

# Whole-genome survey and characterization of MADS-box gene family in maize and sorghum

Yang Zhao · Xiaoyu Li · Wenjuan Chen ·  
Xiaojian Peng · Xiao Cheng · Suwen Zhu ·  
Beijiu Cheng

Received: 5 July 2010 / Accepted: 6 September 2010 / Published online: 17 September 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** MADS-box genes comprise a large gene family, which codes for transcription factors, and play important functions in various aspects of flowering plant growth and development. However, little is known about the MADS-box genes in maize (*Zea mays*) and sorghum (*Sorghum bicolor*). Here, we performed a comprehensive bioinformatics analysis of the MADS-box gene family in the maize and sorghum genomes and identified 75 maize and 65 sorghum MADS-box genes. We subsequently carried out a comparative analysis of these genes, including the gene structure, phylogenetic relationship, conserved protein motifs, gene duplications, chromosomal locations and expression pattern between the two plants. According to these analyses, the MADS-box genes in both maize and sorghum were categorized into five (MIKC<sup>C</sup>, MIKC\*, M $\alpha$ , M $\beta$  and M $\gamma$ ) groups, and the MIKC<sup>C</sup> groups were further divided into 11 subfamilies. In addition, gene duplications of MADS-box genes were also investigated in the maize, sorghum, rice and *Arabidopsis* genomes. We found a higher percentage of MADS-box gene duplications in the maize and sorghum genomes, which contributed to the expansion of the MADS-box gene family. Furthermore, both tandem and segmental duplications played a major role in the MADS-box gene expansion in maize and sorghum. A survey of maize and sorghum EST sequences indicated that MADS-box genes exhibit a various expression pattern, suggesting diverse and novel functions of

MADS-box gene families in the two plants. These results provided a useful reference for selection of candidate MADS-box genes for cloning and further functional analysis in both maize and sorghum.

**Keywords** MADS-box · Maize · Sorghum · Phylogenetic analysis · Duplication · Expression patterns

## Introduction

The MADS-box gene family encodes transcription factors that are widely distributed in eukaryotic organisms (ranging from yeast and plants to insects and mammals), and many of these genes have important biological functions in eukaryotes (Riechmann and Meyerowitz 1997; Becker et al. 2000; Messenguy and Dubois 2003). This gene family is defined by the presence of a conserved 58 ~ 60 amino acids MADS-box domain whose name is based on four members: the *Mini Chromosome Maintenance 1* (*MCMI*) gene in yeast, *Agamous* (*AG*) in *Arabidopsis*, *Deficiens* (*DEF*) in snapdragon and *Serum response factor* (*SRF*) in humans (Passmore et al. 1988; Norman et al. 1988; Yanofsky et al. 1990; Sommer et al. 1990). The MADS domain is involved in binding to DNA sequences (CArG box) based on the consensus CC(A/T)<sub>6</sub>GG sequence (Pellegrini et al. 1995). Indeed, many genes are likely regulated by MADS-box gene family members through CArG boxes that are present in their promoter regions (Shore and Sharrocks 1995). In plants, increasing evidence suggests that MADS-box genes play crucial roles in the processes of flowering plant growth, which influence flowering time (i.e. genes such as *SOC1*, *FLC* and *SVP*), formation and meristem identity of floral organs (*API*, *FUL* and *AG*), fruit ripening (*SHP1* and *SHP2*), seed

---

Yang Zhao and Xiaoyu Li contributed equally to this work.

---

Y. Zhao · X. Li · W. Chen · X. Peng · X. Cheng · S. Zhu ·  
B. Cheng (✉)

Key Laboratory of Crop Biology of Anhui Province,  
Anhui Agricultural University, 230036 Hefei, China  
e-mail: bejiucheng@ahau.edu.cn

pigmentation and embryo development (*TRANSPARENT TESTA16*) (Alwee et al. 2006; Samach et al. 2000; Michaels and Amasino 1999; Hartmann et al. 2000; Mandel et al. 1992; Gu et al. 1998; Liljegren et al. 2000; Nesi et al. 2002; Parenicova et al. 2003; Arora et al. 2007).

Based on phylogenetic analysis, the MADS-box gene family in eukaryotes was divided into *SRF*-type and *MEF2*-type genes (Theißen et al. 1996). As a result of an ancestral duplication that occurred in MADS-box genes before the divergence of plants from animals to fungi, the MADS-box family can be divided into two main groups: type I and type II (MIKC) genes. The members of the type I group, including *SRF*-like genes, have been found in animals, fungi and plants; the type II group includes *MEF2*-like genes, found in animals and yeast, and this group also contains the plant-specific MIKC-type MADS-box genes (Alvarez-Buylla et al. 2000; De Bodt et al. 2003). In plants, the type II MADS-box genes of known function are a group of the most extensively studied and described transcription factors (Purugganan 1997; Theissen et al. 2000; Kaufmann et al. 2005).

Apart from the MADS (M) domain, the MIKC-type genes contain three additional functional domains: the Intervening (I), the kertain-like (K) and the C-terminal (C) domains, from N- to the C-terminus (Norman et al. 1988; Theißen et al. 1996; Cho et al. 1999; Yang et al. 2003). The K domain is a second conserved region among MIKC-type MADS-box genes, which is comprised of ~70 amino acids and mainly contributes to the dimerization by a coiled-coil structure (Ma et al. 1991; Davies et al. 1996). However, the type I genes in plants lack the typical K domain that is found in the type II MADS-box genes (Parenicova et al. 2003; Díaz-Riquelme et al. 2009). The I domain is a less conserved region of ~30 amino acids, which may influence DNA-binding specificity and dimerization (Riechmann et al. 1996). According to the structural divergence of the I domain, the MIKC-type genes have been categorized into MIKC<sup>C</sup>- and MIKC\*-types (Henschel et al. 2002). The C-terminal domain (C-domain) is a highly diverse region, but it has been demonstrated that the C-domain has important functions in transcriptional activation and in the interaction of MADS-box proteins (Kramer and Irish 1999; Honma and Goto 2001). Based on the features of the M domain sequences, the type I MADS-box genes have been further classified into M $\alpha$ , M $\beta$ , M $\gamma$  and M $\delta$  groups. In *Arabidopsis*, the M $\delta$  group consists of six genes: AGL30, 94, 65, 67, 104 and 66, which are considered to be MIKC\*-type genes according to the phylogenetic relationship of the two groups (De Bodt et al. 2003; Kofuji et al. 2003).

Maize and sorghum are globally important agricultural crops for humans and animals. In previous studies, 34 members of the MADS-box gene family were identified

based on specific probes designed from the conserved MADS-box domain in maize (Mena et al. 1995), but a genome-wide analysis of the MADS-box gene family has not been undertaken in maize or in sorghum. Recently, the maize and sorghum genomes have been sequenced (Paterson et al. 2009; Schnable et al. 2009), and the completion of the whole genome sequences of maize and sorghum provided an opportunity for us to analyze the MADS-box gene family of these two plants. The whole genome sequences also provide a valuable resource for comparative analyses, which can be used to further understand the evolutionary history of the MADS-box family in the two plants. In this study, a complete set of 75 and 65 MADS-box genes were identified in the maize and sorghum genomes, respectively. Moreover, gene duplication events and expression patterns of MADS-box genes in maize and sorghum were also surveyed. Our results may be helpful for functional studies of the MADS-box genes, and these bioinformatics analyses also provide a biological reference for cloning of MADS-box genes, which can contribute to genetic breeding in maize and sorghum.

## Materials and methods

### Identification and classification of MADS-box genes

Maize genome sequences were downloaded from <http://www.maizesequence.org/>. The complete genome sequences of sorghum were obtained from <http://genome.jgi-psf.org/Sorbi1/Sorbi1.download ftp.html>. Similar methods were employed to identify the MADS-box genes in maize and sorghum as described in other plants (Zhou et al. 2004; Leseberg et al. 2006).

First, the Hidden Markov Model (HMM) profile of the MADS-box domain was employed as a query to identify all possible MADS-box genes in the maize and sorghum genomes using the BlastP program ( $P$  value = 0.001). In order to identify the maximum number of MADS-box domain-containing sequences, two different HMM profiles were adopted to search for the MADS-box genes. One was obtained from the Pfam database (<http://Pfam.sanger.ac.uk/Software/Pfam>) (Finn et al. 2006), and the other profile was generated by alignments of 107 *Arabidopsis* MADS-box genes (Parenicova et al. 2003). Second, the Pfam database was used to determine if each of the candidate MADS-box sequences was a member of the MADS-box family. Additionally, all of the candidate MADS-box proteins were divided into type I and type II groups according to whether they encoded the K domain. To exclude overlapping genes, all of the candidate MADS-box genes were aligned using ClustalW (Thompson et al. 1994) and

checked manually. All of the nonoverlapping MADS-box genes were used for further analysis.

### Phylogenetic analysis

In order to better understand the phylogenetic relationships, and further classify the type I and type II MADS-box genes, two individual phylogenetic trees (one for type I and one for type II genes) were constructed using the complete amino acid sequences of the MADS-box genes from maize, sorghum and *Arabidopsis* genomes. The phylogenetic trees were generated using MEGA v4.0 (Tamura et al. 2007) with the neighbor-joining (NJ) method and using complete MADS-box protein sequences. The maize and sorghum MADS-box genes were then named according to their phylogenetic relationship with the corresponding *Arabidopsis* MADS-box genes.

### Analysis of conserved motifs and gene duplication

To examine the structural divergence among the MADS-box genes in maize and sorghum, the conserved motifs in the encoded MADS-box proteins were investigated. Their complete amino acid sequences were subjected to MEME (Multiple Expectation Maximization for Motif Elicitation) (Bailey and Elkan 1995) analysis online ([http://meme.sdsc.edu/meme4\\_3\\_0/intro.html](http://meme.sdsc.edu/meme4_3_0/intro.html)). Due to the limitations of the software, and also for analyses of individual conserved motifs in more detail, the MIKC<sup>C</sup>, MIKC\*, M $\alpha$ , M $\beta$  and M $\gamma$  groups were analyzed separately using the MEME search tool with the following parameters: (1) optimum motif width was set to  $\geq 6$  and  $\leq 200$ ; and (2) the maximum number of motifs was set to identify 20 motifs. SMART program (<http://smart.embl-heidelberg.de>) and Pfam database were adopted to annotate the MEME motifs.

Gene duplication events of MADS-box genes in maize and sorghum were also investigated. All of the relevant genes identified in the maize and sorghum genomes were aligned using ClustalW and calculated using MEGA v4.0 (Yang et al. 2008).

### Chromosomal locations of MADS-box genes

In order to determine the physical locations of MADS-box genes, the starting position of all MADS-box genes on each chromosome were confirmed by BlastN searching, using local databases of the complete nucleotide sequences of the maize and sorghum genomes. MapInspect software was used to draw the location images of the maize and sorghum MADS-box genes ([http://www.plantbreeding.wur.nl/uk/software\\_mapinspect.html](http://www.plantbreeding.wur.nl/uk/software_mapinspect.html)).

In silico analysis of expression pattern of maize and sorghum MADS-box genes

We used NCBI EST database to detect the expression patterns of maize and sorghum MADS-box proteins using megablast tool (<http://www.ncbi.nlm.nih.gov/dbEST/>). To search the expression data of MADS-box genes, the following parameters were adopted: maximum identity > 95%, length > 200 bp and *E* value < 10<sup>-10</sup>.

## Results

### Identification and classification of MADS-box genes in maize and sorghum

We used BlastP searches based on the conserved MADS-box domain HMM profile to identify the MADS-box genes in the maize and sorghum genomes. To identify the maximum number of MADS-box genes in the two plants, two different HMM profiles were used to search for the MADS-box genes (see Materials and methods). Interestingly, the total numbers of MADS-box genes identified using the two different HMM profiles were the same in both maize and sorghum. By this approach, 75 and 65 MADS-box genes were identified in maize and sorghum, respectively. The total number of maize MADS-box genes was similar with the number present in sorghum. According to phylogenetic tree analysis, 43 genes were determined to be type II MADS-box genes, and 32 were confirmed as type I MADS-box genes (including M $\alpha$ , M $\beta$  and M $\gamma$  groups) among the 75 maize representatives. In sorghum, 35 type II and 30 type I (consisting of M $\alpha$ , M $\beta$  and M $\gamma$  groups) MADS-box genes were identified. More interestingly, the total numbers of both type I and type II group MADS-box genes in maize and rice are the same, and the total numbers of both type I and type II group genes in *Arabidopsis* are greater than those in maize, sorghum and rice (Table 1).

**Table 1** The total number of MADS-box genes within each group in maize, sorghum, rice and *Arabidopsis*

Category	Maize	Sorghum	Rice <sup>a</sup>	<i>Arabidopsis</i> <sup>b</sup>
Total	75	65	75	107
Total type II genes	43	35	43	45
MIKC <sup>C</sup>	39	33	39	39
MIKC*	4	2	4	8
Total type I genes	32	30	32	61
M $\alpha$	27	26	13	25
M $\beta$	3	2	9	20
M $\gamma$	2	2	10	16

<sup>a</sup> Arora et al. (2007)

<sup>b</sup> Parenicova et al. (2003)

According to the survey of intron numbers, we found that maize and sorghum type II MADS-box genes usually contain multiple introns, and type I genes usually have introns with a range of 0–4 except ZmMADS44, 61, 63 and 71. The same phenomenon was also found in rice and *Arabidopsis* (Parenicova et al. 2003; Arora et al. 2007).

#### Comparative phylogenetic analyses of MADS-box genes in maize, sorghum and *Arabidopsis*

To examine the phylogenetic relationships of MADS-box genes and investigate the evolutionary history of this gene family among maize, sorghum and *Arabidopsis*, two individual phylogenetic trees of type I and type II group genes were constructed, respectively. To better view the phylogenetic relationships, shortened gene names were used in the phylogenetic trees (Table 2). The type II genes phylogenetic tree was constructed from 123 MADS-box sequences (43 maize, 35 sorghum and 45 *Arabidopsis* type II genes) (Fig. 1).

Based on the phylogenetic tree analysis, we divided the tree into 11 subfamilies. All of the MADS-box genes of maize and sorghum can be grouped with their *Arabidopsis* counterparts except for ZmMADS2, 14, 34 and SbMADS32 and 30. The tree shows that there are six maize homologs each for *Arabidopsis* *SEP*, *TT16* and *AP3/PI*, three each for *SOC1*, *SVP* and *AGL12*-like, five for *CAL/FUL/API*, and four for *AG/SHP*. The *CAL/FUL/API*, *SVP* and *AG/SHP* subfamilies each contain four sorghum homologs. Three sorghum homologs each were confirmed for *Arabidopsis* *SOC1*, *AP3/PI* and *AGL12*-like, two for *TT16* and seven for *SEP*. Interestingly, aside from one sorghum homolog grouped with *Arabidopsis* *ANRI*, no maize or sorghum MADS-box genes of the *FLC* and *ANRI* subfamilies were found. The number of *AG/SHP*, *SOC1* and *AGL12*-like homologs in maize and sorghum was the same. Obviously, besides the *AP3/PI* and *TT16* groups, the total number of plant-specific MIKC-type MADS-box genes grouped into each subfamily was similar between maize and sorghum. According to this analysis, we concluded that maize and sorghum MIKC-type genes may stem from a common ancestor, which belongs to the same species of grass. In addition, four maize (ZmMADS40, 41, 42 and 43) and two sorghum (SbMADS35 and 34) MADS-box genes were grouped with the *Arabidopsis* MIKC\*-type genes (Fig. 1).

In rice and *Arabidopsis*, the total number of genes in  $M\alpha$ ,  $M\beta$  and  $M\gamma$  groups was similar. However, unlike the rice and *Arabidopsis* type I MADS-box genes, the  $M\alpha$  group included the most members of type I genes in both maize and sorghum (Fig. 2).

#### Analysis of conserved motif and gene duplication

A total of 140 MADS-box genes from maize to sorghum were subjected to MEME motif analysis to discover conserved motifs shared among related proteins. As previously described (Parenicova et al. 2003; Arora et al. 2007), conserved motifs encoding the MADS-box domain can be found in each maize and sorghum MADS-box gene (Fig. 3). In the maize and sorghum genomes, a total of 72 MADS-box sequences were identified as MIKC<sup>C</sup>-type genes. However, only 50 MIKC<sup>C</sup>-type MADS-box genes were found, including three conserved motifs (motif 5, 2 and 3) of the K domain, while only motif 5 was found in ZmMADS31 and 6 and motif 3 in SbMADS23. This result confirmed the findings in other plants that the K domain is a weakly conserved region (Parenicova et al. 2003; Arora et al. 2007). Four maize and two sorghum MADS-box genes have predicted coiled coil motifs. In addition, several unknown motifs were also identified by MEME motif analysis.

To investigate MADS-box gene duplication events in maize and sorghum, the following criteria was adopted (Yang et al. 2008; Gu et al. 2002): (1) the length of alignable sequence covered  $\geq 80\%$  of the longer gene; and (2) the similarity of the aligned regions was  $\geq 70\%$ . Additionally, a detailed comparison of gene duplication was performed using all of the MADS-box genes from the maize, sorghum, rice and *Arabidopsis* genomes. As a result, 28 MADS-box gene subfamilies (including 62 MADS-box genes) were confirmed among the 75 maize MADS-box genes, and 15 subfamilies (including 36 genes) were found among the 65 sorghum MADS-box genes. The ratio of multi-genes/single-genes of maize (4.77) was significantly higher than in sorghum (1.24), rice (0.53) and *Arabidopsis* (1.74). The average number of members of each MADS-box gene subfamily was 2.21, 2.4, 2.17 and 4.25 in maize, sorghum, rice and *Arabidopsis*, respectively (Table 3). These results suggest that MADS-box gene duplications in maize and sorghum were more prevalent than in rice.

#### Chromosomal locations of MADS-box genes

The physical locations of the MADS-box genes on maize and sorghum chromosomes are depicted in Fig. 4 a and b, respectively. In maize, the largest number of MADS-box genes was located on chromosome 8 (ten). The same number of genes (nine) was located on chromosomes 5, 6 and 9, chromosomes 2 and 7 both contained seven genes, six were found on chromosomes 1 and 4, and eight and four genes were located on chromosomes 3 and 10, respectively (Fig. 4 a).

**Table 2** Characters of maize and sorghum MADS-box genes

Name	Protein ID	Accession no.	Type	Introns	Name	Protein ID	Accession no.	Type	Introns
ZmMADS1	AC233912.1_FGP001	NA	MIKC <sup>C</sup>	16	SbMADS1	Sb01g007780.1	XM_002463818	MIKC <sup>C</sup>	6
ZmMADS2	GRMZM2G001139_P01	BT068004	MIKC <sup>C</sup>	4	SbMADS2	Sb01g030570.1	NA	MIKC <sup>C</sup>	6
ZmMADS3	GRMZM2G003514_P01	BT087751	MIKC <sup>C</sup>	7	SbMADS3	Sb01g042840.1	XM_002465598	MIKC <sup>C</sup>	7
ZmMADS4	GRMZM2G005155_P03	BT069133	MIKC <sup>C</sup>	6	SbMADS4	Sb01g044810.1	NA	MIKC <sup>C</sup>	8
ZmMADS5	GRMZM2G012432_P03	AJ292960	MIKC <sup>C</sup>	6	SbMADS5	Sb01g044810.3	NA	MIKC <sup>C</sup>	7
ZmMADS6	GRMZM2G012432_P04	BT064585	MIKC <sup>C</sup>	5	SbMADS6	Sb01g049020.1	XM_002465916	MIKC <sup>C</sup>	6
ZmMADS7	GRMZM2G018589_P01	BT062110	MIKC <sup>C</sup>	7	SbMADS7	Sb02g001090.1	XM_002459187	MIKC <sup>C</sup>	7
ZmMADS8	GRMZM2G046885_P01	EU966783	MIKC <sup>C</sup>	7	SbMADS8	Sb02g029310.1	XM_002460446	MIKC <sup>C</sup>	7
ZmMADS9	GRMZM2G046885_P02	NA	MIKC <sup>C</sup>	5	SbMADS9	Sb02g038780.1	XM_002460944	MIKC <sup>C</sup>	7
ZmMADS10	GRMZM2G052890_P01	NM_001111851	MIKC <sup>C</sup>	8	SbMADS10	Sb03g002525.1	XM_002454940	MIKC <sup>C</sup>	8
ZmMADS11	GRMZM2G069370_P01	NM_001152058	MIKC <sup>C</sup>	6	SbMADS11	Sb03g041860.1	XM_002458825	MIKC <sup>C</sup>	6
ZmMADS12	GRMZM2G071620_P01	BT087351	MIKC <sup>C</sup>	8	SbMADS12	Sb03g042080.1	XM_002456713	MIKC <sup>C</sup>	6
ZmMADS13	GRMZM2G072582_P02	EU961835	MIKC <sup>C</sup>	7	SbMADS13	Sb04g000500.1	XM_002451301	MIKC <sup>C</sup>	6
ZmMADS14	GRMZM2G079727_P01	NM_001152881	MIKC <sup>C</sup>	7	SbMADS14	Sb04g004736.1	XM_002453325	MIKC <sup>C</sup>	5
ZmMADS15	GRMZM2G097059_P01	EU960801	MIKC <sup>C</sup>	7	SbMADS15	Sb04g028960.1	NA	MIKC <sup>C</sup>	6
ZmMADS16	GRMZM2G099522_P02	BT017740	MIKC <sup>C</sup>	7	SbMADS16	Sb04g031750.1	XM_002454427	MIKC <sup>C</sup>	7
ZmMADS17	GRMZM2G105387_P01	NA	MIKC <sup>C</sup>	5	SbMADS17	Sb04g033930.1	NA	MIKC <sup>C</sup>	7
ZmMADS18	GRMZM2G110153_P01	BT064849	MIKC <sup>C</sup>	6	SbMADS18	Sb06g017660.1	XM_002447855	MIKC <sup>C</sup>	7
ZmMADS19	GRMZM2G117961_P01	NA	MIKC <sup>C</sup>	6	SbMADS19	Sb06g026300.1	XM_002446938	MIKC <sup>C</sup>	9
ZmMADS20	GRMZM2G128953_P01	EU975425	MIKC <sup>C</sup>	5	SbMADS20	Sb06g028420.1	XM_002447045	MIKC <sup>C</sup>	5
ZmMADS21	GRMZM2G129034_P01	NA	MIKC <sup>C</sup>	7	SbMADS21	Sb07g001250.1	XM_002443699	MIKC <sup>C</sup>	6
ZmMADS22	GRMZM2G135018_P01	NA	MIKC <sup>C</sup>	7	SbMADS22	Sb07g026200.1	XM_002445769	MIKC <sup>C</sup>	7
ZmMADS23	GRMZM2G137387_P01	NA	MIKC <sup>C</sup>	11	SbMADS23	Sb08g006430.1	XM_002442975	MIKC <sup>C</sup>	6
ZmMADS24	GRMZM2G137510_P01	NM_001158884	MIKC <sup>C</sup>	7	SbMADS24	Sb08g006460.1	XM_002441944	MIKC <sup>C</sup>	7
ZmMADS25	GRMZM2G137510_P03	NA	MIKC <sup>C</sup>	6	SbMADS25	Sb08g015555.1	XM_002443174	MIKC <sup>C</sup>	2
ZmMADS26	GRMZM2G139073_P01	EU966237	MIKC <sup>C</sup>	6	SbMADS26	Sb09g006360.1	NA	MIKC <sup>C</sup>	8
ZmMADS27	GRMZM2G147716_P01	EU962906	MIKC <sup>C</sup>	7	SbMADS27	Sb09g020770.1	XM_002439779	MIKC <sup>C</sup>	6
ZmMADS28	GRMZM2G148220_P01	NA	MIKC <sup>C</sup>	5	SbMADS28	Sb10g004390.1	XM_002437838	MIKC <sup>C</sup>	7
ZmMADS29	GRMZM2G148693_P01	BT033167	MIKC <sup>C</sup>	7	SbMADS29	Sb10g007380.1	XM_002436665	MIKC <sup>C</sup>	7
ZmMADS30	GRMZM2G152862_P01	NM_001111667	MIKC <sup>C</sup>	6	SbMADS30	Sb10g026690.1	XM_002437376	MIKC <sup>C</sup>	4
ZmMADS31	GRMZM2G152862_P02	BT064585	MIKC <sup>C</sup>	5	SbMADS31	Sb10g029810.1	XM_002438958	MIKC <sup>C</sup>	6
ZmMADS32	GRMZM2G160565_P01	NM_001111862	MIKC <sup>C</sup>	7	SbMADS32	Sb03g033380.1	XM_002458383	MIKC <sup>C</sup>	3
ZmMADS33	GRMZM2G160687_P03	BT036695	MIKC <sup>C</sup>	7	SbMADS33	Sb01g007790.1	XM_002463819	MIKC <sup>C</sup>	7
ZmMADS34	GRMZM2G161666_P01	NA	MIKC <sup>C</sup>	6	SbMADS34	Sb05g025970.1	XM_002449891	MIKC*	8
ZmMADS35	GRMZM2G171365_P01	NM_001111682	MIKC <sup>C</sup>	7	SbMADS35	Sb10g007810.1	XM_002436689	MIKC*	8
ZmMADS36	GRMZM2G171365_P03	NA	MIKC <sup>C</sup>	6	SbMADS36	Sb02g002400.1	XM_002459273	M $\alpha$	0
ZmMADS37	GRMZM2G171365_P04	NM_001111682	MIKC <sup>C</sup>	6	SbMADS37	Sb02g002410.1	XM_002459274	M $\alpha$	0
ZmMADS38	GRMZM2G174005_P01	BT041585	MIKC <sup>C</sup>	7	SbMADS38	Sb02g002430.1	XM_002459275	M $\alpha$	0
ZmMADS39	GRMZM2G359952_P01	NM_001112476	MIKC <sup>C</sup>	8	SbMADS39	Sb02g005005.1	XM_002459416	M $\gamma$	1
ZmMADS40	GRMZM2G441115_P01	NA	MIKC*	5	SbMADS40	Sb02g006700.1	XM_002459522	M $\alpha$	0
ZmMADS41	GRMZM2G334225_P02	NA	MIKC*	9	SbMADS41	Sb02g006710.1	XM_002459523	M $\alpha$	0
ZmMADS42	GRMZM2G152415_P01	NM_001158900	MIKC*	11	SbMADS42	Sb02g006730.1	XM_002459525	M $\alpha$	0
ZmMADS43	GRMZM2G334225_P01	NA	MIKC*	10	SbMADS43	Sb02g006750.1	XM_002459526	M $\alpha$	0
ZmMADS44	AC212823.4_FGP003	NA	M $\alpha$	7	SbMADS44	Sb03g006210.1	XM_002457329	M $\alpha$	1
ZmMADS45	AC216256.3_FGP006	NM_001148155	M $\alpha$	0	SbMADS45	Sb03g027570.1	XM_002455891	M $\alpha$	0
ZmMADS46	AC234185.1_FGP006	EU962654	M $\alpha$	0	SbMADS46	Sb03g027580.1	XM_002455892	M $\alpha$	1
ZmMADS47	GRMZM2G032905_P01	NA	M $\alpha$	3	SbMADS47	Sb03g044170.1	XM_002456815	M $\alpha$	4
ZmMADS48	GRMZM2G035092_P01	NA	M $\alpha$	0	SbMADS48	Sb04g004410.1	XM_002451550	M $\alpha$	0
ZmMADS49	GRMZM2G052045_P01	NA	M $\alpha$	1	SbMADS49	Sb04g024010.1	XM_002452297	M $\alpha$	1
ZmMADS50	GRMZM2G055782_P01	NA	M $\alpha$	1	SbMADS50	Sb05g023351.1	XM_002449746	M $\gamma$	0

**Table 2** continued

Name	Protein ID	Accession no.	Type	Introns	Name	Protein ID	Accession no.	Type	Introns
ZmMADS51	GRMZM2G065234_P01	NA	M $\alpha$	3	SbMADS51	Sb0836s002010.1	XM_002488936	M $\alpha$	0
ZmMADS52	GRMZM2G072582_P03	NA	M $\alpha$	2	SbMADS52	Sb06g015640.1	XM_002446375	M $\alpha$	0
ZmMADS53	GRMZM2G073357_P01	NA	M $\alpha$	0	SbMADS53	Sb06g019040.1	XM_002446577	M $\alpha$	2
ZmMADS54	GRMZM2G098986_P01	NA	M $\alpha$	2	SbMADS54	Sb07g007306.1	XM_002445221	M $\alpha$	0
ZmMADS55	GRMZM2G099577_P01	NA	M $\alpha$	0	SbMADS55	Sb07g007373.1	XM_002445222	M $\alpha$	0
ZmMADS56	GRMZM2G105387_P04	NA	M $\alpha$	0	SbMADS56	Sb07g021110.1	XM_002445494	M $\alpha$	1
ZmMADS57	GRMZM2G129034_P02	NA	M $\alpha$	4	SbMADS57	Sb07g026180.1	XM_002445767	M $\alpha$	4
ZmMADS58	GRMZM2G130382_P01	EU962253	M $\alpha$	0	SbMADS58	Sb08g005892.1	XM_002442939	M $\alpha$	0
ZmMADS59	GRMZM2G152862_P03	BT040464	M $\alpha$	4	SbMADS59	Sb09g016220.1	XM_002440865	M $\alpha$	1
ZmMADS60	GRMZM2G163335_P01	NM_001157552	M $\alpha$	0	SbMADS60	Sb09g029120.1	XM_002441502	M $\alpha$	0
ZmMADS61	GRMZM2G171365_P02	NA	M $\alpha$	6	SbMADS61	Sb10g002680.1	XM_002437747	M $\alpha$	0
ZmMADS62	GRMZM2G320549_P01	NA	M $\alpha$	1	SbMADS62	Sb10g021663.1	XM_002438492	M $\alpha$	0
ZmMADS63	GRMZM2G340130_P01	NA	M $\alpha$	7	SbMADS63	Sb10g019770.1	XM_002437007	M $\alpha$	0
ZmMADS64	GRMZM2G375707_P01	NM_001174288	M $\alpha$	0	SbMADS64	Sb03g013650.1	XM_002455551	M $\beta$	0
ZmMADS65	GRMZM2G446426_P01	NA	M $\alpha$	4	SbMADS65	Sb02g037430.1	XM_002463033	M $\beta$	1
ZmMADS66	GRMZM2G459864_P01	NA	M $\alpha$	0					
ZmMADS67	GRMZM2G471698_P01	NA	M $\alpha$	1					
ZmMADS68	GRMZM2G472096_P01	NA	M $\alpha$	0					
ZmMADS69	GRMZM2G472100_P01	NA	M $\alpha$	0					
ZmMADS70	GRMZM2G475390_P01	NA	M $\alpha$	2					
ZmMADS71	GRMZM2G137289_P01	NA	M $\beta$	6					
ZmMADS72	GRMZM2G038878_P01	NA	M $\beta$	2					
ZmMADS73	GRMZM2G306610_P01	NA	M $\beta$	0					
ZmMADS74	GRMZM2G035836_P01	NA	M $\gamma$	0					
ZmMADS75	GRMZM2G411477_P01	NA	M $\gamma$	0					

ZmMADS: MADS-box genes of *Zea mays*

SbMADS: MADS-box genes of *Sorghum bicolor*

Chr chromosome numbers

NA not available

In sorghum, chromosome 2 had the largest number of genes (eleven) followed by chromosome 3 (nine). Seven genes were located on chromosome 4, eight each on chromosomes 1 and 10, and four each on chromosomes 8 and 9. Five and six genes were found on chromosomes 6 and 7, respectively, and only two genes were located on chromosome 5. Interestingly, only the MIKC<sup>C</sup>-type genes were located on chromosome 1 (Fig. 4 b).

It should be noted that MADS-box genes appeared on each chromosome in both maize and sorghum. This phenomenon indicates that the distributions of MADS-box genes are wide in both genomes, which may represent the distributed characteristics of the MADS-box gene family in grass plants. However, the distribution of MADS-box genes on each chromosome in both maize and sorghum were significantly non-random. Indeed, relatively high densities of MADS-box genes were observed on some chromosomes, such as maize chromosome 8 and sorghum chromosome 2.

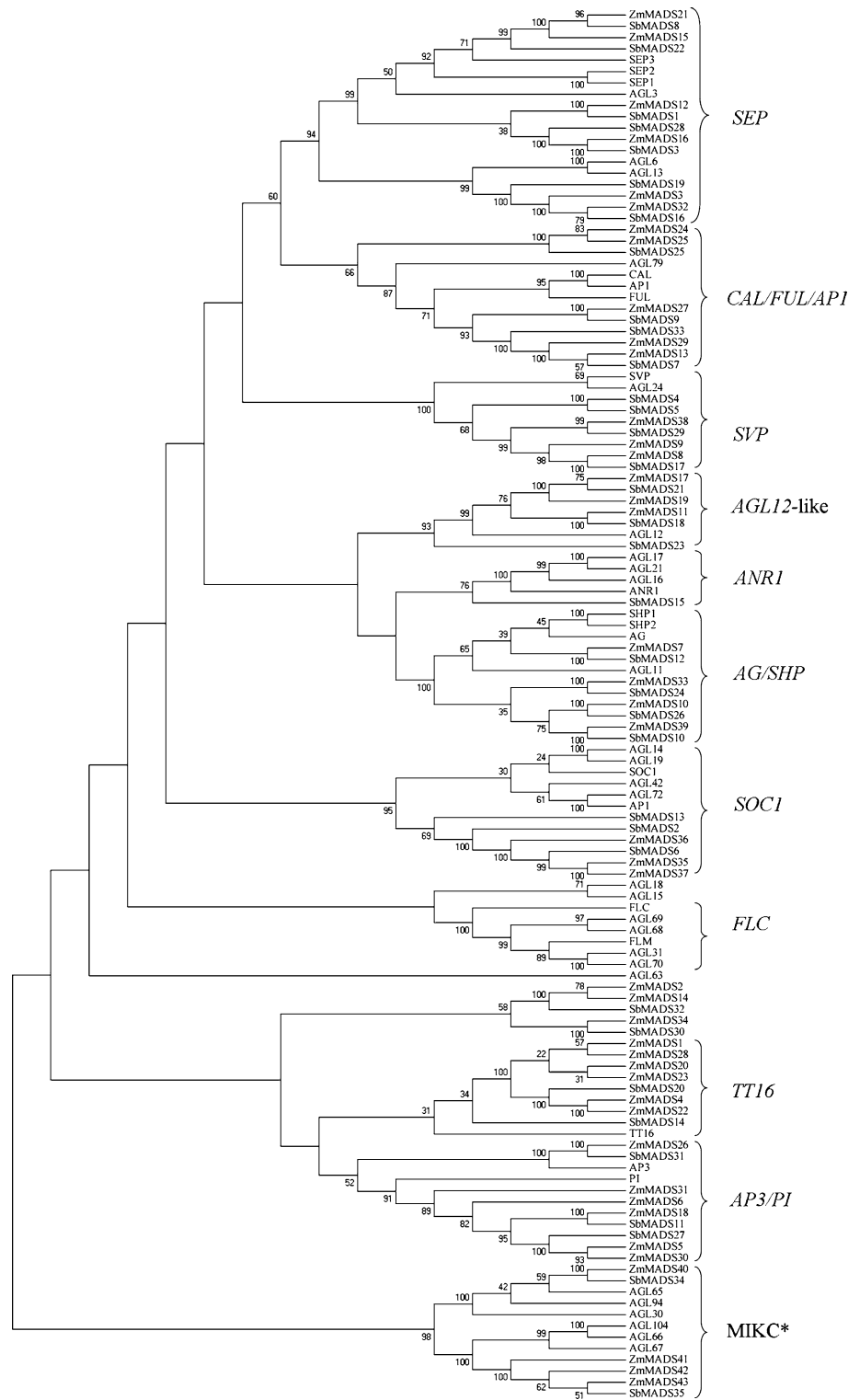
Like *Arabidopsis* and rice MADS-box genes (Parenicova et al. 2003; Arora et al. 2007), several of the MADS-box

genes are physically clustered (Fig. 4) in the maize and sorghum genomes. According to Holub's (2001) definition, a gene cluster is a chromosome region which contains two or more genes within 200 kb (Holub 2001). In maize, we found that 23 MADS-box genes were located in 9 clusters. Among the nine clusters, two were located on chromosome 7 and one each on chromosomes 2, 4, 5, 6, 8, 9 and 10. No cluster was found on chromosomes 1 or 3. In sorghum, nine clusters including 22 genes were found. Chromosomes 1, 2, 3 and 7 each contained two clusters, and only one cluster was found on chromosomes 8. Chromosomes 4, 5, 6, 9 and 10 did not contain any cluster. The largest cluster comprised of five and four MADS-box genes on maize and sorghum chromosomes, respectively.

#### Expression patterns analysis of maize and sorghum MADS-box genes

Since EST data can provide valuable information for gene expression research, we examined the expression patterns of maize and sorghum MADS-box genes in various tissues

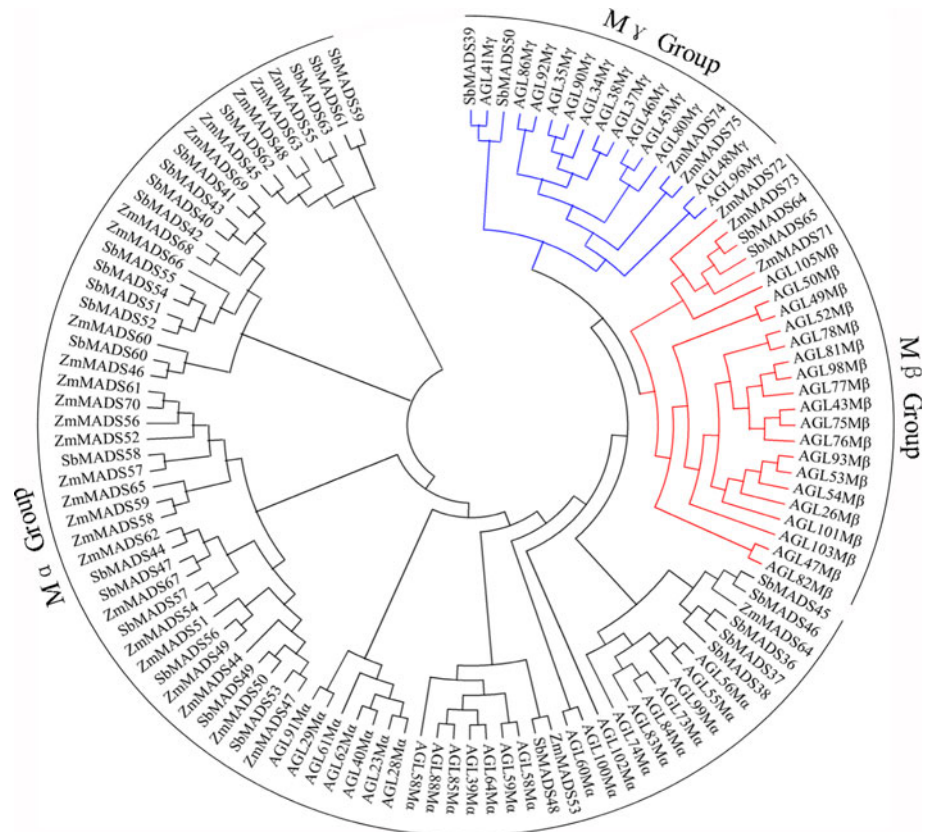
**Fig. 1** Phylogenetic tree of maize, sorghum and *Arabidopsis* type II MADS-box genes. The unrooted tree was generated using the MEGA v4.0 program with the neighbor-joining method. MIKC\*-type MADS-box genes were used as an outgroup. See Table 2 for the maize and sorghum MADS-box gene names



and organs using NCBI EST database. There are 2, 019, 105 maize and 209, 828 sorghum ESTs in the EST data of NCBI. Based on tissue and organ types, we divided the EST data into 12 groups (Fig. 5). In maize, expression of 50 MADS-box genes, including 39 type II and 11 type I

MADS-box genes, was detected from NCBI EST database. We found that most of the maize MADS-box genes have a broad expression spectrum. Interestingly, three genes were found to have specific expression in single tissue and organ. ZmMADS1 was detected to have specific

**Fig. 2** Phylogenetic tree of maize, sorghum and *Arabidopsis* type I MADS-box genes. *Black* (M $\alpha$ ), *Red* (M $\beta$ ), *Blue* (M $\gamma$ ). The unrooted tree was generated using the MEGA v4.0 program with the neighbor-joining method. See Table 2 for the maize and sorghum MADS-box gene names



expression in root, while ZmMADS34 and 60 in endosperm. In addition, total of nine genes (ZmMADS7, 41, 42, 43, 46, 51, 54, 52 and 64) were found in mixed tissue from NCBI EST database.

Using the same method, we also detected the expression pattern of sorghum MADS-box genes from NCBI EST database. 20 of 65 sorghum genes, including 18 type II and 2 type I genes, were shown to express in various tissues and organs. Eleven genes were found to show specific expression in single tissues or organs. For example, SbMADS10, 16, 22, 24, 25 and 57 were expressed only in ovary, suggesting their specific functions in plant growth and development.

More interestingly, type I MADS-box genes with expression data were far less than type II genes in both maize and sorghum. As previously mentioned in this study and other studies, most functionally known MADS-box genes belong to the type II groups, which are the most extensively studied and described transcription factors in plants (Parenicova et al. 2003; Liljegren et al. 2000).

## Discussion

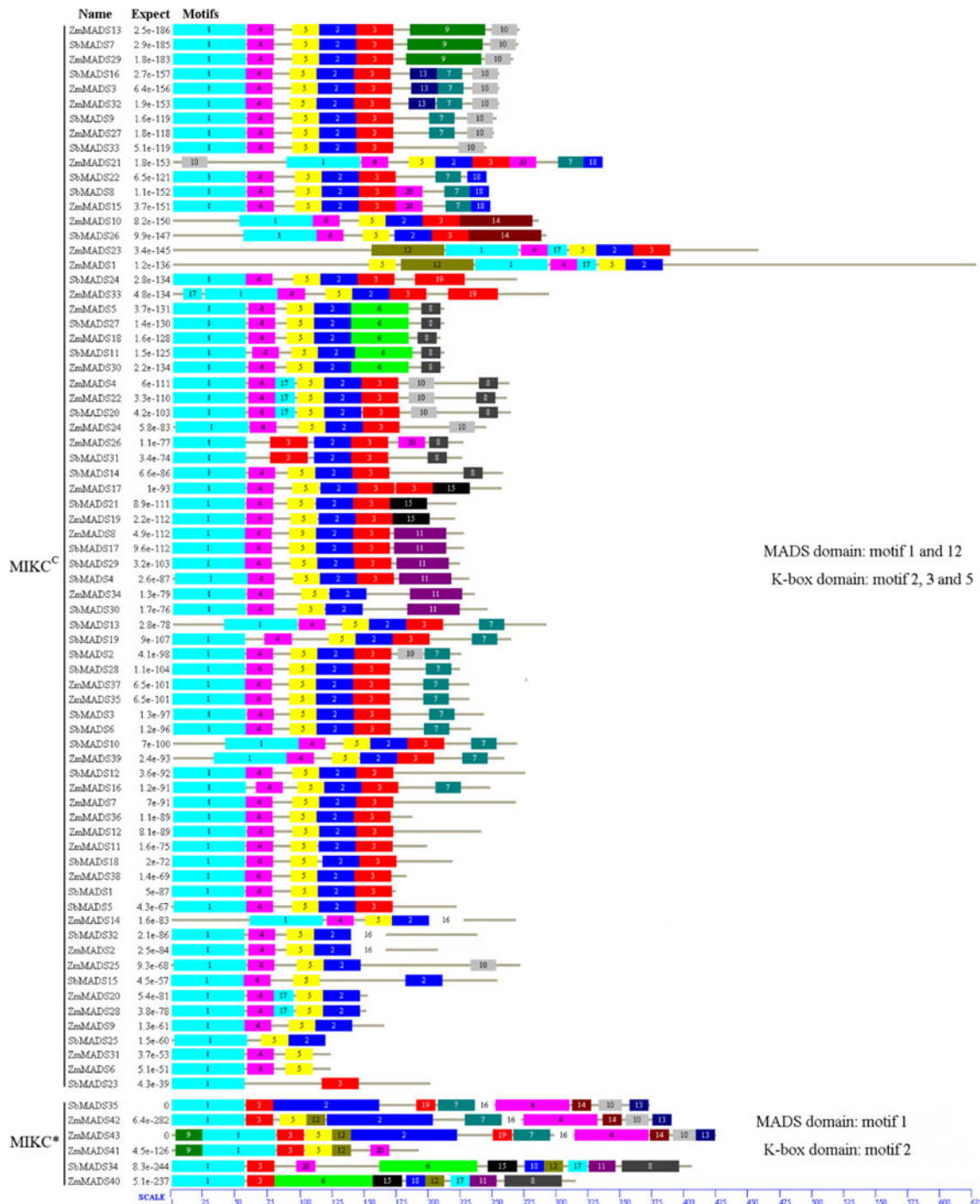
As the key transcription factors in flowering plants, many aspects of the MADS-box genes have been extensively

analyzed in the model plants *Arabidopsis* and rice. The results of present work confirmed many features described in other species. However, some novel characteristics were also found in maize and sorghum compared with *Arabidopsis* and rice.

On the basis of phylogenetic relationships with *Arabidopsis* type II genes (Fig. 1), maize and sorghum type II MADS-box genes can be further divided into 11 subfamilies. Interestingly, no *FLC* homologs were found in the maize or sorghum genomes, such genes have been confirmed to play a major role in controlling flowering time by vernalization and autonomous pathways (Scortecci et al. 2001). This phenomenon was also found in the rice genome (Arora et al. 2007). In addition, we tried to use the coding sequence of *Arabidopsis FLC* genes to search against maize EST database of NCBI but failed. In light of the fact that maize and sorghum do not require vernalization for flowering, we suggested that *FLC*-type genes, which were identified in *Arabidopsis* (Parenicova et al. 2003) have been lost in their genomes.

In addition, *Arabidopsis ANRI* homologs, which play important roles in root development (Zhang and Forde 1998), were also not found in maize. The phenomenon from studies of maize and sorghum phylogeny revealed that MADS-box genes have experienced two rounds of gene loss events, an ancient one before maize-sorghum





**Fig. 3** Twenty putative conserved motifs were identified in the MADS-box family using MEME search tool in the MADS-box family in maize and sorghum. Different motifs are indicated by *different colors*, the names of all members and combined *P* value from

different groups are shown on the *left side* of the figure. The same number in different groups refers to the different motif. The length of motif in each protein represents the actual length and motif sizes are indicated at the *bottom* of the figure

divergence and a recent one after polarization. Previous studies have shown that type I genes have a faster birth and death rate in angiosperms (Nam et al. 2004). However, the number of M $\alpha$ -type genes in both maize and sorghum was higher than those in rice and *Arabidopsis*. We hypothesize

that a fast birth and lower death rate of M $\alpha$ -type genes contributed to this phenomenon.

By analysis of the conserved motifs present in MADS-box proteins, we obtained some essential sources of the evolutionary relationship between maize and sorghum. The

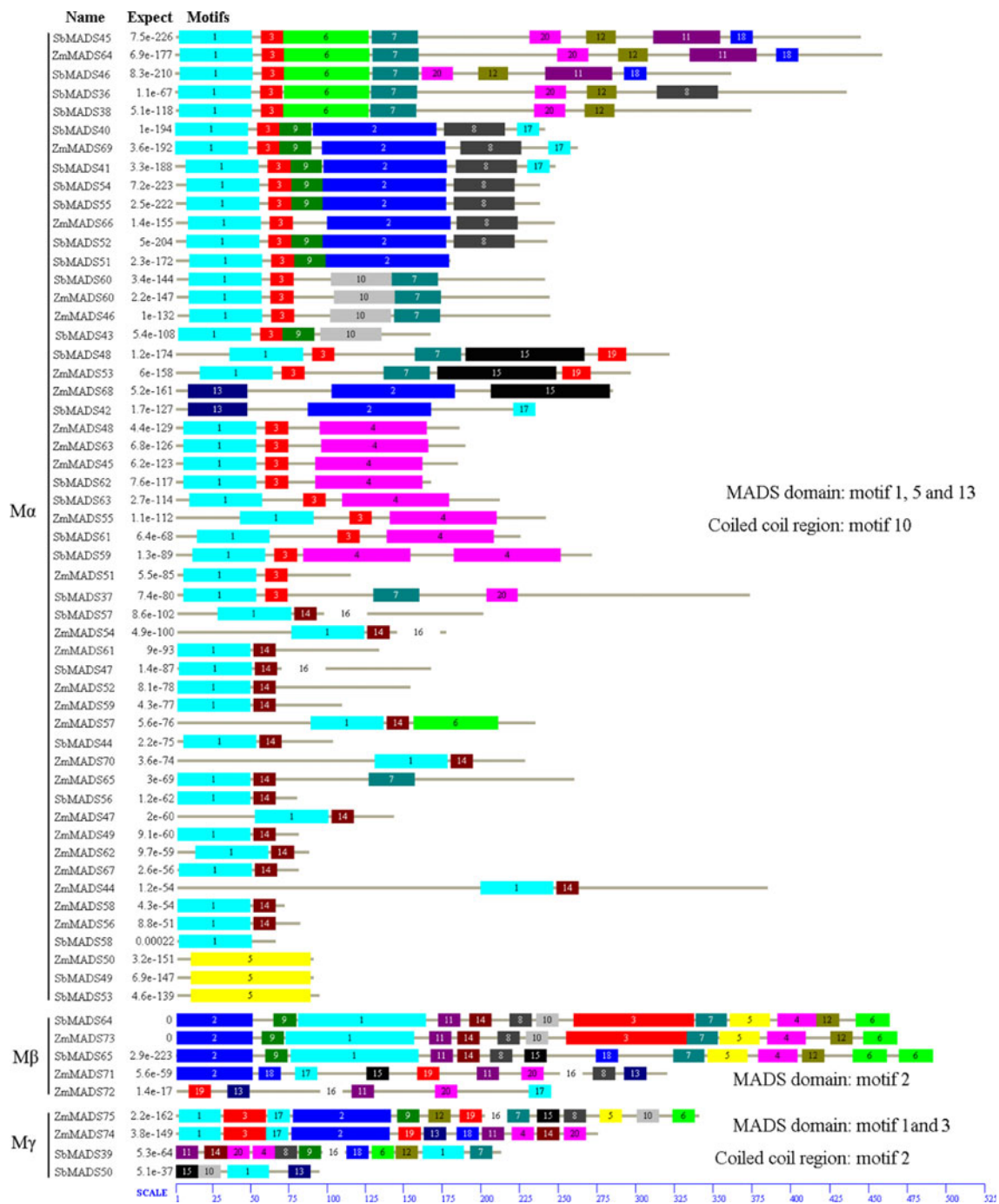


Fig. 3 continued

majority of the members in the same MADS-box subfamily in the phylogenetic trees clearly had similar motifs, which again confirmed their highly sequence similarity again. For example, the *CAL/FUL/API* subfamily contained five maize and four sorghum MADS-box genes, including ZmMADS24, 25, 27, 29, 13 and SbMADS25, 9, 33 and 7. Additionally, all of them shared the common conserved motif 1, 4, 5, 2 and 3 except for ZmMADS25 and

SbMADS25. Particularly, ZmMADS13, 29, and SbMADS7 contained the same motifs, as did ZmMADS27 and SbMADS9 and also SbMADS33 and ZmMADS24. This suggests that these highly conserved motifs play crucial roles in subfamily-specific functions. While the functions of some motifs identified by MEME analysis are unknown (and most of them were also subfamily-specific motifs), we hypothesize that they may also be involved in subfamily-specific

**Table 3** Comparison of MADS-box gene duplications in the four plant genomes

Organization	Maize	Sorghum	Rice <sup>a</sup>	<i>Arabidopsis</i> <sup>b</sup>
Total MADS-box genes	75	65	75	107
Single-genes	13	29	49	39
Multi-genes	62	36	26	68
Multi-genes/single-genes	4.77	1.24	0.53	1.74
Subfamily no.	28	15	12	16
Average members per subfamily	2.21	2.4	2.17	4.25
Maximal members of a subfamily	19	20	5	6

<sup>a</sup> Arora et al. (2007)

<sup>b</sup> Parenicova et al. (2003)

functions, and have important implications for further supporting the subfamily definition. In addition, the MEME analysis also supports the close evolutionary relationship between maize and sorghum, because almost every maize or sorghum protein had one or more orthologous (with similar motifs) in the other genome. Exceptions to this observation were rare.

Although the maize genome is about threefold larger than sorghum (2,300 Mb: 730 Mb), however, in this study, 75 MADS-box genes were identified in the maize genome, which was similar with the number present in sorghum genome. Moreover, the difference in the total number of genes was mainly due to the expansion of type II genes, 43 of which were found in maize and 35 in sorghum. Previous studies show that gene duplications, including tandem and segmental duplications, played a major role in the expansion of the MADS-box gene family (Becker and Theißen 2003; Irish 2003). To investigate the main cause of the possible increase of MADS-box gene number in sorghum, we analyzed both tandem and segmental duplications.

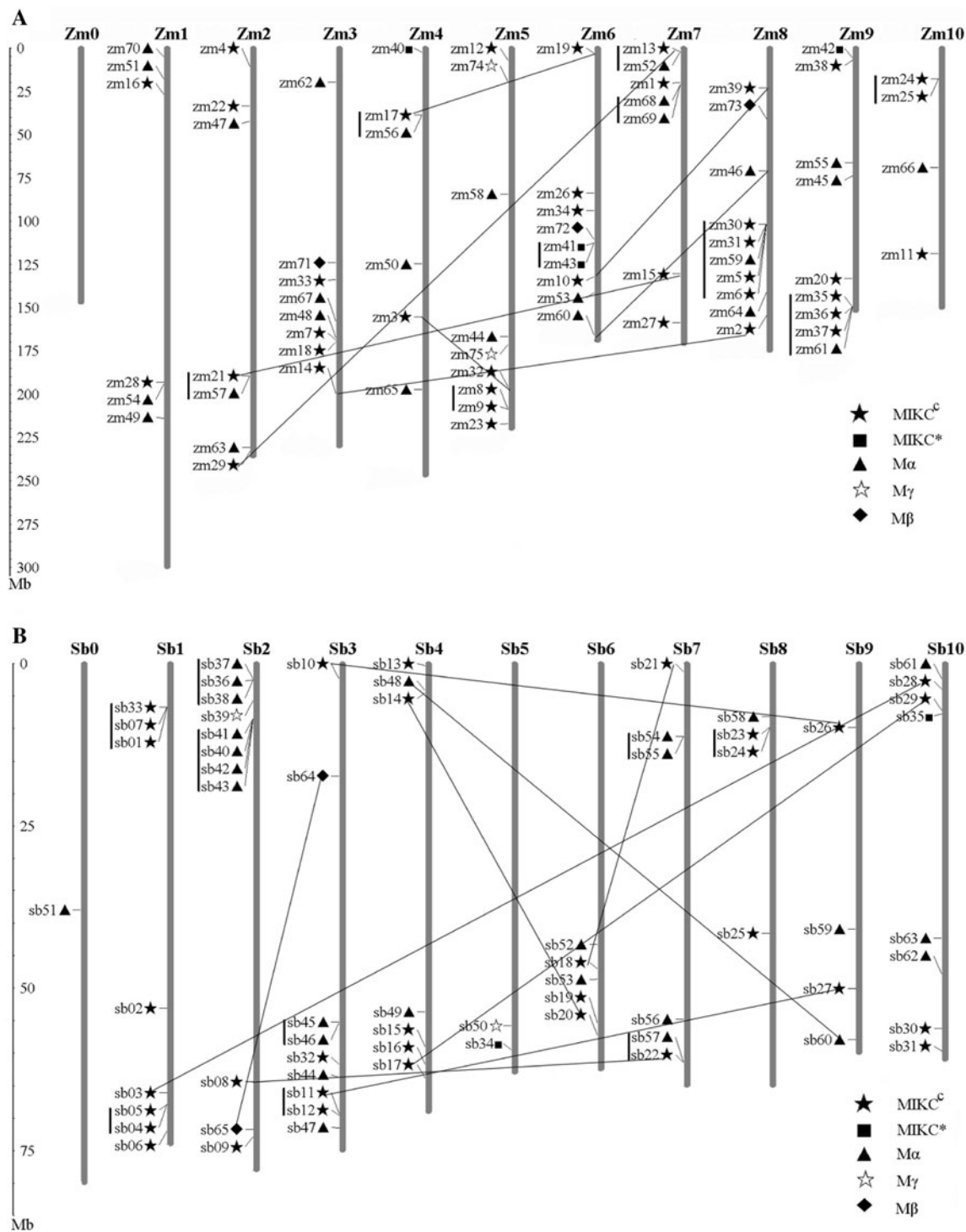
In maize, a total of seven segmental duplication events (14 genes) were found, and eight events (16 genes) presented in sorghum. These results were confirmed by shared phylogenetic clades and the distribution of these MADS-box genes on chromosomes. Obviously, the results also show that the number of MADS-box genes involved in segmental duplication in sorghum (24.6% of the total number of MADS-box genes) is larger than in maize (18.7%). However, about 30.6 and 33.8% of MADS-box genes were clustered on the maize and sorghum chromosomes, respectively. Since MADS-box genes in the same cluster have highly sequence similarity, gene clustering is an outcome of the birth and death process which arises from tandem duplications (Xiong et al. 2005). These results indicate that both tandem and segmental duplications played a major role in MADS-box gene expansion in the maize and sorghum genomes. However, both tandem and

segmental duplications of MADS-box genes in sorghum were more prevalent than in maize.

In *Arabidopsis*, 107 MADS-box genes were identified. According to our previous estimate before this work, the ratio of gene duplications of MADS-box genes in both maize and sorghum should be smaller than in *Arabidopsis*. However, we found that the ratio of of MADS-box gene duplications in *Arabidopsis* was far less than that in maize. This phenomenon was also found when comparing maize, sorghum and rice. Since maize, sorghum and rice are important agricultural crops, the differences in MADS-box gene duplications in all four plant genomes may be due to the artificial selection, which plays a major role in promoting evolution.

In maize, 49.3% of MADS-box genes were located on chromosomes 5, 6, 8 and 9. In sorghum, chromosomes 1, 2, 3 and 10 contained 55.4% of the MADS-box genes, especially the tip of chromosome 2, which contained seven M $\alpha$  and one M $\gamma$  MADS-box genes in a 6 Mb region. We considered this region as a hot spot of MADS-box genes distribution. However, this phenomenon was indistinct in maize. The genes in the sorghum hot spot might originate from tandem duplications, which comprised two gene clusters. In addition, we hypothesize that the fast birth and lower death rate of M $\alpha$ -type genes, led to multiple copies of M $\alpha$ -type genes presented in the hot spot. Interestingly, while most clusters were comprised of the same types of genes, five clusters, containing different types of genes were detected in maize, but only one was found in sorghum. Different types of MADS-box genes located in the same cluster may have originated from ancient duplications (Díaz-Riquelme et al. 2009). Although this phenomenon was rare, it provided informations about the evolution rate and novelty of the maize genome, which has important implications for the evolution of the MADS-box gene family.

Expression patterns of maize and sorghum MADS-box genes were detected using NCBI EST database. The results provided important clues for gene function research. However, this bioinformatic method was only the first step to study the functions of maize and sorghum MADS-box gene family. In this study, no expression data were detected in 25 maize and 45 sorghum MADS-box genes, possibly for the following reasons: First, pseudogenes. During the process of evolution, many genes may lose their functions due to mutations, duplications, indels and other rearrangements (Gojobori et al. 1982; Caetano-Anollés 2001); Second, some genes show temporal and spatial expression pattern. On the other hand, the expression of the MADS-box genes may occur at specific developmental stages or under special conditions. Many transcription factor families have common characters of the temporal and spatial expression pattern, such as rice Aux/IAA genes (Jain and



**Fig. 4** Chromosomal locations of MADS-box genes on maize (a) and sorghum (b) chromosomes. The chromosome numbers are shown at the top of each chromosome (vertical gray bars). The names on the left side of each chromosome correspond to the approximate location of each MADS-box gene (zm: ZmMADS; sb: SbMADS). The markers next to the gene names represent the groups to which each MADS-box gene belongs. The unmapped maize and sorghum

MADS-box genes are shown on Zm0 and Sb0, respectively. The thick black lines on the left side of the names of the MADS-box genes indicate the clusters of tandemly duplicated genes on each chromosome. The thin black lines connect the MADS-box paralogs lying on duplicated chromosomal segments. The images were generated using MapInspect software ([http://www.plantbreeding.wur.nl/uk/software\\_mapinspect.html](http://www.plantbreeding.wur.nl/uk/software_mapinspect.html))

**Fig. 5** The expression profile for maize and sorghum MADS-box genes from NCBI EST database. The *black points* indicate the expression data for maize and sorghum MADS-box genes, and the *blank* shows that no expression could be detected. The names of expressed gene are shown on the *left side* of the figure; genes with no expression data are not shown. The group of “Multiple” indicates the expression data comes from mixed tissues and organs

Gene Name	Root	Leaves	Shoot	Ear	Tassel	Endosperm	Silk	Pollen	Embryo	Callus	Ovary	Multiple
<b>Maize</b>												
ZmMADS1	●											
ZmMADS2				●	●							
ZmMADS3, 32, 38, 39				●								●
ZmMADS4, 20, 22, 23, 28	●											●
ZmMADS5				●		●	●	●				●
ZmMADS6, 30, 31				●		●	●					●
ZmMADS7, 41, 42, 43												●
ZmMADS8			●			●			●			●
ZmMADS9						●			●			●
ZmMADS10				●								●
ZmMADS12, 27				●		●						●
ZmMADS13, 29, 52				●	●							●
ZmMADS14				●	●							●
ZmMADS15				●		●	●		●			●
ZmMADS16						●		●	●			●
ZmMADS18				●	●							●
ZmMADS21				●		●	●	●	●			●
ZmMADS24, 25				●								●
ZmMADS26, 70						●						●
ZmMADS33				●		●			●			●
ZmMADS34, 60						●						●
ZmMADS35, 36, 37		●				●						●
ZmMADS46, 51, 54, 62, 64				●			●		●			●
ZmMADS57		●		●		●			●			●
ZmMADS61		●		●		●			●			●
ZmMADS65	●	●	●	●	●	●						●
<b>Sorghum</b>												
SbMADS2	●		●									
SbMADS3				●							●	
SbMADS4	●	●						●	●			
SbMADS5	●	●							●			
SbMADS7			●								●	
SbMADS8				●							●	
SbMADS9			●					●				
SbMADS10, 16, 22, 24, 25, 57											●	
SbMADS13, 21, 27, 49			●									
SbMADS15								●		●		
SbMADS17	●	●	●					●	●			
SbMADS34								●				

Khurana 2009). Third, due to the limitation of current ESTs of NCBI EST database, especially for sorghum EST database, expression of some genes was not identified in any tissues or organs. MADS-box genes with specific expression can be firstly considered as candidate genes for cloning and further functional analysis, which may play specific functions in specific tissues and organs.

**Acknowledgments** This work was supported by grants from the National High-Tech Research and Development Program (863 Program) (No. 2008AA10Z408) and the National Natural Science Foundation of China (No. 10675002). We thank members of the Key Laboratory of Crop biology of Anhui province for their assistance in this study.

## References

- Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF (2000) MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J* 24:457–466
- Alwee SS, Van der Linden CG, Van der Schoot J, de Folter S, Angenent GC, Cheah SC, Smulders MJM (2006) Characterization of oil palm MADS box genes in relation to the mantled flower abnormality. *Plant Cell Tissue Organ Cult* 85:331–344
- 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: 242–263
- Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* 3:21–29
- 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
- Becker A, Winter KU, Meyer B, Saedler H, Theißen G (2000) MADS-box gene diversity in seed plants 300 million years ago. *Mol Biol Evol* 17:1425–1434
- Caetano-Anollés G (2001) Novel strategies to study the role of mutation and nucleic acid structure in evolution. *Plant Cell Tissue Organ Cult* 67:115–132
- 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
- Davies B, Egea-Cortines M, Silva ED, Saedler H, Sommer H (1996) Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J* 15:4330–4343

- De Bodt S, Raes J, Van de Peer Y, Theißen G (2003) And then there were many: MADS goes genomic. *Trends Plant Sci* 8:475–483
- Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ (2009) Genome-wide analysis of MIKC<sup>C</sup>-type MADS-box genes in grapevine. *Plant Physiol* 149:354–369
- Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34:247–251
- Gojobori T, Li WH, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. *J Mol Evol* 18:360–369
- Gu Q, Ferrándiz C, Yanofsky MF, Martienssen R (1998) The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125:1509–1517
- Gu ZL, Cavalcanti A, Chen FC, Bouman P, Li WH (2002) Extent of gene duplication in the genomes of drosophila, nematode, and yeast. *Mol Biol Evol* 19:256–262
- Hartmann U, Höhmann S, Nettesheim K, Wisman E, Seadler H, Huijser P (2000) Molecular cloning of SVP: a negative regulator of the floral transition in *Arabidopsis*. *Plant J* 21:351–360
- Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theißen G (2002) Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Mol Biol Evol* 19:801–814
- Holub E (2001) Arms race is ancient history in *Arabidopsis*, the wildflower. *Nat Rev Genet* 2:516–527
- Honma T, Goto K (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409:525–529
- Irish VF (2003) The evolution of floral homeotic gene function. *Bioessays* 25:637–646
- Jain M, Khurana JP (2009) Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice. *FEBS J* 276:3148–3162
- Kaufmann K, Melzer R, Theißen G (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347:183–198
- Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M (2003) Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol* 20:1963–1977
- Kramer EM, Irish VF (1999) Evolution of genetic mechanisms controlling petal development. *Nature* 399:144–148
- Leseberg CH, Li A, Kang H, Duvall M, Mao L (2006) Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene* 378:84–94
- Liljgren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404:766–770
- Ma H, Yanofsky MF, Meyerowitz EM (1991) AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Gene* 5:484–495
- Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1. *Nature* 360:273–277
- Mena M, Mandel MA, Lerner DR, Yanofsky MF, Schmidt RJ (1995) A characterization of the MADS-box gene family in maize. *Plant J* 8:845–854
- 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
- Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11:949–956
- Nam J, Kim J, Lee S, An G, Ma H, Nei M (2004) Type I MADS-box genes have experienced faster birth-and-death evolution than typeII MADS-box genes in angiosperms. *Proc Natl Acad Sci USA* 101:1910–1915
- Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L (2002) The TRANSPARENT TESTA16 locus encodes the *Arabidopsis* BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* 14:2463–2479
- Norman C, Runswick M, Pollock R, Treisman R (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55:989–1003
- Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L (2003) Molecular phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15:1538–1551
- Passmore S, Maine GT, Elble R, Christ C, Tye BK (1988) *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. *J Mol Biol* 204:593–606
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Hellsten U, Mitros T, Poliakov P, Schmutz J, Spannag M, Tang HB, Wang XY, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang LF, Carpita NC, Freeling M, Gingle AR, Peterson DG, Mehboob-ur-Rahman WareD, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551–556
- Pellegrini L, Tan S, Richmond TJ (1995) Structure of serum response factor core bound to DNA. *Nature* 376:490–498
- Purugganan MD (1997) The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *J Mol Evol* 45:392–396
- Riechmann JL, Meyerowitz EM (1997) MADS-domain protein in plant development. *Biol Chem* 378:1079–1101
- Riechmann JL, Krizek BA, Meyerowitz EM (1996) Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc Natl Acad Sci USA* 93:4793–4798
- Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G (2000) Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. *Science* 288:1613–1616
- Schnable PS, Ware D, Fulton RS, Wei FS, Pasternak S, Liang CZ, Zhang JW, Clifton SW, Wing RA, Wilson RK et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115
- Scortecchi KC, Michaels SD, Amasino RM (2001) Identification of a MADS-box gene, FLOWERING LOCUS M, that represses flowering. *Plant J* 26:229–236
- Shore P, Sharrocks AD (1995) The MADS-box family of transcription factors. *Eur J Biochem* 229:1–13
- Sommer H, Beltran JP, Huijser P, Pape H, Lönning WE, Saedler H, Schwarz-Sommer Z (1990) Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J* 9:605–613
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA 4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Theißen G, Kim JT, Saedler H (1996) Classification and phylogeny of the MADS-box multigene family suggest defined roles of

- MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J Mol Evol* 43:484–516
- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Munster T, Winter KU, Saedler H (2000) A short history of MADS-box genes in plants. *Plant Mol Biol* 42:115–149
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Xiong YQ, Liu TY, Tian CG, Sun SH, Li JY, Chen MS (2005) Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots. *Plant Mol Biol* 59:191–203
- Yang Y, Fanning L, Jack T (2003) The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, *APETALA3* and *PISTILLATA*. *Plant J* 33:47–59
- Yang SH, Zhang XH, Yue JX, Tian DC, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Genet Genomics* 280:187–198
- Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* 346:35–39
- Zhang H, Forde BG (1998) An *Arabidopsis* MADS-box gene that controls nutrient-induced changes in root architecture. *Science* 279:407–409
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian DC (2004) Genome-wide identification of NBS genes in rice reveals significant expansion of divergent non-TIR NBS genes. *Mol Genet Genomics* 271:402–415