Cloning, expression and characterization of a novel MADS-box gene from kenaf (*Hibiscus cannabinus* L.)

Peng Chen1*, Ru Li2, Jian Liao1, Ruiyang Zhou1*

1 College of Agriculture, Guangxi University, Da-xue Road 100, Nanning 530004, China
2 College of Life Science & Technology, Guangxi University, Da-xue Road 100, Nanning 530004, China
* Author for correspondence (Tel: +86-771-3235612; Fax: +86-771-3235612; E-mail: hustwell@gmail.com)

Key words: Kenaf (*Hibiscus cannabinus* L.); MADS-box; Cytoplasmic male sterility (CMS); Gene cloning

1 SUMMARY

MADS-box genes form a large family of transcription factors and play important roles in flower development and organ differentiation in plants. In this study, a cDNA, encoding a typical MIKC type MADS protein with 245 amino acids was isolated from kenaf flower. Database BLAST search analysis of the full deduced protein sequence yielded high homology with other MADS-box proteins. The sequence was designated as *HcMADS-box* and submitted to the Genbank with accession number of HQ315826. Subsequently, a 2338 bp genomic DNA fragment of *HcMADS-box* gene was isolated by PCR. Compared with its cDNA sequence, 7 introns and 8 exons were found in *HcMADS-box* gene. The nucleotide sequences had no difference between cytoplasmic male sterility (CMS) and maintainer lines. Southern blotting analysis showed it contained one copy in the genome. Phylogenetic analysis suggested the *HcMADS-box* belong to E-type MADS-box gene corresponding to the ABCDE model. The expression pattern of *HcMADS-box* was investigated by RT-PCR. Both in CMS and maintainer lines, the highest level of *HcMADS-box* gene expression was found in floral tissues but very poor in vegetable organs. Further comparing *HcMADS-box* gene expression pattern in CMS and maintainer line, the expression level was found to be reduced in the CMS lines notably.

2 INTRODUCTION

The floral MADS-box genes, named after the archetypal genes (M) MINICHROMOSOME MAINTENANCE 1 from *Saccharomyces cerevisiae*, (A) AGAMOUS from *Arabidopsis thaliana*, (D) DEFICIENS from *Antirrhinum majus* and (S) SERUM RESPONSE FACTOR from *Homo sapiens* share a conserved DNA binding domain (MADS-box). They are essential regulators of the development of the floral meristem and floral organs in plants (*Sommer et al. 1990; Yanofsky et al. 1990; Becker and Theissen 2003; Theissen 2001*). The floral MADS-box genes of *Arabidopsis* have been divided into several major classes (A, B, C, D and E) by phylogenetic analysis (*Becker and Theissen 2003*). The same class genes are usually homologs and are functionally similar (*Gramzow et al. 2010*). In flower development, A-class genes (*AP1, CAL and AP2*) specify the development of sepals; a combination of A-, B- and E-class genes participate in petal
development; B-(AP3 and PI), C- and E-class genes are involved in stamen production; C-(AG) and E-class genes specify carpel formation; D-class genes (SHP1 and SHP2, AGL11 and AGL13) are closely related to C-class genes and control the development of ovules and seeds; and E-class genes (AGL3 and SEP) work with A-, B- and C-class genes and function in different whorls (de Folter et al. 2005; Bemer et al. 2010; Masiero et al. 2011). In high plants, these genes were also called homeotic genes (Munster et al. 1997). The function of these homeotic genes were proved to be conserved among angiosperms (Schwarz-Sommer et al. 2003; Vandenbussche et al. 2003; Kater et al. 2006).

Loss-of-function of any class MADS-box genes may result in the homeotic transformation. For example, in maize, rice and wheat, the class B gene mutants show male sterility due to homeotic conversion of stamens into carpels (Ambrose et al. 2000; Nagasawa et al. 2003; Hama et al. 2004). Furthermore, the variation of MADS-box expression was found to be associated with its cytoplasmic male sterility (CMS) phenotype. Zubko et al (2001) found that a B class AP3-like gene was down-regulated in CMS Nicotiana tabacum. The wheat class-B MADS-box APETALA3 (AP3)-type and SEEDSTICK (STK)-like genes were showed the same expression pattern (Hama et al. 2004; Yamada et al. 2009). In ‘carpeloid' CMS carrot flowers, Linke et al (2003) detected a distinctly reduced expression of DcMADS2 and DcMADS3, homologues of the Antirrhinum genes GLOBOSA and DEFICIENS.

Kenaf (Hibiscus cannabinus L.) is one of the most potential annual crop grown mainly in Africa, Asian and South America. Its fibrous bark has traditionally been used to make ropes, textiles and various types of paper. Being fast growing and multipurpose, kenaf has become one of the important alternative source of fibrous material for the 21st century (Chin and Youisf 2009). Our research team has found and selected several CMS lines from spontaneous mutation kenaf (Zhou et al. 2008). The CMS lines have played a crucial role in utilization the heterosis of kenaf. In the CMS lines, no anthers were observed in male flowers. Our previous work has analyzed the cytological difference between the CMS lines and the maintainer lines in details (Zhu et al. 2007), but little is known about the mechanism on the molecular basis.

In the present study, we report the isolation, expression analysis and characterization of a novel MADS-box gene from kenaf, which may give some clues on understanding the floral development of kenaf.

3 MATERIALS AND METHODS

3.1 Plant materials: The plant material used in this study were P3A (the CMS line) and P3B (the maintainer line), they were the near isogenic lines by backcrossing. The CMS line was natural except it could not form anthers (Fig. 1). Plants were grown in the experimental field at Guangxi University under natural conditions and normal field management.
3.2 RNA isolation and First-Strand cDNA Synthesis: Total RNA was extracted from different tissues (root, stem, leaf and inflorescence) of kenaf with Trizol reagent (Invitrogen, USA) following the manufacturer’s instructions, treated with RNase-free DNase I, the DNase I was then removed according to the instructions. The integrity of each sample was verified by electrophoresis on 1% agarose gels and their concentration was checked using a UV spectrophotometer. Then 2 µg of total RNA was reverse transcribed with M-MLV reverse transcriptase (Promega, Madison, USA) using oligo (dT) 18 primer to obtain the first-strand cDNA.

3.3 cDNA cloning of HcMADS-box gene: According to the homologic sequence in different species, we designed a pair of degenerate primer to clone the MADS-box gene. FP1: 5’-ATGGBGAGAGNNAGAGTGGAGCTG-3’, and RP1: 5’-WTCATWYACNGATGGGTCVTYTGAAG-3’. PCR amplification was conducted in 25 µl of reaction mixture containing 100–150 ng of first-strand cDNA, 0.2 mmol/l of each dNTPs, 0.5 µmol/l of each primer, 2.5 µl 10× PCR buffer, and 1.0 U pfu Taq DNA polymerase (Promega, USA). The reaction mixture was heated to 94 °C for 5 min to denature the template, cycled 33 times at 94 °C for 45 seconds, 54 °C for 50 s and 72 °C for 90 seconds, and then incubated at 72 °C for 10 min. The amplified DNA fragments were cloned into pMD18-T (Takara, Japan) by TA-cloning.

3.4 Genomic DNA cloning of HcMADS-box gene: The cetethyltrimethyl ammonium bromide (CTAB) extraction method was used for genomic DNA extraction. The two primers (FP2: 5’-ATGGGGAGAGGAAGAGTGGAG3’, RP2: 5’-TCCACGGATGGGTCTATTGA-3’) designed according to the cDNA sequence were used to amplify MADS-box genes by PCR using genomic DNA as templates. The method was similar to cDNA cloning except the anneal temperature is 52°C.

3.5 Expression analysis by semi-quantitative RT-PCR: Total RNA was extracted from different tissues (root, stem, leaf and inflorescence) of kenaf CMS and maintainer lines as described before. The synthesized cDNAs were then used as templates to amplify MADS-box genes and the β-actin gene with the following two pair primers, respectively. MADS-box gene RT-F: 5’-AGGTTGCCCTAATCATCTC-3’ and RT-R: 5’-GTCACGGAGAGGTCTATTGA-3’. β-actin gene Actin-F: 5’-CAGGCAGTTCTTTGCTTGT-3’ and Actin-R: 5’-ATCCTCCAATCCAGACACT-3’. RT-PCR reactions were performed at least twice in independent experiments. The numbers of PCR...
cycles were optimized to ensure that the amplification reactions were stopped in the exponential phase of the PCR. Equal volumes of PCR products were separated on 1.0% (w/v) agarose gel. The β-actin gene was used as an internal control. The amplified fragments were predicted about 436bp and 622bp respectively.

3.6 Southern blot analysis: Total volume of 15 µg genomic DNA was digested with three restriction enzymes (Pst I, Hind III and Xho I) at 37°C in a water bath overnight. The resulting genomic DNA fragments were separated on 0.8% agarose gels. After electrophoresis, the agarose gel was soaked in 0.25 M HCl for 10 min and denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 30 min. Then the DNA fragments from the gel were transferred to a nylon membrane by the alkaline transfer method. The transferred DNA was immobilized by UV irradiation. The templates used as probes were prepared by PCR amplification from genomic DNA with gene-specific primers in nonconserved regions. Primer pairs were Sou-F: 5’-CCAAAACGATAAAAAGAACT-3’ and Sou-R: 5’-TCTCCAAGCAATGCCTGATAT-3’. The fragment was predicted about 612bp in length. Then the filters were hybridized with PCR-generated digoxigenin-labelled probe. The hybridized probe DNA was detected using chemiluminescence according to manufacturer-provided protocols (Roche, Germany).

3.7 Sequence comparisons and phylogenetic analysis: The nucleotide and deduced amino acid sequences of kenaf MADS-box gene were carried out with the BLAST network service of NCBI (2011) Among the best BLAST hits, genes for which there were published reports were selected for comparison. Multiple sequence alignment was performed with Clustal W1.81. Phylogenetic and molecular evolutionary analyses were conducted using MEGA4.1 software by the Neighbor-Joining Method with p-distance correction.

4 RESULT

4.1 Cloning and characterization the cDNA of HcMADS-box gene: To isolate MADS-box cDNAs from kenaf, total RNA of various tissues were extracted, reverse-transcripted and PCR amplified using degenerated primers. Both on P3A and P3B, a single band with the size of about 750 bp could be amplified (Fig.2).

![cDNA amplification of HcMADS-box gene from kenaf](image_url)

Fig. 2: cDNA amplification of HcMADS-box gene from kenaf. M: DNA Marker, Lane1: P3B, Lane2: P3A, Lane3: Negative control.
The fragments then were identified by TA cloning and sequencing, and as a result, the two sequences showed 100% identities. The results indicated it comprised a complete open reading frame (ORF) of 735 bp, encoding 245 amino acids with the typical structure of class I plant MADS-box proteins: a highly conserved 60 amino acids MADS-box domain, a conserved intervening (I) domain, a keratin-like (K) sequence and a more variable C-terminal region (Fig. 3 and 4). The putative protein has a molecular mass of 27,861Da and a pI of 9.1.

![Fig.3: HcMADS-box gene and the deduced protein sequence. The MADS, I, K and C regions are indicated.](image)

Database BLAST search analysis of the full deduced protein sequence yielded high homology with other MIKC subfamily MADS-box proteins. The identity to *Gossypium hirsutum* (Genbank Number: ACJ26768), *Carica papaya* (Genbank Number: ACD39984), *Momordica charantia* (Genbank Number: ABE03878), *Vitis vinifera* (Genbank Number: XP_002277624) was 92.7%, 82.3%, 72.5% and 78.3% respectively. While compared using the MADS-box domain, the identity was above 98.3% (Fig.5).
4.2 Isolation of an HcMADS-box genomic DNA fragment: The two primers FP2 and RP2 designed according to the cDNA sequence were used to amplify MADS-box genes by PCR using genomic DNA as templates. A distinctive band was amplified with the length of 2500bp approximately (Fig. 6). Further characterization by DNA sequencing showed the fragment was 2338bp in length. Comparing this DNA fragment with its cDNA sequence, we found HcMADS-box gene comprised of 8 exons and 7 introns (Fig. 7A and B). The MADS-box domain was entirely encoded by exon 1, but the I, K and C domains were encoded by several exons respectively. Southern blot analysis with the non-conservative DNA as a probe showed that HcMADS-box may exist as a single copy in the kenaf genome (Fig. 7C).

Fig. 5: The homology of HcMADS-box and other MADS-box proteins, HQ315826(HcMADS-box Hibiscus cannabinus); AC26768 (MADS-13 Gossypium hirsutum); ACD39984 (MADS3 Carica papaya); ABE03878 (AGL6 protein Momordica charantia); XP_002277624 (MADS-box protein 3 Vitis vinifera)
Fig. 7: Genomic organization of *HcMAD-box* gene and southern blot analysis. (A) Structure of the *HcMAD-box* gene. The boxes indicate exons, and the lines between them represent introns. The exons encoding the protein domains MADS, I, K and C are indicated. (B) The nucleotide length of the exons and introns are indicated. (C) Genomic Southern blot analysis of *HcMAD-box* gene. Total kenaf genomic DNA (15 µg) from maintainer P3B were digested with the indicated restriction enzymes, and submitted to Southern blot analysis with the non-conservative DNA as a probe. B: *BamHI*, E: *EcoRI*, H: *HindIII*, N: *NdeI*.

4.3 Expression patterns of *HcMADS-box*:
To determine the expression profile of *HcMADS-box*, we used semi-quantitive PCR to detect its transcripts in various tissues of kenaf. Both in CMS and its maintainer lines, the highest level of *HcMADS-box* gene expression was found in floral tissues but very poorly in roots, stems and leaves. Further comparing *HcMADS-box* gene expression pattern in CMS and maintainer line, no obvious difference was found in roots, stems and leaves, whereas for floral organ, the expression level was higher in the maintainer notably (Fig.8).

Fig.8: RT-PCR analysis the expression pattern of *HcMADS-box* in different tissue of kenaf. Phylogenetic analysis of *HcMADS-box* gene
The MADS-box genes of *Arabidopsis* were divided into five subfamily, AP1, AP3/PI, AG, AGL2/SEP, and orphans subtribes (Theissen 2001). They each play a distinctive role in plant development corresponding to the ABCDE model. To determine the phylogenetic status of *HcMADS-box* gene in MADS-box gene family, 28 MADS-box genes from different AP1, AP3/PI, AG, AGL2/SEP, and orphans subfamily of *Arabidopsis* were selected for comparison and contrasting a phylogenetic tree. The *HcMADS-box* gene showed a higher homology and classified to AGL2 subfamily (Fig. 9). The results suggested that the newly cloned *HcMADS-box* gene belong to E-type MADS-box gene corresponding to the ABCDE model and may have important function in kenaf floral development.

![Phylogenetic analysis of MADS-box proteins](http://example.com/figure9.png)

**Fig. 9:** Phylogenetic analysis of MADS-box proteins. Based on the amino acid sequence of the full-length protein. The Genbank number for each MADS-box gene was listed behind the protein names. 28 Amino acid sequences of *Arabidopsis thaliana* MADS-box genes were retrieved via the National Center for Biotechnology Information server (http://www.ncbi.nlm.nih.gov/). The *HcMADS-box* protein was indicated by a dark trigon.

5 **DISCUSSION**

In the present study, a full length cDNA sequence was isolated and characterized. Sequence analysis of the deduced protein indicates it has MADS-box typical structures and belongs to the MIKC type MADS-box gene family. The newly cloned *HcMADS-box* gene showed high levels of homology with the genes of *Arabidopsis* which belongs to the AGL2/SEP subfamily of class E MADS-box genes. Phylogenetic analysis also confirmed *HcMADS-box* gene belonged to AGL2/SEP subfamily. To the best of our knowledge, it is the first report cloning and characterization a MADS-box gene from kenaf.

To verify the gene structure of the *HcMADS-box* gene, we further isolated the genomic DNA of *HcMADS-box* gene by PCR. Structure analysis revealed it contained seven introns and eight exons (Fig.7B). The additional intron was positioned between the exons of the K domain. The gene structure was similar to LeMADS6 and OsMADS6 isolated from tomato and rice flower respectively (Li *et al.* 2010; Gaffe *et al.* 2011), but was different to
the major MIK-C-type genes in that they possessed seven introns rather than the typical five or six (Brunner et al. 2000; Johansen et al. 2002; Zhang et al. 2008; Shao et al. 2010). Southern blot analysis with the non-conservative DNA as a probe showed that HcMADS-box may exist as a single copy in the kenaf genome. Further comparison of the cDNA and genomic DNA sequences of HcMADS-box in CMS and maintainer lines, no difference was found. This may imply that the kenaf CMS phenotype was not caused by the nucleotide mutation of HcMADS-box.

Expression analysis showed that HcMADS-box gene is highly expressed in reproductive organs and very lowly in vegetative organs. This is similar to the Arabidopsis thaliana AGL2/SEP-like genes. As in vegetative organs, the expression level showed no difference at different stages (data not show). But comparing the two lines, we found that HcMADS-box gene was down-regulated obviously in CMS lines. The similar results have been reported in many other high plant CMS system, in which their expression pattern of nuclear MADS-box genes, such as AP3-like, GLOBOSA-like, AG-like, SEEDSTICK (STK)-like and DEFICIENS-like were found to be reduced expressed in CMS lines (Linke et al. 2003; Meguro et al. 2003; Murai et al. 2003; Hama et al. 2004; Yamada et al. 2009). In the present study, the alteration of HcMADS-box expression level was close to the kenaf CMS phenotype. This may suggests that HcMADS-box expression may play an important role in pollen development and any alterations can lead to dysfunction in a flower, especially to pollen.

6 ACKNOWLEDGMENTS
This work was supported by the National Natural Science Foundation of China (Grant No.30900912, 31060199), the Important Science and Technology Research Foundation of education ministry of China (Grant No. 209095).

7 REFERENCES


