed SYBR mixture. qRT-PCR was performed using a qRT-PCR Bio-Rad CFX96 platform according to the following cycling protocol: denaturation at 95°C for 30 s; 40 cycles of 95°C for 5 s and 60°C for 30 s; then 95°C for 10 s; followed by 65°C for 5 s and 95°C for 5 s. Three technical replicates for each gene were included for each independent RNA sample of each cultivar at each temperature treatment. Expression levels of MADS-box genes relative to those of the *β-actin* gene were calculated using the $2^{-\Delta\Delta CT}$ method (Pfaffl 2001). The expression level was calculated as the mean of three $2^{-\Delta\Delta CT}$ values obtained from the three replicates. All data analyses were performed with IBM *SPSS Statistics* Version 20.

Results

Identification and classification of the MADS-box genes in C. sinensis

In total, 100 MADS-box sequences encoding putative MADS-box motifs were identified from published *C. sinensis* pollen tube transcriptome data (Pan et al. 2016) and root transcriptome data. Candidate amino acid sequences with an incomplete CsMADS-box domain were removed by querying the NCBI BLAST program. To avoid repeatedly including the same genes, we screened all predicted CsMADS-box sequences with complete MADS-box domains by BLASTing them against the DATF, and by performing multiple alignments among sequences. 45 CsMADS-box genes, designated as *CsMADS1-CsMADS45*, were eventually confirmed as candidates for further analysis (Supplementary Table S2).

Phylogenetic analysis of CsMADS-box genes

To date, most MADS-box genes have been studied in *Arabidopsis*. To clarify the phylogenetic relationship between MADS-box family proteins found in tea plant and *Arabidopsis*, MEGA software was used to create a neighbor-joining phylogenetic tree in order to analyze MADS-box proteins in *C. sinensis* and *Arabidopsis*. There were 4 M-type members (Mα) from *C. sinensis*. The other 41 MADS-box proteins were classified as belonging to the MIKC-type (MIKC^c and MIKC*) subgroup (Figure 1), and of these, 11 MADS-box proteins belonged to the MIKC* subgroup. Notably, According to the MIKC^c subgroup classes of MADS-box proteins in *Arabidopsis*, we identified 7 MIKC^c clades in *C. sinensis*. Among the 7 MIKC^c clades, the SOC1 clade included the most CsMADS-box sequences (10). The AGL12 clade only contained one CsMADS-box sequence.

Conserved motifs analysis in C. sinensis MADS-box proteins

The conserved motifs shared among CsMADS-box proteins were predicted using the MEME program. 18 motifs with E values less than $1.7\,e^{-15}$ were identified from MADS-box protein structures found in *C. sinensis* (Figure 2). The one characteristic domain of MADS-box genes, the MADS-box domain itself (motif 1 and 3), was found in all CsMADS-box genes. And we identified the accurate MADS-box domain by the presence of approximately 58 amino acids (*CArG* boxes) (Figure 3). The K domain, specified by conserved motifs 2, 5, and 7 (Figure 4), were found in diverse combinations in most MIKC^c proteins. However, the CsMIKC* proteins were identified to contain motif 2 of the K-domain less frequently than in the MIKC^c subgroup. Conserved motif 8 specified the C domain (Figure 4). Moreover, in the MIKC^c subgroup clades, only motif 11 was found in SOC1 clades, only motif 14 was found in the SVP clade, and only motif 18 in the SEP clade (Figure 4), respectively. In addition, all Mα proteins were found to have only motif 1 and 3 MADS domains.

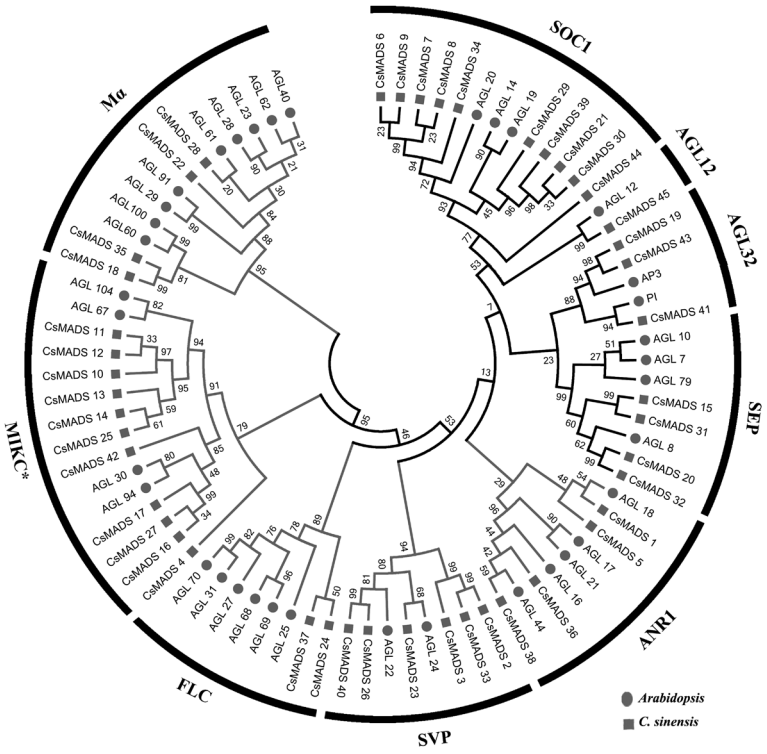

Figure 1. Phylogenetic analysis for *C. sinensis* and *Arabidopsis* MADS-box proteins.

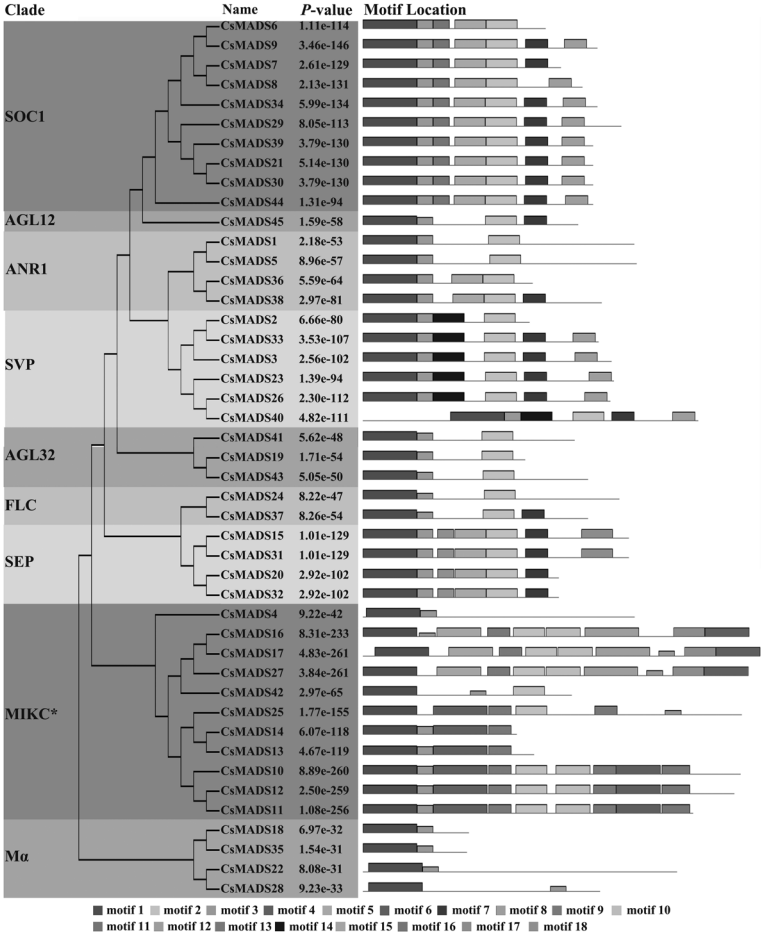

Figure 2. CsMADS-box protein motifs identified by MEME analysis.

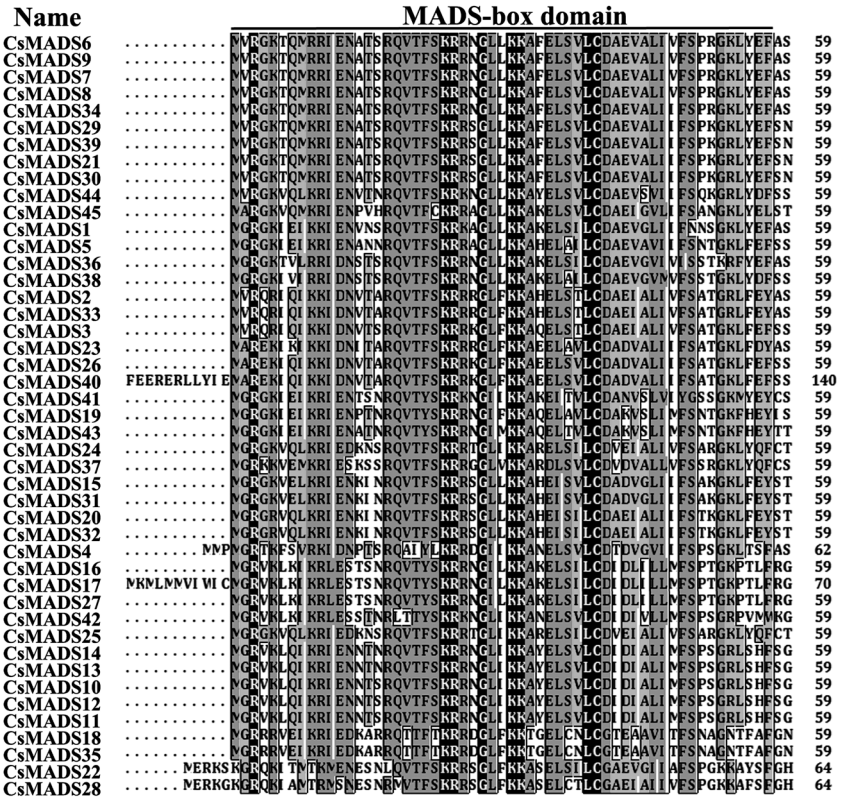

Figure 3. The conserved MADS-box DNA-binding motif. CsMADS-box amino acid sequences are constructed by multiple sequence alignment for searching MADS-box motifs.

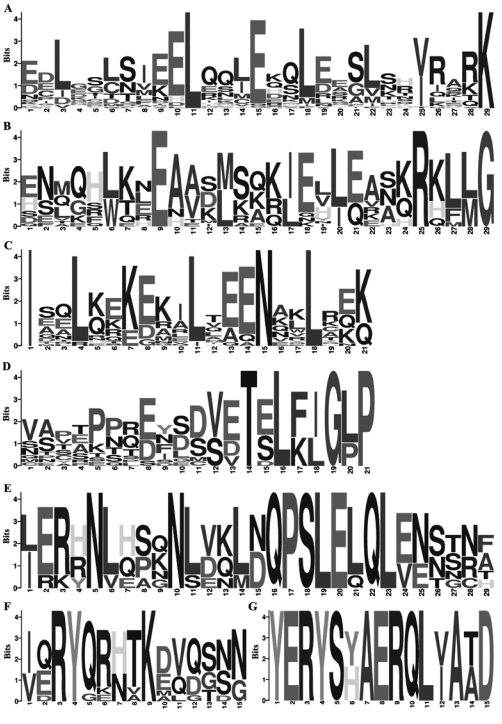

Figure 4. Sequence logos of MADS-box motifs in *C. sinensis*. Stack height shows the information content of that position in the motif. Height of residues within the stack shows the probability of each residue at that position. A: motif 2; B: motif 5; C: motif 7; D: motif 8; E: motif 11. F: motif 14; G: motif 18. Motif 2, motif 5 and motif 7 represent the K domain.

Tissue-specific expression patterns of CsMADS-box genes

Since no MADS-box TFs in tea plant have been previously identified, we selected 14 CsMADS-box genes of each MIKC-type subgroup and investigated the expression pattern of these genes in tea plant tissues. The results showed differential expression of the 14 MADS-box genes throughout the tea plant (Figure 5). Our results indicated that 12 CsMADS-box genes were highly expressed in polarized tip growth and flower formation. For example, three CsMADS-box genes (*CsMADS16*, *CsMADS19*, and *CsMADS27*) exhibited higher expression levels in flower, pollen, and pollen tube tissues. Among other CsMADS-box genes, *CsMADS5*, *CsMADS31*, and *CsMADS38* were highly expressed in the pollen tube, *CsMADS45* in flowers, *CsMADS8* and *CsMADS23* in root tissue, and *CsMADS40* in bud tissue. However, some CsMADS-box genes were highly expressed in other tissues. For example, *CsMADS37* and *CsMADS44* showed higher expression level in the fourth leaves, *CsMADS24* in the first and fourth leaves, and *CsMADS2* in stems. The tissue expression profiles of CsMADS-box genes may provide useful information for investigating the growth and development of *C. sinensis*.

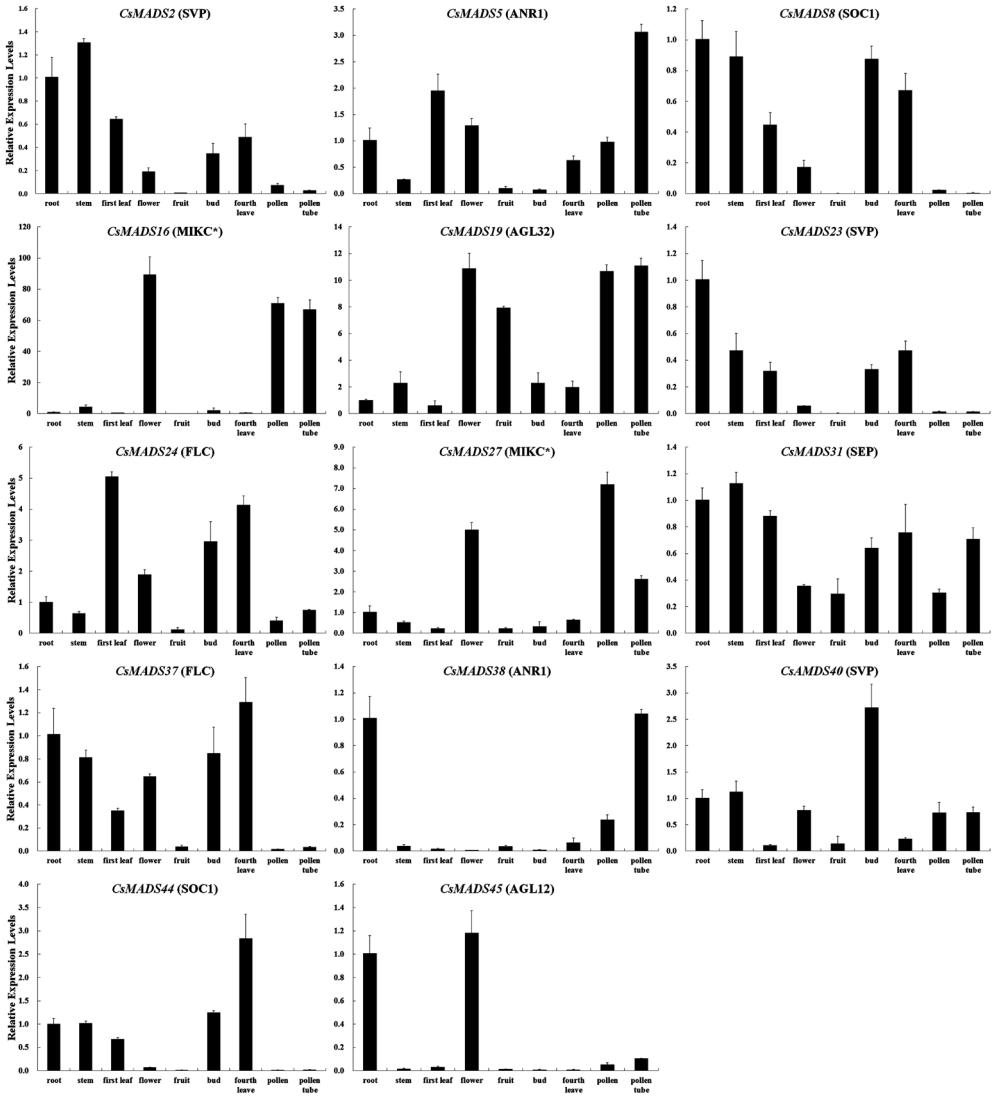

Figure 5. Expression profiles of 14 *CsMADS-box* genes in different *C. sinensis* tissues.

Expression profiles of CsMADS-box genes in response to Al and F treatment

To elucidate the function of CsMADS-box genes in roots and to provide a basis for future functional analysis, qRT-PCR was used to reveal differential expression of the 14 CsMADS-box genes under different Al and F concentration treatments (Figures 6, 7). Preliminary experimental results (unpublished) showed that new root growth of *C. sinensis* was faster at an Al treatment concentration of 0.4 mmol/l than it was at 0 or 2 mmol/l. After Al treatments were performed, six CsMADS-box genes (*CsMADS8*, *-16*, *-37*, *-38*, *-40*, and *-44*) were upregulated at 3 h under the 0 mmol/l treatment, and *CsMADS8*, *-37*, and *-44* expression levels continued to increase after 6 h. Conversely, other CsMADS-box genes were gradually downregulated after the 0 mmol/l treatment. In response to the 0.4 mmol/l Al treatment, expression levels of *CsMADS8*, *-16*, *-37*, *-38*, and *-40* gradually increased, and peak expression levels of *CsMADS16*, *-37*, and *-40* were detected at 3 h, while expression peaks of *CsMADS8* and *CsMADS38* were detected at 6 h and 12 h, respectively. In addition, *CsMADS2*, *-23*, *-24*, *-27*, *-31*, and *-44* were initially downregulated and then later upregulated in two distinct stages. In response to the 2 mmol/l Al treatment, six CsMADS-box genes (*CsMADS8*, *-16*, *-31*, *-37*, *-38*, and *-40*) were upregulated, and expression peaked after 3 h, after which they were gradually downregulated. However, relative expression levels of other genes were gradually downregulated after the first treatment. In general, all genes investigated responded to Al treatment in varying degrees.

In the F treatment (Figure 7), the relative expression trends of *CsMADS19* and *CsMADS24* were consistent across all F treatments. In the 0 mg/l F treatment, expression of six CsMADS-box genes (*CsMADS2*, *-8*, *-16*, *-38*, *-40*, and *-44*) increases, although all six reached peak expression at different times. In addition, four

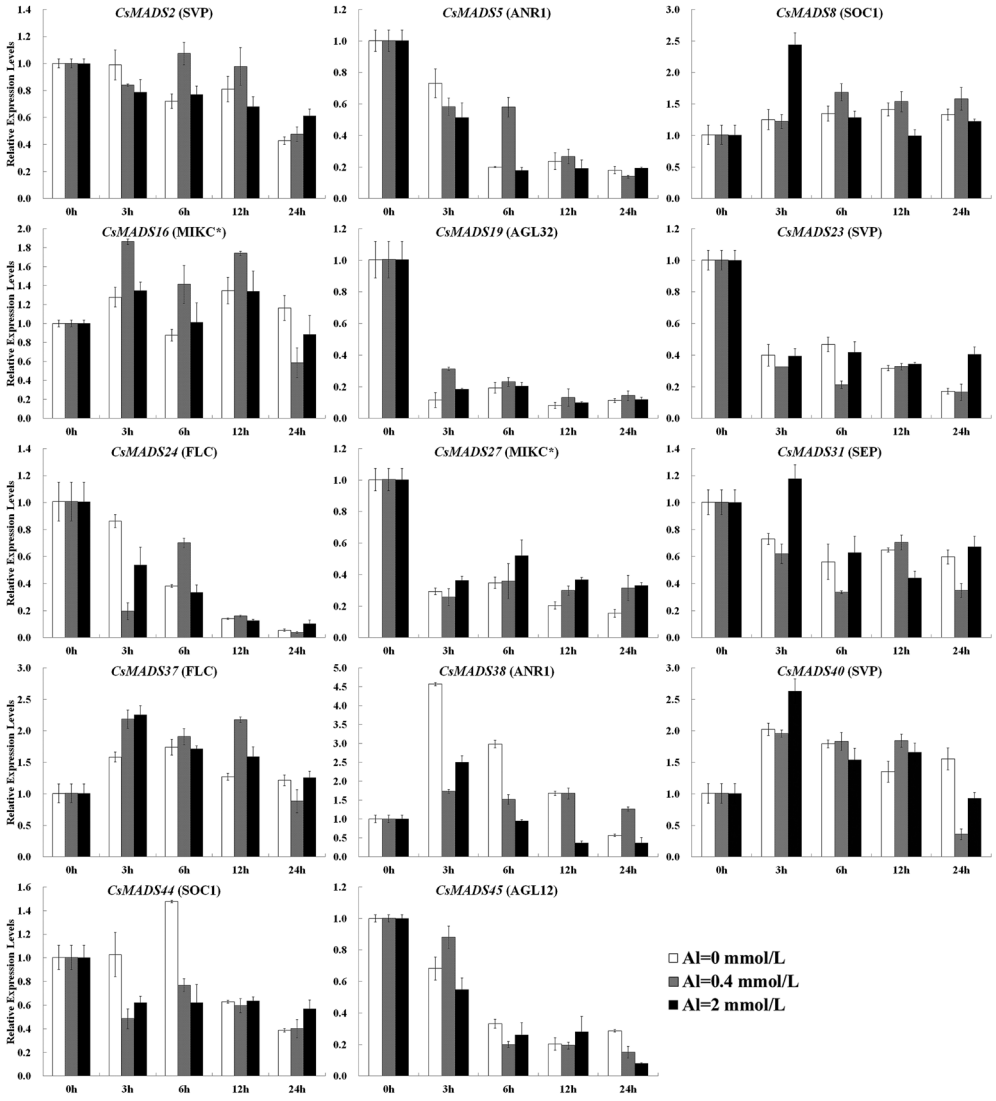

Figure 6. qRT-PCR was used to analyze the expression profiles of *CsMADS-box* genes under different Al concentration treatments.

CsMADS-box genes (*CsMADS5*, *-23*, *-27*, and *-45*) were initially downregulated, but were later upregulated. In the 8 mg/l F treatment, *CsMADS8*, *-37*, and *-38* genes were highly upregulated after 6 h. In addition, the relative expression levels of *CsMADS2*, *-5*, *-16*, *-44*, and *-45* were always lower than those of the 0 mg/l F treatment. In the 16 mg/l F treatment, seven CsMADS-box genes (*CsMADS8*, *-16*, *-37*, *-38*, *-40*, *-44*, and *-45*) were initially upregulated but were then gradually downregulated. Moreover, the relative expression levels of *CsMADS37* and *CsMADS40* were higher in all treatments than they were in the control treatment, i.e., the 0 mg/l F treatment. Conversely, the relative expression levels of *CsMADS2*, *CsMADS5*, and *CsMADS31* were lower in all experimental treatments than they were in the 0 mg/l F control treatment.

Discussion

In previous studies, genome-wide analyses have identified many members of the MADS-box TF family in *Arabidopsis* (Parenicova et al. 2003), *Oryza sativa* (Arora et al. 2007), *Glycine max* (Fan et al. 2013), *Vitis vinifera* (Grimplet et al. 2016), and other plants (Hu and Liu 2012; Leseberg et al. 2006; Wei et al. 2015; Wells et al. 2015; Zhao et al. 2011) (Table 1). However, there are few studies on MADS-box TFs in *C. sinensis* (Gogoi et al. 2015; Wang et al. 2016). Therefore, in the present study we used transcriptomic data from *C. sinensis* in order to identify and analyze putative members of the MADS-box TFs family in the tea plant. Compared with other species, *C. sinensis* had only idendified 45 MADS-box members so far. Furthermore, we found that woody plants—such as *Populus trichocarpa*, *Prunus persica*, and *Vitis vinifera*—generally had fewer MADS-box genes than did herbaceous plants, such as *Glycine max* and

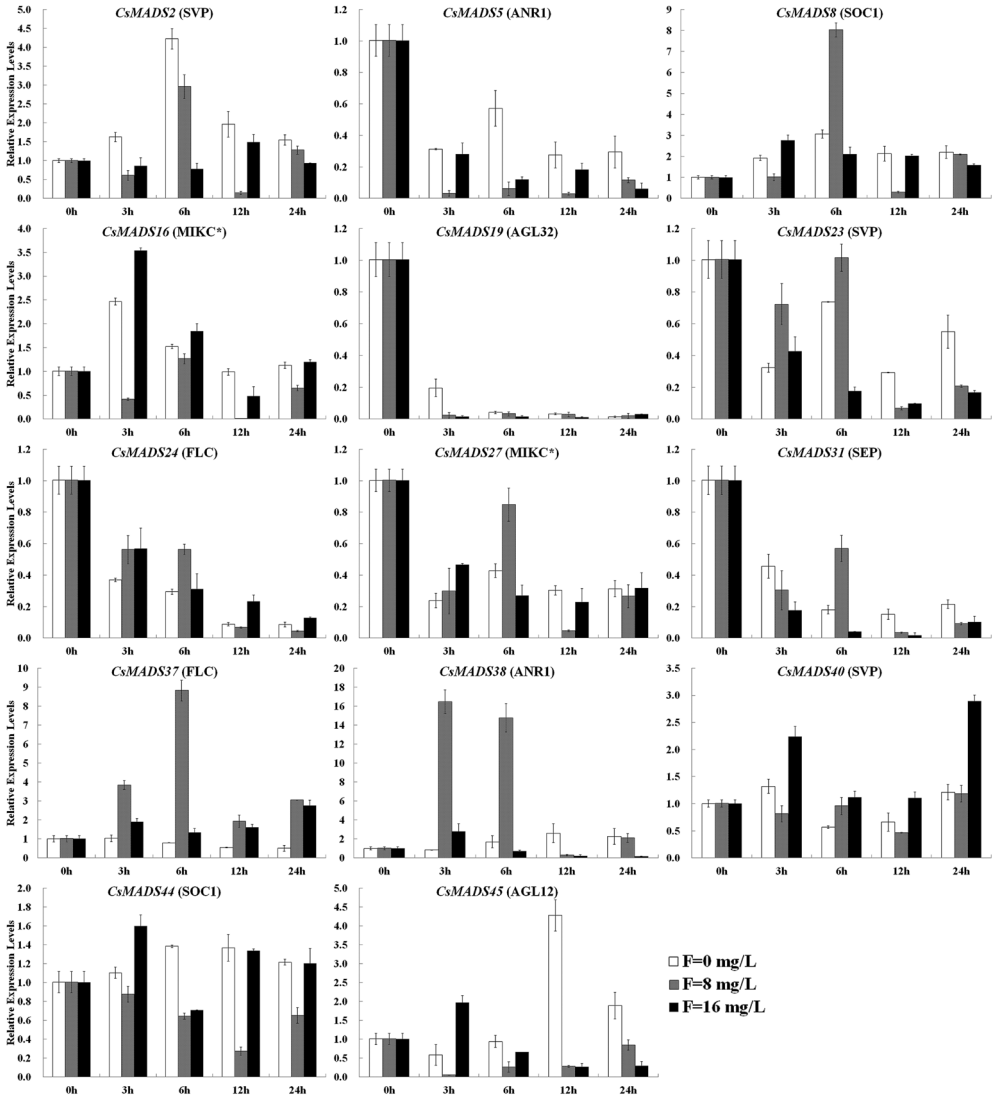

Figure 7. qRT-PCR was used to analyze the expression profiles of *CsMADS-box* genes under different F concentration treatments.

Arabidopsis (Table 1). The number and distribution of MADS-box TFs found in the *C. sinensis* transcriptome were both similar to those found in *Populus trichocarpa*, *Prunus persica*, and *Vitis vinifera*, although it remains true that some MADS-box genes were not captured by our analysis.

In this study, we identified 45 CsMADS-box genes from *C. sinensis* transcriptomic data and classified them on the basis of their phylogenetic similarity to orthologues found in *Arabidopsis*. Furthermore, we analyzed the conserved motifs of *C. sinensis* MADS-box proteins using MEME software. Analysis of 18 motifs from all CsMADS proteins recognized motif 1 and motif 3 as the MADS-box domain, which was found to be present in all CsMADS-box genes. Furthermore, the CsMADS-box DNA-binding motif has identified by the presence of approximately 58 amino acids (*CArG* boxes) (Figure 3) (Messenguy and Dubois 2003). Phylogenetic analysis showed that 4 CsMADS-box genes classified as M-type members of the MADS-box family. The other 41 proteins containing the K domain (motif 2, 5 and 7) were classified as MIKC-type (MIKC^c and MIKC*), with the MIKC^c subgroup itself subdivided into 7 separate clades. Furthermore, we conducted an tissue expression study of individual CsMADS-box genes from MIKC* and MIKC^c clades. The results exhibited a wide variety of expression patterns in each MIKC-type clades.

SOC1-like genes

SOC1-like genes are important transcriptional regulators in the phase change from vegetative to reproductive development in numerous plant species, and are characterized by a specific major expression domain (Becker and Theissen 2003; Liu et al. 2016). We identified ten SOC1-like genes (*CsMADS6*, *-7*, *-8*, *-9*, *-21*, *-29*, *-30*, *-34*, *-39*, and *-44*) in *C. sinensis*. Analysis of conserved motifs showed that only the motif 11 (Figure 4) was found in all SOC1 clades, this was the main difference

Table 1. Summary of MADS-box transcription factors among 11 species.

Species	Type I	Mα	Mβ	Mγ	Type II	MIKC^c	MIKC*	Total
Camellia sinensis	4	4	0	0	41	30	11	45
Populus trichocarpa	41	23	12	6	64	55	9	105
Sesamum indicum	24	14	0	10	33	28	5	57
Glycine max	75	37	14	24	89	82	7	163
Cucumis sativus	10	5	2	3	33	29	4	43
Prunus persica	40	21	7	12	39	35	4	79
Arabidopsis thaliana	62	24	22	16	46	39	7	108
Vitis vinifra	42	23	0	19	48	42	6	90
Zea mays	32	27	3	2	43	39	4	75
Sorghum bicolor	30	26	2	2	35	33	2	65
Oryza sativa	31	12	9	10	41	37	4	72

between the SOC1-like subgroup and other MIKC^c clades. Furthermore, previous work on SOC1-like genes has shown different expression patterns in diverse plant. For example, the *DnAGL19* gene was observed to be highly expressed in *pseudobulbs*, axillary buds, leaves, and roots but not in flowers (Liu et al. 2016). In addition, the *IbAGL20* gene is expressed at similar levels in leaves, vegetative shoots, flowers, and root tissue (Kim et al. 2005). In the present study, we found that *CsMADS8* and *CsMADS44* are most highly expressed in root, stem, bud, and leaf tissues. These results indicated that *C. sinensis* and the above species showed similar patterns of SOC1 subfamily gene expression (Kim et al. 2005; Liu et al. 2016).

AGL12-like genes

AGL12-like genes encode other MADS-box TFs, and have been isolated from root tissues in both monocot and eudicot species (Parenicova et al. 2003). In *Arabidopsis*, *AGL12* is found priority expressed roots. Interesting, it is also expressed in both flowers and shoots (Alvarez-Buylla et al. 2000a). Among the CsMADS-box genes we analyzed, one *CsMADS45* gene was grouped into the AGL12-like clade, and was found to be highly expressed in flower and root tissues of *C. sinensis*. Genes from this subfamily are involved in promoter of the flowering transition and regulator of root cell proliferation in *Arabidopsis* (Tapia-Lopez et al. 2008). Another current study has shown that AGL12 genes are involved in root cell wall synthesis (Montes et al. 2014). Combined with our results, we speculate that *CsMADS45* has a similar role in root and flower development.

AGL32-like genes

AGL32-like clades, including DEFICIENS (DEF) and GLO (PI), genes belong to the B class of floral homeotic genes in eudicots, which participate in specifying petal and stamen tissue identity in flower growth and development period (Winter et al. 2002). However, B-class MADS-box genes in *Physalis floridana* play no role in organ identity determination but continue to be active in pollen maturation (Zhang et al. 2015a). Our analysis revealed three AGL32-like genes (*CsMADS19*, *41* and *43*). *CsMADS19* was abundant in pollen and in pollen tube tissue of *C. sinensis* flowers. Thus, it is tempting to speculate that *CsMADS19* may function as a B-class floral homeotic gene in stamen tissues. However, the specific role played by *CsMADS19* in pollen growth needs to be verified further.

SEP-like genes

SEP-like genes are representative expressed in floral meristems or in inflorescences (Becker and Theissen 2003). Our analyses identified four SEP-like genes (*CsMADS15*, *-20*, *-31*, and *-32*). Motif 18 (Figure 4) was used as a basis for discrimination between the SEP-like genes and other clades. In addition, tissue expression levels of *CsMADS31* were high in root, stem, and first leaf. Our results are similar to those of Gu et al. (1998), where they found the SEP-like gene '*FRUITFULL*,' which regulates meristem and leaf identity in stem and leaf tissue of *Arabidopsis*.

ANR1-like genes

ANR1-like genes, which are thought to be responsible for floral induction, have been identified in many plants (Arora et al. 2007; Fan et al. 2013; Grimplet et al. 2016; Hu and Liu 2012; Leseberg et al. 2006; Parenicova et al. 2003; Wei et al. 2015; Wells et al. 2015; Zhao et al. 2011). We identified four ANR1-like genes in *C. sinensis* (*CsMADS1*, *-5*, *-36*, and *-38*). Becker and Theissen (2003) reported that ANR1-like genes have generally multiple expression patterns, as they are found in root tissues (as are most ANR1-like genes), in pollen (*DEFH125*), in both roots and pollen (*ZmMADS2*), or in leaf guard cells and trichromes (*AGL16*). In the present study, the measured expression levels of *CsMADS5* genes were higher in pollen tubes than in pollen; thus *CsMADS5* may be involved in polarized growth of pollen tubes. In addition, the expression pattern of the *CsMADS38* gene was broadly consistent with that of *ZmMADS2*.

SVP-like genes

The genome of *Arabidopsis* contains two SVP-like genes, *AGL24* and *SVP*. These genes regulate floral transition, with AGL24 promoting floral meristem identity, and SVP repressing the floral transition (Hartmann et al. 2000; Yu et al. 2002). Six SVP-like genes (*CsMADS2*, *-3*, *-23*, *-26*, *-33*, and *-40*) were detected in *C. sinensis*. This clade has a unique motif 14 (Figure 4) that distinguishes it from other clades. *CsMADS2* and *CsMADS40* were both highly expressed in stem and bud tissue. Previous work showed that *SVP* is highly expressed during the shoot apical meristem is inhibited in the inflorescence apical meristem development process of *Arabidopsis* (Hartmann et al. 2000). We speculate that *CsMADS2* and *CsMADS40* may be similar to *SVP*, but the actual function of *CsMADS2* and *CsMADS40* remains unknown.

FLC-like genes

FLOWERING LOCUS C (FLC)-like genes repress flowering in *Arabidopsis* by their concentrated action on environmental and endogenous pathways (Sheldon et al. 2000). We found two FLC-like genes in *C. sinensis* (*CsMADS24* and *37*). Both *CsMADS24* and *CsMADS37* were more highly expressed in newly rooted seedlings (e.g., with fourth leaves) than were CsMADS-box genes from other gene clades. Since MADS-box genes are well known for the role they play in floral development, meristem identity, and polarized cell growth (Ma and dePamphilis 2000), it is possible that *CsMADS24* and *CsMADS37* may, under normal circumstances, be expressed at a relatively low level. However, both the structure and function of *CsMADS24* and *CsMADS37* require further elucidation before any conclusions can be drawn.

MIKC-like genes*

Our analyses revealed the presence of eleven MIKC*-like genes (*CsMADS4*, *-10*, *-11*, *-12*, *-13*, *-14*, *-16*, *-17*, *-25*, *-27*, and *-42*) in *C. sinensis*, and phylogenetic comparisons indicated that these were distinct from other MIKC genes. Previous work has established that expression of MIKC* genes during the bud stage of the life cycles of *B. rapa* and *Arabidopsis* is associated with male reproductive organ development (Saha et al. 2015; Verelst et al. 2007). In the present study, relatively high expression of *CsMADS16* and *CsMADS27* in pollen and pollen tube tissue also showed that these genes may play a role in male organ development in *C. sinensis*.

Al and F accumulation-responsive CsMADS-box genes

C. sinensis is known to be an Al hyperaccumulator. Previous results have shown that low concentrations of Al^{3+} promote root growth in *C. sinensis*, but high concentrations of Al^{3+} do not, and are instead a stress (Hajiboland and Poschenrieder 2015). In this study, we examined gene expression levels of 14 *CsMADS-box* genes in *C. sinensis* root under different concentrations of Al. While SOC1-, SEP-, AGL12-, and ANR1-like gene subgroups have all been known to be closely linked to root growth (Becker and Theissen 2003; Gu et al. 1998), our results revealed differential expression patterns of the *C. sinensis* orthologues of these MADS-box genes (*CsMADS5*, *-8*, *-31*, *-38*, *-44*, and *-45*) in response to Al accumulation. *CsMADS38* and *-44* were involved in *C. sinensis* root growth but were inhibited by Al. However, the expression of *CsMADS5* and *-8* was higher at a low concentration of Al, but was lower in response to the high concentration Al treatment. This result, combined with pre-experimental data, suggests that *CsMADS5* and *-8* genes accelerate root growth of *C. sinensis* in response to stimulation from low, but not high concentrations of Al. In addition, *SVP* (*CsMADS2*), *FLC* (*CsMADS37*), and *MIKC** (*CsMADS16*) genes were also highly expressed when exposed to a low concentration of Al. Previous work has shown that the SVP-like genes of *I. batatas* (*IbMADS3* and *IbMADS4*) were mainly expressed in the vascular cambium region of roots (Kim et al. 2005). We speculate that the *CsMADS2* may be a close association with *C. sinensis* root growth and development in response to low concentration Al stimulation. Moreover, *MIKC** regulates male reproductive organ development (Verelst et al. 2007) and *FLC* acts as a repressor of the floral transition (Sheldon et al. 2000), so both are implicated in flower growth of *Arabidopsis*. However, to date there is no evidence that *FLC* and *MIKC** also regulate root growth. The function of *FLC* (*CsMADS37*) and *MIKC** (*CsMADS16*) in *C. sinensis* root growth in response to Al stimulation requires further investigation.

In accumulator plants such as *C. sinensis*, root uptake of F is an active, energy dependent process (Zhang et al. 2013). However, the molecular mechanism responsible for high F accumulation in *C. sinensis* roots is unknown. In this study, the relatively high expression levels of *CsMADS8*, *-37*, and *-38* genes in response to the 8 mg/l F treatment suggests that these genes may somehow influence F accumulation in tea root. Conversely, root uptake of F may inhibit *CsMADS2*, *-5*, and *-45* gene expression. In addition, *CsMADS16* and *-40* were relatively highly expressed in response to the 16 mg/l F treatment. This data combined with knowledge of the root growth patterns of *C. sinensis* in response to different concentrations of F, leads us to speculate that *CsMADS16* and *-40* are also involved in F accumulation of *C. sinensis* roots. However, we acknowledge that gene expression levels do not perfectly correlate with protein expression, and thus relying only on qPCR data may not perfectly describe the roles *CsMADS-box* genes play in response to F accumulation. The specific function

of *CsMADS-box* genes therefore need to be examined further.

Conclusion

This work represents a comprehensive analysis of 45 CsMADS-box TFs identified from analyses of transcriptome sequences. We also performed phylogenetic analysis to identify conserved sequences and potential functional motifs of CsMADS-box genes in order to provide a basis for further research into their structure and function. Furthermore, we investigate the expression patterns of CsMADS-box genes in various tissues. These results provide a foundation for future study on the function of CsMADS-box genes. Finally, the expression of CsMADS-box genes in response to Al and F treatments of different concentrations showed that CsMADS-box genes are involved in *C. sinensis* root growth in response to mild Al and F accumulation. These findings provide a basis for investigating for future studies on CsMADS-box genes involved in Al and F accumulation in *C. sinensis*.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 31770733), the Fundamental Research Funds for the Central Universities (No. KYZ201509), the Opening Project of State Key Laboratory of Tea Plant Biology and Utilization (No. SKLTOF20150113), the earmarked fund for China Agriculture Research System (CARS-19) and Modern Agriculture Industry System in Jiangsu (SXGC [2017] 279).

Conflict of interest

The authors have no conflict of interest to declare.

Author contributions

Conceived and designed the experiments: Yuhua Wang, Junting Pan. Performed the experiments: Junting Pan, Pinpin Chang, Xiaoli Ye, Jiaojiao Zhu. Analyzed the data: Junting Pan, Chuanlei Cui, Xiaoli Ye, Dongqin Li. Wrote and revised the manuscript: Junting Pan, Bo Wen, Yuanchun Ma, Yuhua Wang, Xujun Zhu, Wanping Fang.

References

Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF (2000a) MADS-box gene evolution beyond flowers: Expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J* 24: 457–466

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana LR, Martinez-Castilla L, Yanofsky MF (2000b) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA* 97: 5328–5333

Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S (2007) MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* 8: 242

Becker A, Theissen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol* 29: 464–489

Burgeff C, Liljegren SJ, Tapia-Lopez R, Yanofsky MF, Alvarez-Buylla ER (2002) MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots. *Planta* 214: 365–372

Cao J, Zhao Y, Liu JW, Xirao RD, Danzeng SB, Daji DW, Yan Y (2003) Brick tea fluoride as a main source of adult fluorosis. *Food Chem Toxicol* 41: 535–542

Chen F, Zhang X, Liu X, Zhang LS (2017) Evolutionary analysis of MIKC^c-Type MADS-Box genes in gymnosperms and angiosperms. *Front Plant Sci* 8: 895

Cho S, Jang S, Chae S, Chung KM, Moon YH, An G, Jang SK (1999) Analysis of the C-terminal region of *Arabidopsis thaliana* APETALA1 as a transcription activation domain. *Plant Mol Biol* 40: 419–429

Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S, Roberts K, Scheres B (1993) Cellular-organization of the *Arabidopsis-Thaliana* root. *Development* 119: 71–84

Erdemoğlu SB, Pyrzyniska K, Güçer S (2000) Speciation of aluminum in tea infusion by ion-exchange resins and flame AAS detection. *Anal Chim Acta* 411: 81–89

Fan CM, Wang X, Wang YW, Hu RB, Zhang XM, Chen JX, Fu YF (2013) Genome-wide expression analysis of soybean MADS genes showing potential function in the seed development. *PLoS One* 8: e62288

Forester SC, Lambert JD (2011) The role of antioxidant versus pro-oxidant effects of green tea polyphenols in cancer prevention. *Mol Nutr Food Res* 55: 844–854

Fung KF, Carr HP, Zhang J, Wong MH (2008) Growth and nutrient uptake of tea under different aluminium concentrations. *J Sci Food Agric* 88: 1582–1591

Fung KF, Zhang ZQ, Wong JWC, Wong MH (1999) Fluoride contents in tea and soil from tea plantations and the release of fluoride into tea liquor during infusion. *Environ Pollut* 104: 197–205

Gan Y, Filleur S, Rahman A, Gotensparre S, Forde BG (2005) Nutritional regulation of ANR1 and other root-expressed MADS-box genes in *Arabidopsis thaliana*. *Planta* 222: 730–742

Gauthier E, Fortier I, Courchesne F, Pepin P, Mortimer J, Gauvreau D (2000) Aluminum forms in drinking water and risk of Alzheimer's disease. *Environ Res* 84: 234–246

Gogoi M, Borchetia S, Bandyopadhyay T (2015) Computational identification and analysis of MADS box genes in *Camellia sinensis*. *Bioinformation* 11: 115–121

Grimplet J, Martinez-Zapater JM, Carmona MJ (2016) Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genomics* 17: 80

Gu Q, Ferrandiz C, Yanofsky MF, Martienssen R (1998) The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125: 1509–1517

Guo A, He K, Liu D, Bai SN, Gu XC, Wei LP, Luo JC (2005) DATF: A database of *Arabidopsis* transcription factors. *Bioinformatics* 21: 2568–2569

Hajiboland R, Poschenrieder C (2015) Localization and compartmentation of Al in the leaves and roots of tea plants. *Phyton-Int J Exp Bot* 84: 86–100

Hartmann U, Hohmann S, Nettesheim K, Wisman E, Saedler H, Huijser P (2000) Molecular cloning of SVP: A negative regulator

of the floral transition in *Arabidopsis*. *Plant J* 21: 351–360

Hu L, Liu S (2012) Genome-wide analysis of the MADS-box gene family in cucumber. *Genome* 55: 245–256

Kim S, Hamada T, Otani M, Shimada T (2005) Isolation and characterization of MADS box genes possibly related to root development in sweetpotato (*Ipomoea batatas* L. Lam.). *J Plant Biol* 48: 387–393

Leseberg CH, Li AL, Kang H, Duvall M, Mao L (2006) Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene* 378: 84–94

Li C, Xu H, Xu J, Chun X, Ni D (2011) Effects of aluminium on ultrastructure and antioxidant activity in leaves of tea plant. *Acta Physiol Plant* 33: 973–978

Liu XR, Pan T, Liang WQ, Gao L, Wang XJ, Li HQ, Liang S (2016) Overexpression of an orchid (*Dendrobium nobile*) SOC1/TM3-like ortholog, DnAGL19, in *Arabidopsis* regulates HOS1-FT expression. *Front Plant Sci* 7: 99

Ma H, dePamphilis C (2000) The ABCs of floral evolution. *Cell* 101: 5–8

Messenguy F, Dubois E (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* 316: 1–21

Montes RAC, Coello G, González-Aguilera KL, Marsch-Martínez N, de Folter S, Alvarez-Buylla ER (2014) ARACNe-based inference, using curated microarray data, of *Arabidopsis thaliana* root transcriptional regulatory networks. *BMC Plant Biol* 14: 97

Nagata T, Hayatsu M, Kosuge N (1993) Aluminum kinetics in the tea plant using ^{27}Al and ^{19}F NMR. *Phytochemistry* 32: 771–775

Pan J, Wang W, Li D, Shu Z, Ye X, Chang P, Wang Y (2016) Gene expression profile indicates involvement of NO in *Camellia sinensis* pollen tube growth at low temperature. *BMC Genomics* 17: 809

Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, 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

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: e45

Ruan JY, Ma LF, Shi YZ, Han WY (2004) The impact of pH and calcium on the uptake of fluoride by tea plants (*Camellia sinensis* L.). *Ann Bot* 93: 97–105

Ruan JY, Wong MH (2001) Accumulation of fluoride and aluminium related to different varieties of tea plant. *Environ Geochem Hlth* 23: 53–63

Saedler H, Becker A, Winter KU, Kirchner C, Theissen G (2001) MADS-box genes are involved in floral development and evolution. *Acta Biochim Pol* 48: 351–358

Saha G, Park JI, Jung HJ, Ahmed NU, Kayum MA, Chung MY, Hur Y, Cho YG, Watanabe M, Nou IS (2015) Genome-wide identification and characterization of MADS-box family genes related to organ development and stress resistance in *Brassica rapa*. *BMC Genomics* 16: 178

Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES (2000) The molecular basis of vernalization: The central role of FLOWERING LOCUS C (FLC). *Proc Natl Acad Sci USA* 97: 3753–3758

Simpson A, Shaw L, Smith AJ (2001) The bio-availability of fluoride from black tea. *J Dent* 29: 15–21

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30: 2725–2729

Tapia-López R, García-Ponce B, Dubrovsky JG, Garay-Arroyo A, Pérez-Ruíz RV, Kim SH, Acevedo F, Pelaz S, Alvarez-Buylla ER (2008) An AGAMOUS-related MADS-box gene, XAL1 (AGL12), regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. *Plant Physiol* 146: 1182–1192

Verelst W, Saedler H, Munster T (2006) MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific *Arabidopsis* promoters. *Plant Physiol* 143: 447–460

Wan Q, Xu RK, Li XH (2012) Proton release by tea plant (*Camellia sinensis* L.) roots as affected by nutrient solution concentration and pH. *Plant Soil Environ* 58: 429–434

Wang L, Wang Y, Cao H, Hao X, Zeng J, Yang Y, Wang X (2016) Transcriptome analysis of an anthracnose-resistant tea plant cultivar reveals genes associated with resistance to *Colletotrichum camelliae*. *PLoS One* 11: e0148535

Wei X, Wang LH, Yu JY, Zhang YX, Li DH, Zhang XR (2015) Genome-wide identification and analysis of the MADS-box gene family in sesame. *Gene* 569: 66–76

Wells CE, Vendramin E, Tarodo SJ, Verde I, Bielenberg DG (2015) A genome-wide analysis of MADS-box genes in peach [*Prunus persica (L.) Batsch*]. *BMC Plant Biol* 15: 41

Wenzel WW, Blum WEH (1992) Fluorine speciation and mobility in f-contaminated soils. *Soil Science* 153: 357–364

Winter KU, Weiser C, Kaufmann K, Bohne A, Kirchner C, Kanno A, Saedler H, Theissen G (2002) Evolution of class B floral homeotic proteins: Obligate heterodimerization originated from homodimerization. *Mol Biol Evol* 19: 587–596

Wong MH, Fung KF, Carr HP (2003) Aluminium and fluoride contents of tea, with emphasis on brick tea and their health implications. *Toxicol Lett* 137: 111–120

Xie ZM, Ye ZH, Wong MH (2001) Distribution characteristics of fluoride and aluminum in soil profiles of an abandoned tea plantation and their uptake by six woody species. *Environ Int* 26: 341–346

Yang Y, Liu Y, Huang CF, de Silva J, Zhao FJ (2016) Aluminium alleviates fluoride toxicity in tea (*Camellia sinensis*). *Plant and Soil* 402: 179–190

Yu H, Xu YF, Tan EL, Kumar PP (2002) AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. *Proc Natl Acad Sci USA* 99: 16336–16341

Zhang L, Li Q, Ma L, Ruan J (2013) Characterization of fluoride uptake by roots of tea plants (*Camellia sinensis* (L.) O. Kuntze). *Plant and Soil* 366: 659–669

Zhang S, Zhang JS, Zhao J, He C (2015a) Distinct subfunctionalization and neofunctionalization of the B-class MADS-box genes in *Physalis floridana*. *Planta* 241: 387–402

Zhang XC, Gao HJ, Wu HH, Yang TY, Zhang ZZ, Mao JD, Wan XC (2015b) Ca^{2+} and CaM are involved in Al^{3+} pretreatment-promoted fluoride accumulation in tea plants (*Camellia sinesis* L.). *Plant Physiol Biochem* 96: 288–295

Zhao Y, Li X, Chen W, Peng X, Cheng X, Zhu S, Cheng B (2011) Whole-genome survey and characterization of MADS-box gene family in maize and sorghum. *Plant Cell Tiss Organ Cult* 105: 159–173