Bionature, 38(6) 2018 : 366-375 ISSN: 0970-9835 (P), 0974-4282 (O)

In silico ANALYSIS OF FLORAL MADS-BOX GENE IN *Brachypodium distachyon*

ABBAS SAIDI[1*] AND ZAHRA HAJIBARAT[1]
[1]Department of Plant Sciences and Biotechnology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, G.C, Tehran, Iran.
Email: abbas.saidi@gmail.com

Article Information
Reviewers:
(1) Samuel I. Ogenyi, Nnamdi Azikiwe University, Nigeria.
(2) Alexis Valerio, Universidad Nacional Experimental del Táchira, Venezuela.
(3) Raul Antonio Sperotto, Universidade do Vale do Taquari – UNIVATES, Brazil.
(4) S.A.C.N. Perera, University of Peradeniya, Sri Lanka.

Received: 23 February 2018
Accepted: 22 May 2018
Published: 05 January 2019

Original Research Article

ABSTRACT

MADS-box plays a fundamental role in all stages of the processes leading to flower formation and other diverse developmental processes in plants. *Brachypodium distachyon is* an excellent model for bioinformatics analysis in temperate grasses and cereals. In this study, analysis of 43 sequences of MADS-box genes, generated from Grassius database was performed in *Brachypodium*. In order to characterise MADS-box genes involved in plant development and to detect cis-element in MADS-box gene, we analysed conserved motifs, physico-chemical characteristics and the rate of amino acid in *Brachypodium*. MADS-box was distributed on all *Brachypodium* chromosomes, but gene clusters were observed on four chromosomes except chromosome 5. Twenty four types of elements were identified in the promoter regions of MADS-box with the G-box containing the highest number of cis elements involved in MADS-box gene family. Our results suggest that G-box can bind to FLC gene which has a key role in flowering control. Our results should help to provide useful data for further functional studies of MADS-box genes in *Brachypodium.*

Keywords: MADS-box; *Brachypodium distachyon*; bioinformatics; cis element.

INTRODUCTION

MADS-box encodes transcription factors that play vital roles in plant development. Members of MADS-box gene play regulatory roles during growth phases of embryo, root and leaf development [1]. It has been reported that MADS-box is not only involved in flower or fruit development [2] but also can have a variety of functions including stress resistance and organ development [3]. The MADS-box is a conserved sequence motif that encodes the DNA-binding MADS domain [4]. MADS domain have high similarity to *CArG-box* which attach to DNA sequences of high homology to the motif $CC[A/T]_6GG$. Based on their protein domain type, MADS-box gene family contain two kinds of lineages, type I and type II, [4]. The two types of domain proteins are different namely the SRF-like (Type I) MADS-domain proteins and the MEF2-like (after MYOCYTE-ENHANCER-FACTOR2) (Type II) MADS-domain proteins [5]. MADS-box genes are involved in controlling all phases of growth including male and female

gametophytic developments and root, flower and fruit development [5]. The MADS-box gene family in plants is significantly larger than that recognized in fungi or yeast, with over 100 genes detected in genomes of flowering plants [6].

Evolutionary, MADS-box gene family was expanded via duplication events and whole genome duplications (rice) [7]. MADS-box genes subfamilies regulate the vegetative-to-floral transmission, showing the many complexities of the life history approaches in flowering plants ranging from annuals to perennial [8]. *Brachypodium distachyon,* a grass species native to Europe, northern Africa and Asia, belongs to the cereal grain species (i.e., wheat, rice sorghum, maize) [9]. It is considered an experimental model organism for providing an insight to the genetic and molecular biology of temperate grasses. Because of its small genomic size and rapid life cycle, *B. distachyon* is considered a useful tool for genetic researches.

MADS-box genes bioinformatics analysis was performed in Arabidopsis and coffee [10,11]. Classification of genes is essential in order to perform the functional analysis of a gene family [12]. In order to find interactions among related genes involved in flowering, the expression of flowering genes is regulated by regions of DNA sequences known as cis-regulatory elements. The transcription factor binding sites in the cis-regulatory elements are responsible for the regulation of gene transcription [13]. In this study, 43 putative MADS-box genes from the *B. distachyon* genome were analyzed. Also, characterisation of MADS-box family in *B. distachyon* including phylogenetic analysis, a conserved motif, chromosomal positions, and promoter analysis were performed. *This research aimed to identify functional genes, to characterize MADS-box proteins, and to detect cis elements involved in MADS-box gene for flowering.*

MATERIALS AND METHODS

Phylogenetic Analysis and Mapping *BdMADS* of *Brachpodium* Genes

The BdMADS sequences were downloaded from www.grassius.org. *Originally, a higher number of gene sequences were selected but only 43 were used based on the chromosomal location of the gene sequences.* Alignment of the sequences of the MADS-box genes was performed using CLUSTALW program with MEGA software version 6 [14]. The phylogenetic tree was constructed using the Neighbour-Joining (NJ) method. Diagrams of phylogenetic trees were drawn with MEGA6 (http://www.megasoftware.net/mega.php) (Fig. 1). Also, verification and test of the phylogeny were performed using the bootstrap method with 1000 replicates. The MapChart software version 2.3 [15] was downloaded from https://www.wageningenur.nl/en/show/Mapchart-2.30.html to show chromosomal locations of *B. distachyon* MADS-box genes (Fig. 2).

Analysis of Gene Structure and Conserved Motifs

In order to identify MADS-box protein motifs, MEME database was used in *Brachypodium.* This database was utilized to predict conserved motifs, a number of repeated motifs, maximum motif, and range of motifs (Fig. 3). SMART software (Simple Motif Architecture Research Tool, http://smart.embl-heidelberg.de/) version 5.0 [16] was used to confirm motifs in MEME. In this study, physico-chemical characteristics

such as number of amino acids, molecular weight, and an isoelectric point in *Brachypodium* were studied using ProtParam tool at ExPASy (website http://web.expasy.org/protparam).

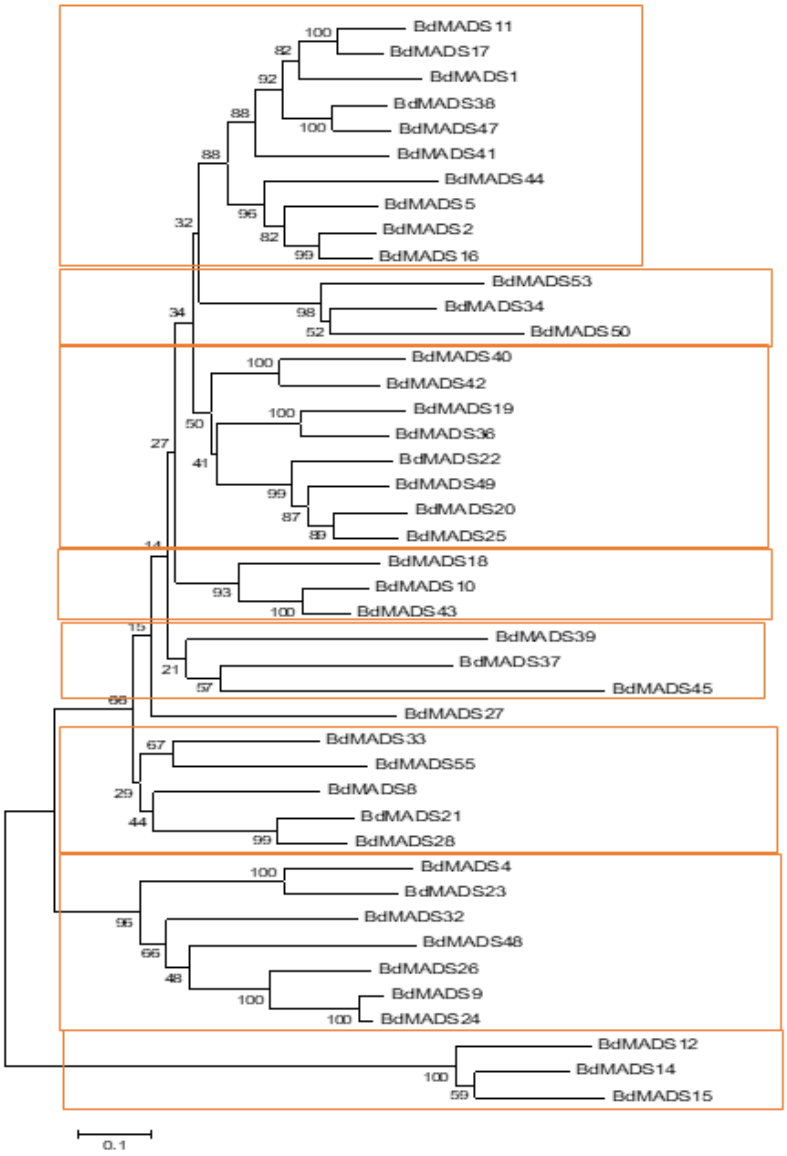

Fig. 1. Phylogenetic tree of *B. distachyon* BdMADS. The phylogenetic tree was drawn using the Neighbor-Joining (NJ) method with 1000 bootstrap replicates in MEGA 6.0 software

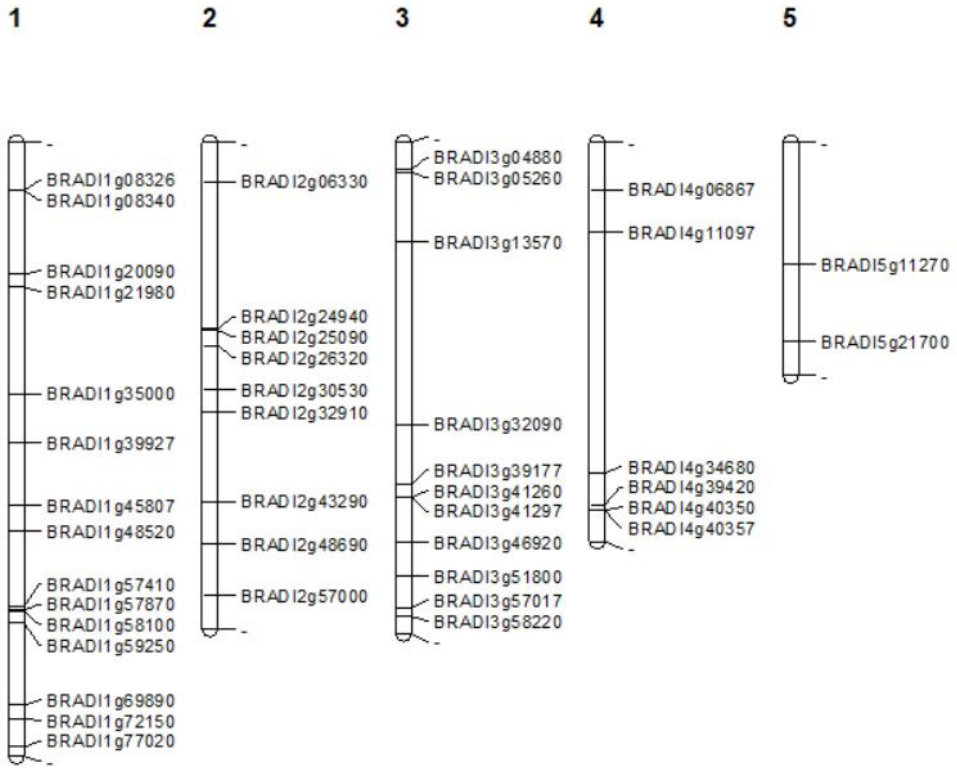

Fig. 2. Chromosomal locations of *B. distachyon*'s *BdMADS* genes. The chromosome numbers were shown at the top of each chromosome. The location of each MADS-box gene was indicated by a line.

Promoter Analysis

Sequence analysis of the 1500bp upstream of the transcription start site (TSS) was performed using PlantCARE [17] database. The analysis allowed the identification of cis-regulatory elements including Transcription Factor Binding Sites (TFBS) (Fig. 4).

RESULTS AND DISCUSSION

Phylogenetic Analysis and Chromosomal Distribution of MADS-box Family

Multiple alignment analysis of MADS-box domains and phylogenetic tree construction were performed. The MADS-box domain phylogenetic tree can be divided into five groups: Ia, Ib, IIa, IIb, III, and IV. Physical locations of the MADS-box genes on the five chromosomes of *Brachpodium* were mapped. The highest numbers of MADS-box genes were found on the chromosome 1 while chromosome 5 contained the lowest. Fifteen MADS were located on chromosome 1, 10 on chromosomes 2, 11 on chromosomes 3, 6 on chromosome 4 , and only two genes on chromosomes 5 (Fig. 2). All chromosomes contained MADS encoding sequences. The clustering patterns of these sequences were compared with their placement on each

chromosome, confirming the presence of the sequences of the same chromosomal origin in the same phylogeny [18]. The researcher suggested that a gene cluster is defined as a chromosome region with two or more genes located within 200 kb sequence [19]. Accordingly, based on the Holub's criterion, we found four MADS gene clusters containing a total of eight genes [19]. Only one gene cluster was found on each of the chromosomes 1, 2, 3 and 4. No cluster was found on chromosome 5. Gene cluster analysis showed that the number of genes on chromosomes was positively correlated with the number of gene clusters (Pearson correlation = 0.882, P-Value = 0.001). Duplicated genes and gene cluster have similarly appeared in *Brachpodium*, *Coffea Arabica*, and banana [10,20,12].

Protein Properties and Sequence Analysis of MADS-box Genes

In this study, a set of MADS-box genes was studied. The ExPASy proteomics server (http://expasy.org/) was used to detect physico-chemical characteristics. Results showed that MADS-box genes were different for the number of amino acids, molecular weight, and isoelectric point (Table 1). Protein length in this gene family ranged from 183 to 603 bp, and molecular weight ranged from 20669.32 to 48315.86 Kd. Highest and lowest isoelectric point was 5.1 (*BdMADS26*) to 9.67 (*BdMADS48*), respectively. Using the MEME program (http://meme.nbcr.net/meme/cgi-bin/meme.cgi), the conserved motifs in the MADS proteins were detected with the following factors: maximum and minimum number of motifs was 10 and 3, respectively, with an optimum motifs width of 50. Subsequently, all identified motifs were annotated with the aid of SMART (http://smart.embl-heidelberg.de/). Our findings showed that the conserved motifs patterns identified in our analyzed genes belonged to the same group, as also reported by other research results [21,20]. Conserved MADS-box motifs have also been detected in other plants such as grapevine and rice [22,23]. Therefore, proteins containing lower domain complexity are also simpler in function.

Cis-regulatory Element Analysis

The analysis of cis-elements showed that twenty-four responsive elements were identified in the promoter regions. A survey on regulatory cis-elements makes it possible to understand gene expression under different conditions better. These elements were divided into several groups: essential elements, responsive to light, expression of a specific tissue, abiotic stresses, hormonal and other types. The largest group of cis elements activated in response to light included ACE, G-Box, Sp1, Box 4, GATA-motif, TCCC-motif, AE box, GT1-motif, and Box I. Responsive elements to hormones were the second largest group involved in hormonal response which included TGA-element, TCA-element, GARE-motif, TGACG-motif, CGTCA-motif, motif IIb, and ABRE (Fig. 4). Stress-responsive elements included CCAAT-box, TC-rich repeat, HSE, and MBS (Liu *et al.,* 2017). Core promoter elements contained CAAT-box and TATA-box [17]. Motifs involved in cellular development included O2-site, Skn-1 motif, CAT-box, and Circadian (Liu *et al.,* 2017). Other motifs involved in tissue-specific meristem included as-2-box, CCGTCC-box, and A-box.

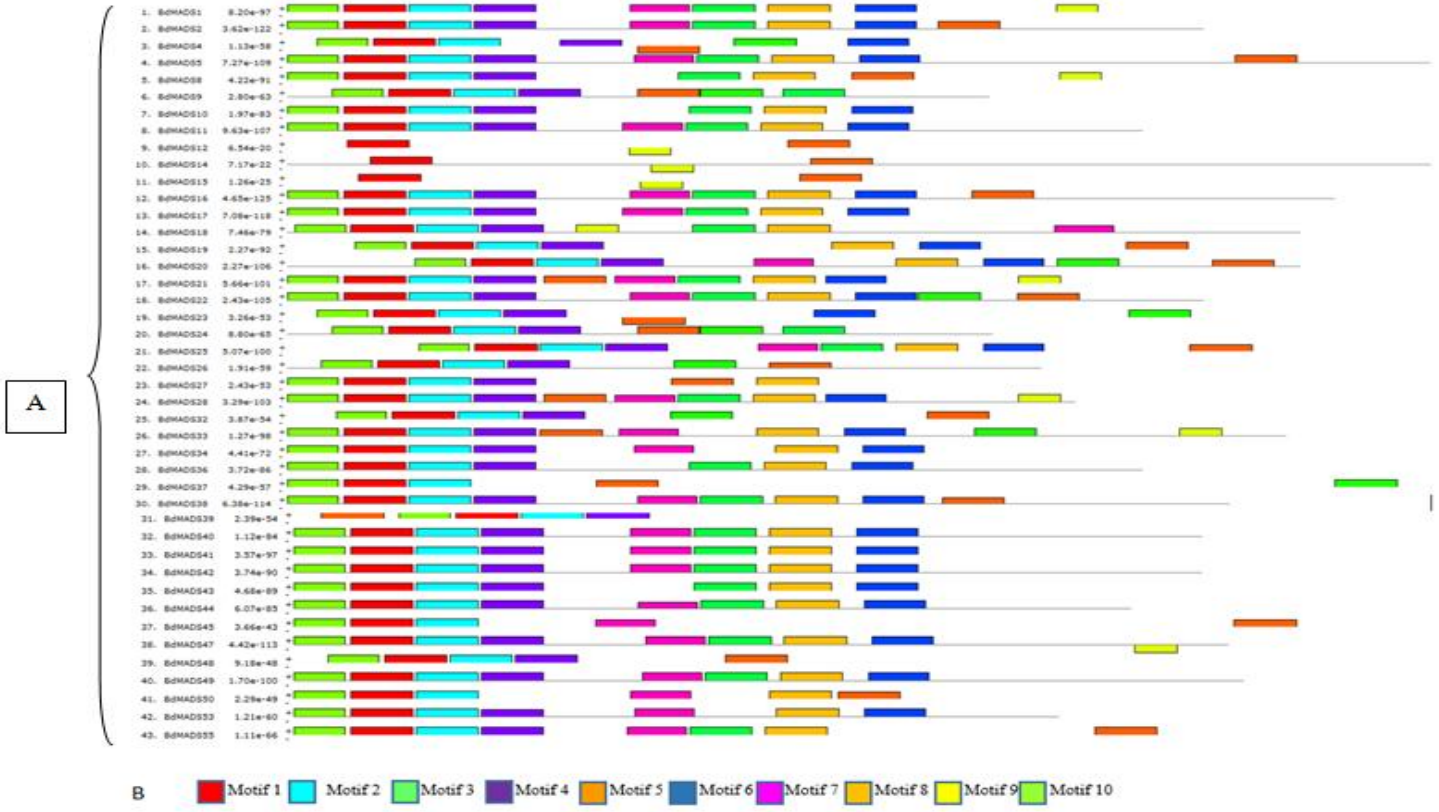

Fig. 3. Schematic representation of motifs identified in *B. distachyon* MADS-box proteins using MEME. Different motifs are indicated by different colors. (A) Distribution of conserved motifs of BdMADS from different groups. (B) The motifs, numbered 1–10, were displayed in different colored boxes

Table 1. Characteristics of MADS-box gene family in *B. distachyon*

Protein name	Gene locus	Number of amino acid	Molecular weight	Theoretical pI	Protein name	Gene locus	Number of amino acid	Molecular weight	Theoretical pI
BdMADS1	Bradi1g08326	233	26565.43	7.69	*BdMADS27*	Bradi2g48690	196	22381.65	8.6
BdMADS2	Bradi1g08340	243	27900.7	9.17	*BdMADS28*	Bradi2g57000	209	24112.64	7.81
BdMADS4	Bradi1g20090	283	30315.37	6.36	*BdMADS32*	Bradi3g04880	365	38244.93	5.58
BdMADS5	Bradi1g21980	309	35159.5	8.49	*BdMADS33*	Bradi3g05260	265	29010.7	6.61
BdMADS8	Bradi1g35000	230	26034.57	8.95	*BdMADS34*	Bradi3g13570	232	26121.02	6.18
BdMADS9	Bradi1g39927	186	20283.19	9.45	*BdMADS36*	Bradi3g32090	227	25690.51	9.21
BdMADS10	Bradi1g45807	224	25047.15	5.92	*BdMADS37*	Bradi3g39177	330	36020.52	5.24
BdMADS11	Bradi1g48520	227	26133.54	7.65	*BdMADS38*	Bradi3g41260	250	28915.86	8.74
BdMADS12	Bradi1g57410	373	40667.46	4.62	*BdMADS39*	Bradi3g41297	203	23543.49	5.54
BdMADS14	Bradi1g57870	359	39885.58	4.88	*BdMADS40*	Bradi3g46920	240	27494.11	7.79
BdMADS15	Bradi1g58100	183	20669.32	5.85	*BdMADS41*	Bradi3g51800	268	30255.26	8.73
BdMADS16	Bradi1g59250	278	31748.88	9.24	*BdMADS42*	Bradi3g57017	240	27334.18	8.88
BdMADS17	Bradi1g69890	253	29191.92	7.6	*BdMADS43*	Bradi3g58220	229	25601.83	5.9
BdMADS18	Bradi1g72150	269	30248.29	5.82	*BdMADS44*	Bradi4g06867	221	24727.12	9.25
BdMADS19	Bradi1g77020	258	29272.34	8.82	*BdMADS45*	Bradi4g11097	372	41623.84	6.24
BdMADS20	Bradi2g06330	269	30317.06	9.25	*BdMADS47*	Bradi4g34680	247	28368.23	8.87
BdMADS21	Bradi2g24940	209	24434.97	9	*BdMADS48*	Bradi4g39420	258	27844.17	9.67
BdMADS22	Bradi2g25090	243	27455.93	9.1	*BdMADS49*	Bradi4g40350	251	28004	9.31
BdMADS23	Bradi2g26320	314	33279.51	9.18	*BdMADS50*	Bradi4g40357	188	21330.62	9.51
BdMADS24	Bradi2g30530	187	20300.33	9.65	*BdMADS53*	Bradi5g11270	202	23289.87	7.78
BdMADS25	Bradi2g32910	267	30521.34	9.06	*BdMADS55*	Bradi5g21700	240	27931.18	6.53
BdMADS26	Bradi2g43290	603	48315.86	5.1					

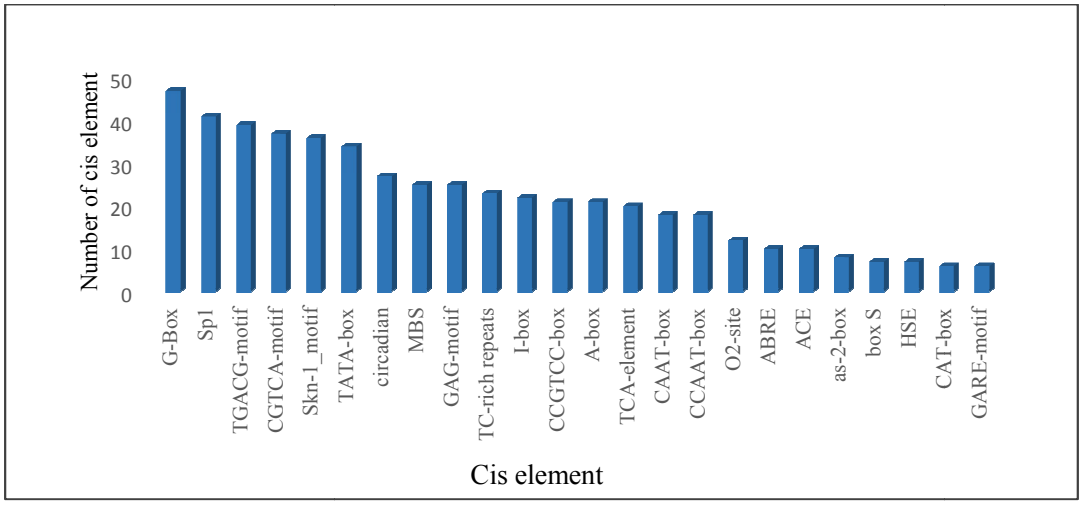

Fig. 4. Cis regulatory elements detected in the region of the prefix genes and their frequency in each gene, the 1000-bp upper region of the upper hand codon of each gene were searched using the PlantCARE database.

G-box had the highest number of cis-elements involved in MADS-box gene family. G-box (CACGTG) element is involved in many reactions such as response to light, abscisic acid, methyl-jasmonate, and anaerobiosis. In addition, it has a role in ethylene induction as well as in seed-specific expression independently of other regulatory sequences [24]. G-box, also known as ABRE (ABA-responsive element), is a binding site for bZIP proteins which have been identified as binding proteins [24]. Basic leucine zipper domain (bZIP transcription factor), bHLH, and G-box increase the possibility of being co-regulated by other transcription factors when combined with FLC in different forms [25]. The key roles of *FLC are* in flowering control, the timing of vernalization, and reproductive transition [26]. The G-box is necessary for transcriptional activation by a phenyl- propanoid-pathway intermediate. Sequences related to the G-box has been found in promoters of various flavonoid biosynthetic genes [27].

CONCLUSION

In the present study, we carried out an *in silico* survey of *B. distachyon* and characterised MADS-box on the basis of phylogenetic relationships, conserved protein motifs, chromosomal locations, gene duplications, and promoter regions. We used *Brachpodium* MADS gene sequences and applied various bioinformatics software to find information about the biological and physiological processes. Our findings indicate that the association of floral MADS proteins into higher-order MADS complexes might be the principal mode of combinatorial control for floral-organ specification.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS,

Vergara-Silva F, Yanofsky MF. MADS-box gene evolution beyond flowers: Expression in pollen, endosperm, guard cells, roots and trichomes. The Plant Journal. 2000; 24(4):457-466.
2. Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theißen G. Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. Molecular Biology and Evolution. 2002;19(6):801-814.
3. Causier B, Kieffer M, Davies B. MADS-box genes reach maturity. Science. 2002;296(5566):275-276.
4. Krizek BA, Fletcher JC. Molecular mechanisms of flower development: An armchair guide. Nature Reviews Genetics. 2005;6(9):688.
5. Gramzow L, Theissen G. A hitchhiker's guide to the MADS world of plants. Genome Biology. 2010; 11(6):214.
6. De Bodt S, Raes J, Florquin K, Rombauts S, Rouzé P, Theißen G, Van de Peer Y. Genome-wide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. Journal of Molecular Evolution. 2003;56(5):573-586.
7. Yang Y, He M, Zhu Z, Li S, Xu Y, Zhang C, Wang Y. Identification of the dehydrin gene family from grapevine species and analysis of their responsiveness to various forms of abiotic and biotic stress. BMC Plant Biology. 2012;12(1):140.
8. Li Z, Reighard GL, Abbott AG, Bielenberg DG. Dormancy-associated MADS genes from the EVG locus of peach [*Prunus persica* (L.) Batsch] have distinct seasonal and photo-periodic expression patterns. Journal of Experimental Botany. 2009;60(12): 3521-3530.
9. Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, Tice H. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature. 2010;463(7282):763.
10. De Oliveira RR, Chalfun-Junior A, Paiva LV, Andrade AC. In silico and quantitative analyses of MADS-Box genes in *Coffea arabica*. Plant Molecular Biology Reporter. 2010; 28(3):460-472.
11. Raes J. Power in numbers*: In silico* analysis of multigene families in *Arabidopsis thaliana* (Doctoral dissertation, Ghent University); 2003.
12. Wei B, Zhang RZ, Guo JJ, Liu DM, Li AL, Fan RC, Zhang XQ. Genome-wide analysis of the MADS-box gene family in *Brachypodium distachyon*. Plos One. 2014;9(1):e84781.
13. Wittkopp PJ, Kalay G. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics. 2012;13(1):59.
14. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution. 2013;30(12): 2725-2729.
15. Voorrips RE. Map Chart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity. 2002;93(1):77-78.
16. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Bork P. Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Research. 2002;30(1):242-244.
17. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rombauts S. Plant care, a database

of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research. 2002;30(1): 325-327.
18. Liu J, Zhang J, Zhang J, Miao H, Wang J, Gao P, Jin Z. Genome-wide analysis of banana MADS-box family closely related to fruit development and ripening. Scientific Reports. 2017; 7(1):3467.
19. Holub EB. The arms race is ancient history in Arabidopsis, the wildflower. Nature Reviews Genetics. 2001;2(7): 516-527.
20. Liu L, Zhang X, Chen F, Mahi AAE, Wu X, Chen Q, Fu YF. Analysis of promoter activity reveals that GmFTL2 expression differs from that of the known Flowering Locus T genes in soybean. The Crop Journal. 2017; 5(5):438-448.
21. de Silva WSI, Perera MMN, Perera KLNS, Wickramasuriya AM, Jayasekera GAU. *In silico* analysis of osr40c1 promoter sequence isolated from indica variety pokkali. Rice Science. 2017;24(4):228-234.
22. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S. MADS-box gene family in rice: genome-wide identification, organisation and expression profiling during reproductive development and stress. Bmc Genomics. 2007;8(1):242.
23. Grimplet J, Martínez-Zapater JM, Carmona MJ. Structural and functional annotation of the MADS-box trans-cription factor family in grapevine. BMC Genomics. 2016;17:80.
24. Hatton D, Sablowski R, Yung MH, Smith C, Schuch W, Bevan M. Two classes of cis sequences contribute to tissue-specific expression of a PAL2 promoter in transgenic tobacco. The Plant Journal. 1995;7(6):859-876.
25. Menkens AE, Cashmore AR. Isolation and characterization of a fourth *Arabidopsis thaliana* G-box-binding factor, which has similarities to Fos oncoprotein. Proceedings of the National Academy of Sciences. 1994;91(7):2522-2526.
26. Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES. The FLF MADS box gene: A repressor of flowering in Arabidopsis regulated by vernalization and methylation. The Plant Cell. 1999; 11(3):445-458.
27. Sibéril Y, Doireau P, Gantet P. Plant bZIP G-box binding factors. The FEBS Journal. 2001;268(22):5655-5666.