Tracing the footprints of the ABCDE model of flowering in *Phalaenopsis equestris* (Schauer) Rchb.f. (Orchidaceae)

Himani · Thakku R. Ramkumar · Shivi Tyagi · Himanshu Sharma · Santosh K. Upadhyay · Jaspreet K. Sembhi

Received: 13 August 2019 / Revised: 23 December 2019 / Accepted: 23 December 2019
© Korean Society for Plant Biotechnology

Abstract Orchids are indispensable to the floriculture industry due to their unique floral organization. The flowers have two outer whorls of tepals including a lip (labellum), and two inner whorls, pollinia and gynostemium (column). The floral organization and development is controlled at the molecular level, mainly by the *MADS-box* gene family, comprising homeotic genes divided into type I and type II groups. The type I group has four sub-groups, $M\alpha$, $M\beta$, $M\gamma$, and $M\delta$, playing roles in seed, embryo, and female reproductive organ development; the type II group genes form classes A, B, C, D, and E, which are a part of the $MIKC^C$ subgroup with specific roles in florigenesis and organization. The coordinated functioning of these classes regulates the development of various floral whorls. The availability of genome and transcriptome sequence data for *Phalaenopsis equestris* offers an opportunity to validate the ABCDE model of flower development. Hence, this study sought to characterize the *MADS-box* gene family and elucidate of the ABCDE model. A total of 48 identified *MADS-box* proteins, including 20 type I [$M\alpha$ (12), $M\gamma$ (8)] and 28 type II [$MIKC^C$ (27), $MIKC^*$ (1)] members, were characterized for physico-chemical features and domains and motifs organization. The exon-intron distribution and the upstream *cis*-regulatory elements in the promoter regions of *MADS-box* genes were also analysed. The discrete pace of duplication events in type I and type II genes suggested differential evolutionary constraints between groups. The correlation of

spatio-temporal expression pattern with the presence of specific *cis*-regulatory elements and putative protein–protein interaction within the different classes of *MADS-box* gene family endorse the ABCDE model of floral development.

Keywords *MADS-box* genes, *Phalaenopsis equestris*, Characterization, Expression, Flowering, ABCDE model

Introduction

The *MADS-box* genes family encode transcription factors that partake in numerous developmental processes and signal transductions (Riechmann and Meyerowitz 1997; Shore and Sharrocks 1995; Theissen et al. 1996, 2000). The name *MADS-box* is derived from four different proteins from taxonomically diverge species, namely MINI CHROMOSOME MAINTENANCE 1 (M) of *Saccharomyces cerevisiae* (Passmore et al. 1988), AGAMOUS (A) of *Arabidopsis thaliana* (Yanofsky et al. 1990), DEFICIENS (D) of *Antirrhinum majus* (Sommer et al. 1990) and SERUM RESPONSIVE FACTOR (S) of *Homo sapiens* (Norman et al. 1988). The pivotal role of *MADS-box* genes in florigenesis was initially established upon studying the genetic mutants in *A. thaliana* and *A. majus* (Coen and Meyerowitz 1991). Originally, the members of *MADS-box* gene family were functionally characterised into three classes (A, B and C) and this laid the foundation for development of the ABC model of floral development (Coen and Meyerowitz 1991). The whorl 1 (calyx with sepals) is controlled by class A genes (*APETALA1*; *API*), while the whorl 2 (corolla with petals) development is controlled by a combinatorial effect of class A (*APETALA3*; *AP3*) and class B (*PISTILLATA*; *PI*) genes. The class B and class C (*AGAMOUS*; *AG*) genes regulate the development of the whorl 3 (androecium with stamens), and the interior most whorl (gynoecium with carpels) development is controlled by C class genes (Bowman

Himani · T. R. Ramkumar · S. Tyagi · H. Sharma ·
S. K. Upadhyay · J. K. Sembhi (✉)
Department of Botany, Panjab University, Chandigarh, India
e-mail: jaspreet.sembhi@pu.ac.in

T. R. Ramkumar
Agronomy department, IFAS, University of Florida, Gainesville,
FL, 32611, USA

H. Sharma
IKG Punjab Technical University, Jalandhar

et al. 1991; Coen and Meyerowitz 1991; Meyerowitz et al. 1991). The ABC model was later extended by including class D and E genes. The D-class genes (*SEEDSTICK*, *STK* and *SHATTERPROOF*, *SHP*) are shown to be involved in ovule formation in *Petunia* (Angenent and Colombo 1996) and the E-class genes (*SEPALLATA*: *SEP1/2/3/4*) identified in *A. thaliana* with roles in coordinated functioning of all the other classes genes, leading to the synchronised development of floral whorls (Pelaz et al. 2000). Thus, the ‘ABC model’ was revised to ‘ABCDE model’ (Theissen 2001).

The genomic organisations, evolutionary lineages, developmental functions and their functional redundancy led to the division of *MADS-box* genes into two major groups, as type I and type II (Alvarez-Buylla et al. 2000a). Parenicova et al. (2003) classified the type I *MADS-box* genes further into four sub-groups, M α , M β , M γ and M δ , based on their phylogeny and conserved nature of MADS-box domain. The type I genes are involved predominantly in female gametophyte, embryo and seed development (Masiero et al. 2011). Proteins of type II *MADS-box* genes carry characteristic domains, MADS-box (M), intervening (I), keratin (K) and C-terminus (C), and hence often termed as MIKC-type genes. The M domain is highly conserved followed by K domain (Davies et al. 1996; Fischer et al. 1995; Ma et al. 1991; Pnueli et al. 1991) whereas, the I and C domains are least conserved (Parenicova et al. 2003). The K domain is involved in protein-protein interaction, while I domain is responsible for DNA binding and dimerization of proteins (Riechmann et al. 1996). The variable C domain is considered to play a role in transcription initiation (Davies et al. 1996; Kramer et al. 1998; Riechmann et al. 1996). The highly conserved nature of MADS-box and K domain indicates that both domains were subjected to very strong constraints, structurally as well as functionally in comparison to the less conserved I and C domains (De Bodt et al. 2003). The type II (MIKC) genes are further divided into two sub-groups MIKC^C and MIKC* based on their protein structural organization (Henschel et al. 2002). MIKC* proteins consist of longer I domain and less conserved K domain than MIKC^C proteins (Adamczyk and Fernandez 2009; Henschel et al. 2002). The MIKC^C have been further divided phylogenetically into multiple sub-classes, 12 in *A. thaliana* (Parenicova et al. 2003), 13 in grape (Diaz-Riquelme et al. 2009) and 14 in *Oryza sativa* (Arora et al. 2007). These MIKC^C proteins are reported to be involved in multiple developmental processes, especially in floral organ and fruit development (Alvarez-Buylla et al. 2000b; Kaufmann et al. 2005),

whereas the MIKC* genes essentially participate in male gametophyte development (Zobell et al. 2010). Considering the genomic organization, the type I genes are either intron-less or mono-intronic, while the type II genes have multiple introns (Parenicova et al. 2003).

The characteristic N-terminus MADS-box domain (~60 aa) binds to the DNA element CArG box [CC(A/T)₆GG] and forms either homodimer or heterodimer (Riechmann and Meyerowitz 1997). The dimerization is necessary for the formation of floral quartet complexes (AP1/AP1/SEP/SEP, AP1/SEP/AP3/PI, AG/SEP/AP3/PI and AG/AG/SEP/SEP) that induces floral organ formation (Theissen and Saedler 2001). The MIKC proteins form dimer preferentially with other type II members, and scarcely with type I members (de Folter et al. 2005). Similarly, the type I MADS proteins preferentially associate with other type I members. Within the type I proteins, M α , M β , and M γ types form heterodimer with each other. The M α interacts predominantly with M β and M γ and rarely with itself. This indicates that type I protein complexes might get stabilized by M α and establish as higher order complexes (Immink et al. 2009). Masiero et al. (2011) have proposed that type I MADS-box factors probably do not bind to DNA.

MADS-box gene family has been characterised in several plant species across the plant kingdom from algae to angiosperms. Recent whole genome sequencing and global transcriptome profiling of floriculturally important orchids have enabled genome-wide characterization and expression profiling of important gene families. Genome-wide identification of *MADS-box* gene has been reported in orchids like *Apostasia shenzhenica* (Zhang et al. 2017), *Dendrobium catenatum* (Zhang et al. 2016), *Dendrobium officinale* (Yan et al. 2015) and *P. equestris* (Cai et al. 2015). Lin et al. (2016) carried out transcriptome-wide analysis and characterization of *MADS-box* gene family in *Erycina pusilla*. Functional characterization of several *MADS-box* genes in *P. equestris* have been reported earlier (Tsai et al. 2004, 2005; Chen et al. 2012). However, analyses for protein characterization, gene structure, spatio-temporal expression and upstream *cis*-regulatory elements of *PeMADS* genes at a genome-wide scale were not performed earlier. The present study is an attempt to fill this lacuna and consolidate the fragmented information available on *PeMADS* gene family, and further attempts to physico-chemically characterize the *PeMADS* proteins and elucidate the ABCDE model of flowering through development of interaction network and gene expression profiling in *P. equestris*.

Material and Methods

Structural analysis

The PeMADS protein sequences and CDS sequences for *PeMADS* genes as reported by Cai et al. (2015), were retrieved from Orchidbase 3.0 database (<http://orchidbase.its.ncku.edu.tw/>). The genomic DNA and 1500 upstream promoter sequences for *PeMADS* genes were extracted from NCBI database (<https://www.ncbi.nlm.nih.gov/>). The intron-exon arrangement in *PeMADS* genes was drawn using GSDS - Gene Structure Display Server 2.0 (gsds.cbi.pku.edu.cn/). The *PeMADS* promoter sequences were analysed for occurrence of upstream *cis*-regulatory elements using PLACE (<https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?lang=en&pj=640&action=page&page=newplace>). The presence of MADS-box domain (PF00319) and K-box domain (PF01486) in *PeMADS* protein sequences was confirmed by SMART server (<http://smart.embl-heidelberg.de/>). The domain architecture was generated using IBS (Illustrator for biological sequences) tool (<http://ibs.biocuckoo.org/online.php>). The MULTALIN tool (multalin.toulouse.inra.fr/multalin/) was employed to locate the domain position by multiple sequence alignment. The conserved motifs were identified by MEME suite server (<http://meme-suite.org/tools/meme>) with preset parameters (maximum number of motifs - 10, number of repetitions - any, optimum motif width - ≥ 6 to ≤ 200).

Phylogenetic analysis

Full length *PeMADS* protein sequences were pre-aligned using inbuilt MUSCLE program and the phylogenetic tree was created using MEGA7 tool (<http://www.megasoftware.net/>) by neighbour-joining method at 1000 bootstrap replicates.

Protein characterization

The physico-chemical properties of *PeMADS* proteins like molecular weight, aliphatic index, instability index, pI and GRAVY value were calculated using PROTPARAM (<http://web.expasy.org/protparam/>). The sub-cellular localization of proteins was examined by CELLO v. 2.5 (<http://cello.life.nctu.edu.tw/>) and WoLF PSORT (<http://www.genscript.com/wolf-psort.html>). The presence of signal peptide and transmembrane region was explored respectively by SignalP.4.0 server (<http://www.cbs.dtu.dk/services/signalp/>), and TMHMM server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). GLOBPLOT (<http://globplot.embl.de/>) was employed to

evaluate the globularity of proteins.

Interaction network

Interaction input data was generated by analysing the structural basis of protein-protein interaction using Struct2Net tool (<http://cb.csail.mit.edu/cb/struct2net/webserver/>) with specific parameters (query method: Thread sequences onto all templates and struct2net prediction algorithm). The generated data was used to construct the interaction network among specific *PeMADS* proteins using Cytoscape tool (<http://www.cytoscape.org/>).

Duplication events and ortholog prediction

The sequence similarity within *PeMADS* CDS sequences were analysed using MAFFT tool (<https://www.ebi.ac.uk/Tools/msa/mafft/>) and genes sharing sequence similarity $\geq 80\%$ in CDS were considered duplicated (Sharma et al. 2019). The orthologous genes for *PeMADS* genes were predicted by BLASTn search against *MADS* genes of *A. thaliana* (*AtMADS*) and *O. sativa* (*OsMADS*) retrieved from their respective databases (<https://www.arabidopsis.org/>; <http://rice.plantbiology.msu.edu/>).

Expression Analysis

Spatio-temporal expression profiling of *PeMADS* genes was done using RPKM values retrieved from Orchidbase 3.0 database (<http://orchidbase.its.ncku.edu.tw>) for different vegetative and reproductive plant developmental stages such as leaf, root, floral stalk, flower, sepal, petal, labellum and gynostemium. The genes having null RPKM expression value in all the tissues were not included for evaluation. The heat map was generated using Hierarchical Clustering Explorer 3.5 (<http://www.cs.umd.edu/hcil/hce/>).

Results

Structural analysis

The *P. equestris* genome carries 51 *MADS-box* (*PeMADS*) genes, of which 48 genes were grouped into type I (20) [$M\alpha$ (8) and $M\gamma$ (12)] and type II (28) [$MIKC^C$ (27) and $MIKC^*$ (1)]. No $M\beta$ and $M\delta$ type genes were reported (Cai et al. 2015). Data was insufficient for one C class gene (*PeMADS1*) and two $M\alpha$ genes (*PeMADS5* and *PeMADS6*), hence these were not included for further

Fig. 1 Intron-exon arrangement in *MADS-box* genes. (A) Type I. (B) Type II. Exons and introns are represented as boxes and thin lines, respectively. 0 = phase 0 intron; 1 = phase 1 intron; 2 = phase 2 intron

analysis. Majority (16/20) of type I genes were intron-less, while others carried small-stretch and lesser number of introns (Fig. 1a). On the other hand, type II genes harboured multiple and long-stretch introns. The single MIKC* (*PeMADS9*) gene had the maximum of 10 introns (Fig. 1b). The intronic phases were majorly biphasic in type I, while in type II, 0 phase introns dominated (Fig. 1).

All the *PeMADS* proteins carried the signature *MADS-box* domain (~60aa). The type II *PeMADS* (except *PeMADS7/14/16/20/22/25*) additionally carried a less conserved K-box domain (~90aa), C-terminal domain and intervening I domain (Fig. 3, 4; Table 1). Motif analysis by MEME-suite server identified 10 conserved motifs. Motif 1 representing the *MADS-box* was present in all the proteins. Surprisingly, one protein (*PeMADS22*) lacked any of the identified motifs. Motifs 3, 4, 5, 8, 9 and 10 were restricted to type I, while motif 2 and 7 were restricted to type II proteins. Motif 2, representing the K-box domain, was

only present in MIKC^C group of type II proteins. Motif 3 and 4 represented internal repeats, and motif 5 and 9 represented the low complexity region, while the motif 6, 7, 8 and 10 were of unknown domains (Fig. 5).

Phylogenetic analysis

Phylogenetically, the type I proteins clearly diverged from the type II proteins and formed separate clades of Ma (*PeMADS37/39/40/43/47/48/50/53*) and My (*PeMADS29/30/31/32/33/34/54/55/56/57/58/60*) (Fig. 6). On the other hand, type II proteins congregated to form clear groups pertaining to classes, A (*PeMADS20/21/22*), B (*PeMADS2/3/4/5/6*), C (*PeMADS24/25*), D (*PeMADS7/38*) and E (*PeMADS8/10/11/12/13*). The other type II proteins grouped into subclades: OsMADS32-like (*PeMADS36*), SVP (*PeMADS23*), ANR1 (*PeMADS26/27*), Bs (*PeMADS28*), SOC1 (*PeMADS18/19*) AGL6 (*PeMADS14/16*) and MIKC*

Fig. 2 Distribution of ten most frequently occurs cis-regulatory elements on MADS-box genes. (A) Type I. (B) Type II. Five were annotated while five remained unknown

(PeMADS9). Noteworthy, PeMADS15 protein did not group with other AGL6 subclade and remained an outlier.

Protein characterization

PeMADS had an average length of 195aa and molecular weight of 21.92kDa that ranged from 61aa to 428aa and 7kDa to 35kDa. PeMADS56 (type I Ma) was the largest protein (428aa with 49.2kDa), while the PeMADS14 (type

II MIKC^C) represented the smallest protein (61 aa with 7kDa). The isoelectric point ranged from 5.9 to 9.7 in type I, and 4.9 to 10.1 in type II proteins, indicating that most of the PeMADS proteins are alkaline in nature as reported earlier in *O. sativa* (Arora et al. 2007). The aliphatic index value ranged from 67.7 to 101.8 in type I, and ranged from 56.5 to 106.02 in type II (File S1). All the proteins were predicted unstable except for the three proteins (PeMADS03/23/27) of MIKC^C type II. The grand average

Fig. 3 Domain Organization in MADS-box proteins. (A) Single MADS domain shown in red within type I MADS-box proteins. (B) Additional K domain in blue within type II MADS-box proteins. Green color represents the protein length

of hydropathy (GRAVY) value ranged from -0.922 to 0.263 with all the proteins having a negative GRAVY value, except PeMADS37 (type I) and PeMADS22 and PeMADS25 (type II) indicated that most of the PeMADS proteins are hydrophilic in nature (File S1). The globularity of proteins was also evaluated and it was found that most of them were predicted to be globular except for PeMADS22. None of the proteins harboured a signal peptide sequence. Except for few PeMADS proteins, all proteins were predicted to be localised in nucleus (File S1). Two proteins, PeMADS37 and PeMADS25 were predicted to be plasma

membrane localised and displayed a transmembrane helix region, while PeMADS54 (MY), PeMADS36 (OsMADS32-like) and PeMADS22 (A-class) were localised in mitochondria, cytoplasm and extracellular region, respectively.

Protein-protein interaction (PPI)

The putative physical interaction network among ABCDE classes of MADS-box proteins confirmed their role in development of specific floral whorls. The constructed interaction network amongst various proteins of classes A, B, C, D

Fig. 4 Multiple sequence alignments of MADS-box proteins (A) Type I. (B) Type II. The green box represents the MADS domain and the black over lined region represents the signal motif (KR(K/R)₄KK) required for nuclear localization. In type II, the less conserved I domain is present in between MADS and the variable C-domain at the c-terminus region. Asterisks represent the conserved amino acid residues

and E (Fig. 7) showed direct interaction with each other. The class A proteins (PeMADS20/21) showed interaction with other classes of MADS-box protein. With respect to interactions between members of class A and B, PeMADS21 of class A displayed interaction with all B class proteins (PeMADS2/3/4/5/6), while the PeMADS20 of A class showed interaction only with PeMADS3/4/5 (Fig. 7). A and B class proteins interact to regulate the development of petals. Interestingly, class A proteins showed interactions with proteins of class C and class D as well of the C class

proteins, only PeMADS24 showed interaction with both the A class proteins (PeMADS20/21), while PeMADS25 did not show any interaction. Both the A class proteins (PeMADS20/21) showed interaction with D class protein (PeMADS38) (Fig. 7). Notable interactions were observed among class A and E proteins, and this can be correlated to their combined expression during sepal development. Both A class proteins showed to interact only with two E class (PeMADS12/13) proteins (Fig. 7). Intra-class interaction between both A class proteins was also observed. Within

Table 1 Domain Organization in MADS-box proteins

GENE IDs	MADS box Domain	PFAM:kbox	Transmembrane domain	Internal repeat 1 (RPT)	Coiled coil region	Low complexity region	Unknown region
PeMADS60	1-60						61-175
PeMADS58	1-60			119-130 175-186		191-1278	61-118 131-174
PeMADS57	1-60				89-119		61-88 69-148
PeMADS56	9-68			149-191 247-294 141-189		308-323	192-246 295-307 324-428
PeMADS55	1-60			189-239		249-257	61-140 240-248 258-305
PeMADS54	1-60						61-214
PeMADS53	4-63					100-111	64-99 112-125
PeMADS50	4-61					83-94	62-82 95-148
PeMADS48	4-61					83-94	62-82 95-148
PeMADS47	4-63					100-111	64-99 112-125
PeMADS43	4-69						70-149
PeMADS40	10-69						70-116 69-93
PeMADS39	9-68				94-176	188-203	117-187 204-226
PeMADS37	4-63		147-169				64-146 170-215
PeMADS34	1-49						50-157
PeMADS33	1-60						61-127
PeMADS32	1-60				86-115		61-85 116-151
PeMADS31	1-60						61-111
PeMADS30	1-60			221-226 255-357			61-120 227-254 358-384
PeMADS29	2-61						62-186
PeMADS8	1-60	80-172					61-79 173-250
PeMADS7	1-60						61-69
PeMADS6	1-60	72-165					61-71 166-210
PeMADS5	1-60	77-169					61-76 170-219
PeMADS4	1-60	77-169					61-76 170-224
PeMADS38	3-62	81-172					63-80
PeMADS36	1-60	78-155				157-172	61-77
PeMADS3	1-60	75-169				174-185	61-74 186-222
PeMADS28	1-60	75-140					61-74
PeMADS27	1-60	83-181					61-82
PeMADS26	1-60	63-171					182-194 172-192

Table 1 Domain Organization in MADS-box proteins

GENE IDs	MADS box Domain	PFAM:kbox	Transmembrane domain	Internal repeat 1 (RPT)	Coiled coil region	Low complexity region	Unknown region
PeMADS25	8-67		70-92				93-103
PeMADS24	6-65	85-176					66-84
PeMADS23	1-60	85-171					177-219
PeMADS22	5-53						61-84
		81-173					172-228
PeMADS21	1-60					230-253	54-74
							61-80
						276-289	174-229
							254-275
PeMADS20	1-60						290-301
						164-182	61-163
PeMADS2	27-86	104-195					183-290
						6-21	87-103
PeMADS19	1-60	80-172					196-253
							61-79
PeMADS18	1-60	80-171					173-222
PeMADS16	6-65						61-79
PeMADS15	1-60	78-167					172-223
PeMADS14	1-60						66-80
PeMADS13	1-60	62-182					
							61-81
PeMADS12	1-60	77-172					163-175
							61-76
PeMADS11	1-60	79-172					173-246
							61-78
PeMADS10	1-60	80-172					173-244
							61-79
		264-289					173-244
PeMADS9	1-60				107-142		61-106
							143-263
							290-321

B class and C class proteins, PeMADS3/4/5 of B class, showed interaction with only PeMADS24 of C class but not with PeMADS25 (Fig. 7). PeMADS2/6 of B class did not show interaction with C class members. All the B class proteins show interaction with all E class proteins (PeMADS10/12/13), except PeMADS11 (Fig. 7). The C class and D class interaction was restricted with interaction of PeMADS24 (C class) with PeMADS38 (D class) (Fig. 7). Considering C class interaction with E class, all the C class genes interacted with three E class proteins (PeMADS10/12/13), but not with PeMADS11. E class proteins, PeMADS10/11/13, also showed interaction with each other (Fig. 7).

Duplication events and ortholog prediction

Nine duplication events (DEs) were predicted, out of

which only one (*PeMADS8-PeMADS11*) was from type II, and the rest were from type I genes. In Ma type, three DEs (*PeMADS48-PeMADS50*, *PeMADS47-PeMADS53* and *PeMADS40-PeMADS39*) and in Mγ type, five DEs (*PeMADS60-PeMADS58*, *PeMADS33-PeMADS55*, *PeMADS33-PeMADS56*, *PeMADS55-PeMADS56*, and *PeMADS34-PeMADS30*) were predicted (File S2). The orthologous *MADS-box* genes were predicted from *A. thaliana* and *O. sativa*. A total of 29 and 21 orthologous genes were found in *A. thaliana* and *O. sativa*, respectively (File S3).

Promoter analysis

The identification of *cis*-acting promoter elements disclosed the significance of *MADS-box* genes in growth and developmental processes. Several elements that direct spatio-temporal expression of genes identified showed correlation

Table 2 Distribution of Type I and Type II genes in different plant genome

Type / Plant	<i>P. equestris</i>	<i>A. sherzhenica</i> (Zhang et al. 2017)	<i>D. officinale</i> Lindl. (Zhang et al. 2016)	<i>A. thaliana</i> (Parenicova et al. 2003)	<i>O. sativa</i> (Arora et al. 2007)	<i>Z. mays</i> (Zhao et al. 2011)	<i>B. distachyon</i> (Wei et al. 2014)	<i>G. max</i> (Shu et al. 2013)	<i>M. domestica</i> (Tian et al. 2015)	<i>S. bicolor</i> (Zhao et al. 2011)	<i>T. aestivum</i> (Ma et al. 2017)
Ma	16.66%	13.8%	23.80	18.69%	17.33%	36%	15.78%	22.69%	15.06%	40%	17.77%
Mβ	-	-	-	15.88%	12%	4%	12.2%	8.58%	5.47%	3.07%	
Mγ	25%	11.11%	20.63%	19.62%	13.33%	2.66%	3.50%	14.72%	16.43%	3.07%	2.77%
Mδ	-	-	-	3.73%	-	-	-	-	6.84%		2.77%
Total Type I	41.66%	25%	44.44%	57.92%	42.66%	42.66%	31.48%	46%	43.8%	46.1%	23.33%
MIKC	56.25%	69.44%	50.79%	40.18%	52%	52%	56.14%	49.69%	51.36%	50.76%	76.66%
MIKC*	2.08%	5.55%	4.76%	1.86%	5.53%	5.53%	12.28%	4.29%	4.79%	3.07%	
Total Type II	58.33	75%	55%	42%	57.53%	57.53%	68.42%	54%	56.15%	53.84	76.66%
Total MADS-box genes	48	36	63	107	75	75	57	163	146	65	180

with the expression patterns of the concerned genes (File S4). Type I gene promoters predominantly carried elements (ACGTABOX, ACGTCBOX, ABRELATERD1, SORLIP1AT, SEF3MOTIFGM, SEF1MOTIF, SEF4MOTIFGM7S, AGAMOUSATCONSENSUS, AGATCONSENSUS) that are responsible for gynoeceium specific expression. The promoter region of *PeMADS39* gene showed presence of root and flower specific elements (ASF1MOTIFCAMV, CCAATBOX1, and CIACADIANLELHC). The *Mγ* gene promoters carried elements for flower specific expression (CCAATBOX1, MYB26PS, LEAFYATAG, and WUSATAG). The *PeMADS54* gene harboured leaf specific expression element (CCA1ATLHCBI, DOFCOREZM) in its promoter region.

A similar trend was observed in promoter elements distribution in type II *MADS-box* genes as expected. The A-class *MADS-box* genes mainly carried elements (CCAATBOX1, AGL2ATCONSENSUS, CARGCW8GAT, AGATCONSENSUS, WUSATAG, and CCAATBOX1), that directs expression in sepal, and elements that are responsible for high expression in gynostemium (such as ACGTABOX, ACGTCBOX, ABRELATERD1, BOXIIPCCHS, GADOWNAT, SORLIP1AT, SORLIP5AT, SEF3MOTIFGM, SEF1MOTIF, and SEF4MOTIFGM7S). The *PeMADS20* gene carried flower specific promoter elements (CARGCW8GAT, CARGNCAT, CCAATBOX1, CCAATBOX1, and MYB26PS). The *PeMADS21/22* genes displayed root specific element (ASF1MOTIFCAMV).

The B-Class genes, with significant expression in reproductive tissues (flower, sepal, and petal), carried a combination of regulatory elements (CCAATBOX1, MYB26PS, AGL2ATCONSENSUS, and CARGCW8GAT)

that have direct role in reproductive tissues. The C/D-class genes with role in female reproductive organ development harboured regulatory elements that are essential for expression in gynostemium (AGAMOUSATCONSENSUS, AGATCONSENSUS, and LEAFYATAG). Promoters of the E-class genes that participate in coordination of flower development showed predominant presence of flower development specific elements such as CCAATBOX1, and CARGCW8GAT. The other type II genes carried diverse elements in upstream region. The *PeMADS36* (*OsMADS32*-type) gene promoter had root and leaf specific elements ASF1MOTIFCAMV, AGL1ATCONSENSUS, AGL3ATCONSENSUS, while the *PeMADS28* (Bs-type) promoter carried flower and gynostemium specific elements (CCAATBOX1, ACGTCBOX, ABRELATERD1, BOXIIPCCHS, GADOWNAT, SEF3MOTIFGM, AGAMOUSATCONSENSUS, AGATCONSENSUS). The *PeMADS23* gene showed expression in leaf, root, and floral stalk, and carried elements such as ASF1MOTIFCAMV, CCAATBOX1, CARGNCAT, AGL3ATCONSENSUS, LEAFYATAG, and CARGCW8GAT. The *ANR1* (*PeMAD27/26*) and *SOC1* type (*PeMADS18/19*) genes had leaf and root specific promoter elements (ASF1MOTIFCAMV, LEAFYATAG) (File S4).

Expression analysis

The expression profile of *PeMADS* genes in different tissues such as leaf, root, floral stalk, flower, sepal, petal, labellum and gynostemium was analysed on the basis of RPKM value. Genes having null RPKM values were not considered for analysis. The remaining 14 type I genes had high

Fig. 5 Organization of Conserved motifs of MADS-box proteins. (A) Type I. (B) Type II. (C) Ten conserved motif sequences of MADS-box proteins showing e-value and site count

expression in gynostemium. Apart from gynostemium expression, the *Ma* genes, *PeMADS37* expressed in floral stalk and floral parts, while *PeMADS40* showed expression in flower. Interestingly, *PeMADS39* exhibited expression

in root as well as in flower. Similarly, in addition to gynostemium, majority of the *My* genes displayed significant expression in flower and petal; *PeMADS54* had high expression in leaf as well (Fig. 8A).

Fig. 6 Phylogenetic relationship among type I and type II MADS-box proteins. The MADS-box family proteins involved in floral whorls formation are grouped together

The three A class genes (*PeMADS22/21/20*) had high expression in gynostemium. The genes *PeMADS22/21* had least expression in sepals and high expression in root, with *PeMADS21* also expressing moderately in leaf (Fig. 8B). The five B class genes (*PeMADS6/5/2/3/4*) had least expression in vegetative parts and high expression in flower. Additionally, the *PeMADS4* gene expressed in gynostemium, labellum and petal, while the *PeMADS6/5/2* genes showed maximum expression in sepals and petals, and the *PeMADS3* gene showed high expression in petal and labellum (Fig. 8). Of the two C class genes (*PeMADS24/25*), *PeMADS24* gene showed expression in root, flower, sepal and gynostemium, and the *PeMADS25* gene had expression in flower, petal and gynostemium (Fig. 8). Similarly, the two D genes (*PeMADS38/7*) had intense expression in gynostemium and petals. The *PeMADS38* gene also had maximum expression in flower. The five E class genes (*PeMADS13/11/8/12/10*) had maximum expression in flower, sepal and petal. The *PeMADS13/11* genes also showed maximum expression in labellum, and the *PeMADS12/10* genes also showed moderate expression in labellum and gynostemium (Fig. 8).

The only SVP sub-family gene (*PeMADS23*) showed

high expression in leaf, root and floral stalk. The Bs sub-family gene (*PeMADS28*) showed maximum expression in flower and gynostemium and moderate expression in petal. The ANR1 sub-family genes (*PeMADS26/27*) and SOC1 sub-family genes (*PeMADS19/18*) showed maximum expression in root (Fig. 8). The AGL6-like sub-family genes (*PeMADS16/15*) had maximum expression in sepal and petal and moderate expression in labellum and gynostemium. The single OsMADS32-like sub-family gene (*PeMADS36*) gene had expression in leaf, root and weak expression in flower. The only MIKC* type II gene (*PeMADS9*) had maximum expression in flower, petal and gynostemium (Fig. 8)

Discussion

The *MADS-box* genes are floral homeotic genes with critical role in plant growth and development and form the basis of ABCDE model of flowering. The highly complicated and advanced floral organization of *Phalaenopsis equestris* makes it an ideal organism to study florigenesis and the interrelationship of *MADS-box* genes in the

Fig. 7 Protein-Protein Interaction (PPI) Network among MADS-box proteins of ABCDE classes

process. In the present study, *MADS-box* (*PeMADS*) gene family is characterized in *P. equestris* extensively for various genomic and proteomic attributes. The size of *MADS-box* gene family varied widely across different plant species. The *P. equestris* genome carried more of type II *MADS-box* genes than type I *MADS-box* genes. A similar trend was reported with closely related genomes *Apostasia shenzhenica* (Zhang et al. 2017) and *Dendrobium catenatum* (Zhang et al. 2016) and in other plant genomes such as *Glycine max* (Fan et al. 2013), *Brachypodium distachyon* (Wei et al. 2014) and *Malus domestica* (Tian et al. 2015) as well (Table 2). This indicates, presumably, that the type II group members are evolutionarily more advanced than the type I group members (Lin et al. 2016). However, *Arabidopsis thaliana* consisted of more type I than the type II *MADS-box* genes. This can be attributed to unrelated gene duplication pattern of type I and type II groups (Parenicova et al. 2003), indicating different evolutionary strain in different plant groups. Such duplication events bring gain and loss of function events and has been proposed as an important mechanism for neo/non-functionalization and sub-functionalization (Irish and Litt 2005). The lesser number of intron or intron-less genes in type 1 can be due to the simple gene structure and reverse transcribed organization (Parenicova et al. 2003; Wells et al. 2015). While the complex gene structure and homo- and hetero- dimer formation for specifying the floral organ identity have resulted in the more number of

introns in type II genes (Theissen 2001; Egea-Cortines et al. 1999; Eckardt 2003). This further corroborates that the type II genes are evolutionary more advanced than the type I genes (Lin et al. 2016). The intron phase analysis indicated the occurrence of maximum 0 phase followed by 2 and 1 phase, suggesting the conserved nature of splicing phases in the *MADS-box* genes (Fig. 1A-B). The physico-chemical characters of *PeMADS* were also comparable in accordance with studies in *Erycina pusilla* (Lin et al. 2016), *A. thaliana* (Parenicova et al. 2003), *Oryza sativa* (Arora et al. 2007) and *G. max* (Shu et al. 2013).

In *P. equestris*, type I *MADS-box* proteins were grouped into Ma and My. Interestingly, in *E. pusilla* only one group i.e. My type is present (Lin et al. 2016), while in *B. distachyon* and *A. thaliana*, Ma, Mβ and My groups are present (Wei et al. 2014; Parenicova et al. 2003) and all the four groups (Ma, Mβ, My, and Mδ) are reported in *Cucumis sativus* (Hu and Liu, 2012). This inconsistency within the type I groups in different plant species indicate the varying birth and death rate of type I genes (Nam et al. 2004). Among type II *MADS-box* proteins, nine subgroups were formed in *P. equestris* (Cai et al. 2015), while nine in *E. pusilla* (Lin et al. 2016) and *B. distachyon* (Wei et al. 2014), 10 in *C. sativus* (Hu and Liu 2012), eight in *M. domestica* (Tian et al. 2015), 14 in *O. sativa* (Arora et al. 2007) and 13 were reported in *Vitis vinifera* (Diaz-Riquelme et al. 2009). The variability in number of subgroups in different plant species might be

Fig. 8 Expression profiles of *MADS-box* genes in various tissues. (A) Type I genes. (B) Type II subfamily genes (A, B, C, D and E class gene) (C) Type II subfamily genes (*SVP*, *Bs*, *ANR1*, *SOC1*(or *AGL20*), *AGL6*-like, *OsMADS32*-like and *MIKC**)

due to the requirement of specific functions of these groups (Shimeld 1999). Interestingly, the *FLC* genes involved in vernalization (Michaels and Amasino 1999) and the *AGL12*-like genes with a role in root proliferation and floral transition (Tapia-Lopez et al. 2008) were absent in *P. equestris* (Cai et al. 2015). But it is also reported that *FLC* are difficult to isolate owing to their selectively smaller size and highly divergent nature (Rouse et al. 2002). *FLC* and *AGL12*-like proteins probably have a different mechanism from proteins belonging to *ANR1* and *SOC1* (or *AGL20*) subgroups which are also responsible for root proliferation and floral transition (Hepworth et al. 2002; Liu et al. 2008; Zhang et al. 2000). The functioning of these genes in *P. equestris* might have been lost or diverged into a separate group during the course of evolution.

Reports on type I genes are sporadic and mostly focussed towards elucidating their role in plant reproduction and developmental stages, chiefly in female gametophyte, embryo and endosperm development (Bemer et al. 2008; Bouyer et al. 2011; Kang et al. 2008; Masiero et al. 2011; Portereiko et al. 2006). *PeMADS* type I genes expressed predominantly in gynostemium with overlapping expression in sepal, labellum and flower (Fig. 8a) which is similar to earlier reports from *B. distachyon* (Wei et al. 2014), *E. pusilla* (Lin et al. 2016) and *A. thaliana* (Parenicova et al. 2003), suggesting their role in female gametophyte development. The Ma gene, *PeMADS39*, showed high expression in root and flower which is in line with the expression of *Raphanus sativus* genes (*RsMADS093* and *RsMADS111*; Li et al. 2016) and *C. sativus* genes (*CsMADS34/35*; Hu and Liu 2012). The My *PeMADS54* is highly expressed in leaf,

which is comparable with expression of *PpeMADS31* of *Prunus persica* (Wells et al. 2015) and *CsMADS36/37/38* in *C. sativus* (Hu and Liu 2012). Additionally, this gene also has leaf specific elements in its promoter region (CCA1ATLHCB1, DOFCOREZM).

The type II genes clearly diverged into separate groups pertaining to A, B, C, D and E classes (Fig. 6). Malcomber and Kellogg (2005) suggests the evolution of A and E class proteins involved gene duplication events, and they coordinate with each other for sepal development (Coen and Meyerowitz 1991). Predicted interaction network projects intense interaction within A and E class members (Fig. 7). The expression profile also confirms that class A and E genes predominantly express in similar tissues (Fig. 8B). The expression of A class genes in gynostemium is also justified by the presence of gynostemium specific elements (ACGTABOX, ACGTCBOX, ABRELATERD1, BOXIIPCCHS, GADOWNAT, SORLIP1AT, SORLIP5AT, SEF3MOTIFGM, SEF1MOTIF, and SEF4MOTIFGM7S). Expression of A class genes in reproductive tissues was also been reported in *Dendrobium* (*DoMADS2*; Aceto and Gaudio 2011) and *E. pusilla* (*EpMADS12*; Lin et al. 2016). All genes of E class were highly expressed in sepals (Fig. 8B) which is similar to the expression of *AdOMI* in *Aranda* Deborah (Lu et al. 1993) and *EpMADS9* in *E. pusilla* (Lin et al. 2016). The putative interaction network also suggested that class A and E genes are associated with each other (Fig. 7) supporting their synergistic role for whorl 1 development (Pelaz et al. 2000). In addition, *PeMADS21* and *PeMADS22*, the class A genes had root specific promoter element (ASF1MOTIFCAMV) and showed

high expression in root tissues, similar to *OsMADS18*, a class A gene in *O. sativa* root (Fornara et al. 2004) indicating their additional role in root development.

The combined role of class A, B and E genes in whorl 2 development can be reflected in interaction studies (Fig. 7) which is in accordance with earlier reports (Pelaz et al. 2000). *PeMADS22* (class A) gene had high expression in petals (Fig. 8B), similar to expression of *DOMADS1* in *Dendrobium* (Yu and Goh 2000) and *RsMADS68* in *R. sativus* (Li et al. 2016). All the genes of class B exhibited high expression in flower, similar to their expression in *C. sativus* (*CsMADS21/22/23*) (Hu and Liu 2012). The expression of *PeMADS2/5/6* (class B) genes in petals was in conformity with class B genes of *E. pusilla* (*EpMADS14/15/16*) (Lin et al. 2016) and *B. distachyon* (*BdMADS5*) (Wei et al. 2014). All genes of class E of *P. equestris* were found to be upregulated in petals which is in concordance with expression of *DcSEP1* in *Dendrobium crumenatum* (Xu et al. 2006), *PhaMADS7* in *Phalaenopsis* ‘Athens’ (Acri-Nunes-Miranda and Mondragón-Palomino 2014) and *RsMADS033* in *R. sativus* (Li et al. 2016). Furthermore, the remaining two class B genes, *PeMADS3* and *PeMADS4* expressed in labellum, as also reported for genes in *E. pusilla* (*EpMADS13/14/16*) (Lin et al. 2016) and *Oncidium* Gower Ramsey (*OMADS8*) (Chang et al. 2010). The expression of *PeMADS4* gene in gynostemium tissue (Fig. 8C), can be related to the expression of *BdMADS20* gene in carpel of *B. distachyon* (Wei et al. 2014).

Studies on *P. equestris* peloric mutants along with wild type by Tsai et al. (2004) revealed roles of specific genes in specific organ development. The sepals of peloric mutant had expression of only *PeMADS2* in place of *PeMADS2* and *PeMADS5* for wild type flower, confirming critical role of *PeMADS5* in sepal development. The expression of all four *AP3-like* genes were expressed in petals of wild flower except *PeMADS4*. In case of peloric mutant, *PeMADS5* was not detected (lip-like petals), suggesting the role of *PeMADS5* in petal formation. Moreover, *PeMADS4* was highly expressed in labellum and gynostemium of wild flower. The expression of gene was noted in lip-like petals of peloric mutant flower, signifying the role of *PeMADS4* in labellum formation. Furthermore, *PeMADS3* gene have similar expression pattern in both wild and peloric mutant flower, indicating their role in inner perianth whorl. The *PI-like* gene (*PeMADS6*), found to be having no significant role in peloric mutant phenotype (Chen et al. 2012).

Many models have been proposed for understanding the functioning of these B class genes. Mondragon-Palomino

and Theißen, (2008, 2009) proposed a model called ‘Orchid Code’ which implicit that combination of different genes to specify the different organ development. This model includes four different clades of *DEF-like* genes, which comes from the duplication of four *DEF-like* genes. Each specific clade expressed in specific organ of perianth. Clade 1 and 2 involve *PeMADS2-like* and *OMADS3-like* genes, respectively, which are present in all the tepals (Mondrago-Palomino and Theißen 2011), while clade 3 (*PeMADS3-like*) specifically expressed in inner tepals and clade 4 (*PeMADS4-like*) entirely shows expression in labellum. This Orchid Code described the expression of clade 1 and 2 in the outer tepals of flower and clade 1, 2 and 3 show presence in lateral inner tepals while clade 1, 2, 3 and 4 stipulate the formation of lip. Later, the Orchid Code was refined by Mondrago Palomino and Theißen (2011). The model includes the comparison of expression of four *DEF-like* genes in outer tepals, inner lateral tepals, labellum, gynostemium and ovary of different species (*Vanilla planifolia*, *Phragmipedium longifolium*, wild-type and peloric flowers of *Phalaenopsis* hybrid ‘Athens’). Hsu et al. (2015) proposed another model named ‘Perianth or P-code’ based on the expression, specifying the perianth formation in *Oncidium* species. It involves the interaction of *OPI* with *AP3/AGL6* that lead to the formation of two protein complexes, SP (sepal/petal) and L (Lip). The *OPI* gene is universally expressed in perianth. The sepal/petal development is determined by SP complex with *OAP3-1* and *OAGL6-1* proteins, while the L complex with *OAP3-2* and *OAGL6-2* determines the lip development. Together the negative interaction of SP and L complex governs the non-reproductive part of flower. This model is specifically applicable to perianth development, not for stamen and carpel formation.

The androecium (whorl 3) and gynoecium (whorl 4) in orchids are fused to form gynostemium. Theoretically, whorl 3 is controlled by Class B, C, and E. *PeMADS4* (class B) gene and *PeMADS24/25* (class C) genes had high expression in gynostemium and was comparable with class B genes *PpeMADS56* of *P. persica* (Wells et al. 2015) and *RsMADS078* of *R. sativus* (Li et al. 2016) in stamens, and with class C genes *CeMADS1* of *Cymbidium ensifolium* in gynostemium (Wang et al. 2011), *RsMADS036* in *R. sativus* stamens (Li et al. 2016) and *OMADS4* in *Oncidium* Gower Ramsey column (Hsu et al. 2010). Chen et al. (2012) reported that, *PeMADS1* ectopic expression mutants turned petals to gynostemium-like structures. For E class genes *PeMADS10/12* were expressed in gynostemium as similar to *EpMADS6* gene of *E. pusilla* (Lin et al.

2016), *SEP* genes of *A. thaliana* (Pelaz et al. 2000) and *EgMADS3* of *Eucalyptus* (Southerton et al. 1998). Conventionally, whorl 4 includes class C, D and E genes. Class C/D genes (*PeMADS7/38/24/25*) were highly expressed in gynostemium (Fig. 8B). This class genes, *EpMADS20/21/22* of *E. pusilla* (Lin et al. 2016), *PhaMADS8/10* of *Phalaenopsis* ‘Athens’ (Aciri-Nunes-Miranda and Mondragón-Palomino 2014), *PhalAG1* of *Phalaenopsis* Hatsuyuki (Song et al. 2006) and *DthyrAG1* of *D. thyrsiflorum* (Skipper et al. 2006) were reported to have similar expression. The C class genes are also reported to have a role in fruit formation of *P. persica* (*PpeMADS11/24/45*) (Wells et al. 2015). Class D genes (*PeMADS38*, *PeMADS7*) were expressed in gynostemium, thus resembling the expression pattern of other class D genes of *RsMADS047* in *R. sativus* (Li et al. 2016), *BdMAD2* in *B. distachyon* (Wei et al. 2014) and *EpMADS23* in *E. pusilla* (Lin et al. 2016). The E class genes (*PeMADS12* and *PeMADS10*) have comparable expression pattern with *VvMADS5* of *V. vinifera* (Boss et al. 2002). The control of whorl 3 and whorl 4 by class B, C, D and E genes, has also been revealed by the significant interaction among these genes on the basis of predicted protein structure (Fig. 7).

Class E genes (*PeMADS8/10/11/12/13*) were expressed in all the floral organs (Fig. 8B–E) which is in accordance with earlier reports on class E genes [*RsMADS017/023* in *R. sativus* (Li et al. 2016), *BdMADS11/26* in *B. distachyon*, *EpMADS6* in *E. pusilla* (Lin et al. 2016) and *OMADS6/11* in *Oncidium* Gower Ramsey (Chang et al. 2009)]. These results suggest that class E genes are required for the functioning of all floral whorls.

A significant interaction between class A and C genes was observed in the protein-protein interaction network (Fig. 7). This interaction has been reported to be quite significant but mutually antagonistic (Drews et al. 1991; Pelaz et al. 2000). Their mechanism of action can be explained on the basis of a number of studies. Class A genes are reported to be restricted to outer two whorls and prevent the activity of class C genes (Irish and Sussex 1990). Activity of class C genes, however, is reported in the inner two whorls restricting the expression of class A genes (Yanofsky et al. 1990). This is evident in their expression pattern (Fig. 8b). Interestingly, class A comprises of *AP1* and *AP2* genes, but the *AP2* gene does not encode typical MADS-box domain (Riechmann and Meyerowitz 1997). The normal development of whorl 1 and 2 requires a significant expression of *AP2* to prevent the accretion of *AG* transcripts, which could in turn trigger the formation of carpels instead (Drews et al. 1991; Jofuku et al. 1994).

Similarly, the activity of class C genes is essential for floral determinacy; in the absence of which, there is a formation of indeterminate number of floral whorls (Pelaz et al. 2000). The antagonist interaction between class A and C was also reported in the *ap2* mutant of *A. thaliana* (Coen and Meyerowitz 1991). Ectopic expression of a class C gene (*PeMADS1*) was reported to be associated with development of gynostemium-like petal in *P. equestris* (Chen et al. 2012).

The SVP gene subfamily in *C. sativus* (Hu and Liu 2012) and *O. sativa* (Arora et al. 2007) had crucial role in leaf and root development. The presence of leaf specific elements (LEAFYATAG) and expression of *PeMADS23* reflects the same (Fig. 8F). The B sister gene *EpMADS24* of *E. pusilla* (Lin et al. 2016) and *VvBS2* of *V. vinifera* (Diaz-Riquelme et al. 2009) proposed to have a crucial role in cell differentiation during ovule and seed development, and similar expression was observed with *PeMADS28* (B sister) gene in *P. equestris*. The *SOC1* (or *AGL20*) subfamily genes (*PeMADS18/19*) were expressed in root and carried root and leaf specific promoter elements (ASF1MOTIFCAMV, LEAFYATAG) and the expression was similar to *AGL14/19* of *A. thaliana* (Parenicova et al. 2003), and *PpeMADS64* of *P. persica* (Wells et al. 2015). This up-regulation in root might also be associated with nutrient deficiency response (Gan et al. 2005). The role of *SOC1* (or *AGL20*) genes in controlling flowering time has also been reported (Duan et al. 2015). In addition, *PeMADS18* gene was also strongly detected in leaf as reported in *V. vinifera* (Diaz-Riquelme et al. 2009). The *ANR1* subfamily (*PeMADS26/27*) was found to have similar expression pattern and similar promoter elements (ASF1MOTIFCAMV, LEAFYATAG) in root like *CsMADS27/28/29* of *C. sativus* (Hu and Liu 2012) and *AGL16/17/21* of *A. thaliana* (Parenicova et al. 2003), suggesting their function in lateral root growth in response to nutrients and root development (Gan et al. 2005; Zhang and forde 2000). The *AGL6*-like genes (*PeMADS14/15/16*) were expressed in sepal, petal and labellum (Fig. 8F), related to *OsMADS6* of *O. sativa*, which known to regulate floral meristem determinacy and floral organ identity (Ohmori et al. 2009). *OsMADS32*-like gene has similar expression like *OsMADS32* of *O. sativa* (Arora et al. 2007) and *TaAGL14/15* in *T. aestivum* (Zhao et al. 2006) and probably have role in plant developmental processes. The MIKC* genes have similar expression pattern as in *CsMADS30* in *C. sativus* (Hu and Liu 2012), but the functional significance of MIKC* is not much clear in development of floral whorls. The expression of MIKC* in gynostemium of *P. equestris*

and preceding studies in *A. thaliana* (Parenicova et al. 2003), and *E. pusilla* (Lin et al. 2016) have suggested a role for MIKC* genes in gametophyte development.

Conclusions

In the present study, an attempt was made to validate ABCDE model in *P. equestris*, on the basis of structural and functional characterization of *MADS-box* genes with special thrust to expression profiling and protein-protein interaction. This provides a platform to functionally elucidate the mechanism of flower development and open newer vistas to functional validation of the MADS-box gene family in florigenesis.

Acknowledgements

Authors are grateful to Department of Science and Technology, Government of India for partial financial support under Promotion of University Research and Scientific Excellence (PURSE) grant scheme. JKS is thankful to Bio-CARe scheme (BT/Bio-CARe/647/2011-12) of Department of Biotechnology, Government of India, and Science and Engineering Research Board (SERB), Government of India for Early Career Research Award (ECR/2016/001190).

References

- Aceto S, Gaudio L (2011) The MADS and the Beauty: Genes Involved in the Development of Orchid Flowers. *Curr Genomics* 12:342–356
- Acri-Nunes-Miranda R, Mondragón-Palomino M (2014) Expression of paralogous SEP-, FUL-, AG- and STK-like MADS-box genes in wild-type and peloric *Phalaenopsis* flowers. *Front. Plant sci* 5:76
- Adamczyk BJ, Fernandez DE (2009) MIKC* MADS Domain Heterodimers Are Required for Pollen Maturation and Tube Growth in Arabidopsis. *PLANT Physiol.* 149:1713–1723
- Alvarez-Buylla ER, Liljegren SJ, Pelaz S, et al. (2000b) MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J.* 24:457–66
- Alvarez-Buylla ER, Pelaz S, Liljegren SJ, et al. (2000a) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* 97:5328–33
- Angenent GC, Colombo L (1996) Molecular control of ovule development. *Trends Plant Sci.* 1:228–232
- Arora R, Agarwal P, Ray S, et al. (2007) MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* 8:242
- Bemer M, Wolters-Arts M, Grossniklaus U, Angenent GC (2008) The MADS Domain Protein DIANA Acts Together with AGAMOUS-LIKE80 to Specify the Central Cell in Arabidopsis Ovules. *Plant Cell Online* 20:2088–2101
- Boss PK, Sensi E, Hua C, Davies C, Thomas MR (2002) Cloning and characterisation of grapevine (*Vitis vinifera* L.) MADS-box genes expressed during inflorescence and berry development. *Plant Sci* 162:887–895
- Bouyer D, Roudier F, Heese M, et al. (2011) Polycomb Repressive Complex 2 Controls the Embryo-to-Seedling Phase Transition. (GP Copenhagen, Ed.). *PLoS Genet.* 7:e1002014
- Bowman JL, Smyth DR, Meyerowitz EM (1991) Genetic interactions among floral homeotic genes of Arabidopsis. *Development* 112:1–20
- Cai J, Liu X, Vanneste K, et al. (2014) The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* 47:65–72
- Chang YY, Chiu YF, Wu JW, Yang CH (2009) Four Orchid (*Oncidium* Gower Ramsey) AP1/AGL9-like MADS Box Genes Show Novel Expression Patterns and Cause Different Effects on Floral Transition and Formation in *Arabidopsis thaliana*. *Plant Cell Physiol.* 50:1425–1438
- Chang YY, Kao NH, Li JY, et al. (2010) Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. *Plant physiol.* 152:837–53
- Chen YY, Lee PF, Hsiao YY, et al. (2012) C- and D-class MADS-Box Genes from *Phalaenopsis equestris* (Orchidaceae) Display Functions in Gynostemium and Ovule Development. *Plant Cell Physiol.* 53:1053–1067
- Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353:31–37
- Davies B, Egea-Cortines M, Andrade Silva E de, Saedler H, Sommer H (1996) Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J.* 15:4330–43
- 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
- Diaz-Riquelme J, Lijavetzky D, Martinez-Zapater JM, Carmona MJ (2009) Genome-Wide Analysis of MIKC^C-Type MADS Box Genes in Grapevine. *Plant physiol.* 149:354–369
- Drews GN, Bowman JL, Meyerowitz EM (1991) Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. *Cell* 65:991–1002
- Duan W, Song X, Liu T, et al. (2015) Genome-wide analysis of the MADS-box gene family in *Brassica rapa* (Chinese cabbage). *Mol. Genet. Genomics* 290:239–255
- Eckardt NA (2003) MADS Monsters: Controlling Floral Organ Identity. *PLANT CELL ONLINE* 15:803–805
- Egea-Cortines M, Saedler H, Sommer H (1999) Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *EMBO J.* 18:5370–9
- Fan CM, Wang X, Wang YW et al. (2013) Genome-Wide Expression Analysis of Soybean MADS Genes Showing

- Potential Function in the Seed Development. (T Zhang, Ed.). PLoS ONE 8:e62288
- Fischer A, Baum N, Saedler H, Theißen G (1995) Chromosomal mapping of the MADS-box multigene family in *Zea mays* reveals dispersed distribution of allelic genes as well as transposed copies. *Nucleic Acids Res.* 23:1901–1911
- Folter S de, Immink RGH, Kieffer M, et al. (2005) Comprehensive Interaction Map of the Arabidopsis MADS Box Transcription Factors. *Plant Cell Online* 17:1424–1433
- Fornara F, Parenicová L, Falasca G, et al. (2004) Functional Characterization of OsMADS18, a Member of the AP1/SQUA Subfamily of MADS Box Genes. *Plant Physiol* 135:2207–2219
- Gan Y, Filleur S, Rahman A, Gotensparre S, Forde BG (2005) Nutritional regulation of ANR1 and other root-expressed MADS-box genes in *Arabidopsis thaliana*. *Planta* 222:730–742
- 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.* 19:801–14
- Hepworth SR, Valverde F, Ravenscroft D, Mouradov A, Coupland G (2002) Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs. *The EMBO J.* 21:4327–37
- Hsu HF, Hsieh WP, Chen MK, Chang YY, Yang CH (2010) C/D Class MADS Box Genes from Two Monocots, Orchid (*Oncidium Gower Ramsey*) and Lily (*Lilium longiflorum*), Exhibit Different Effects on Floral Transition and Formation in *Arabidopsis thaliana*. *Plant Cell Physiol.* 51:102
- Hsu HF, Hsu WH, Lee YI, et al. (2015) Model for perianth formation in orchids. *Nature Plants* 1:15046
- Hu L, Liu S, Somers DJ (2012) Genome-wide analysis of the MADS-box gene family in cucumber. *Genome* 55:245–256
- Immink RG, Tonaco IA, Folter S de, et al. (2009) SEPALLATA3: the ‘glue’ for MADS box transcription factor complex formation. *Genome Biol.* 10:R24
- Irish VF, Litt A (2005) Flower development and evolution: gene duplication, diversification and redeployment. *Curr. Opin. Genet. Dev.* 15:454–460
- Irish VF, Sussex IM (1990) Function of the *apetala-1* gene during Arabidopsis floral development. *Plant cell* 2:741–53
- Jofuku KD, Boer BG den, Montagu M Van, Okamura JK (1994) Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6:1211–1225
- Kang IH, Steffen JG, Portereiko MF, Lloyd A, Drews GN (2008) The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant cell* 20:635–47
- Kramer EM, Dorit RL, Irish VF (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. *Genetics* 149:765–83
- Li C, Wang Y, Xu L, et al. (2016) Genome-Wide Characterization of the MADS-Box Gene Family in Radish (*Raphanus sativus* L.) and Assessment of Its Roles in Flowering and Floral Organogenesis. *Front. Plant Sci.* 07:1390
- Lin CS, Hsu CT, Liao DC, et al. (2016) Transcriptome-wide analysis of the MADS-box gene family in the orchid *Erycina pusilla*. *Plant Biotechnol. J.* 14:284–298
- Liu C, Chen H, Er HL, et al. (2008) Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis. *Development* 135:1481–1491
- Lu ZX, Wu M, Loh CS, Yeong CY, Goh CJ (1993) Nucleotide sequence of a flower-specific MADS box cDNA clone from orchid. *Plant Mol. Biol.* 23:901–904
- Malcomber ST, Kellogg EA (2005) SEPALLATA gene diversification: brave new whorls. *Trends Plant Sci.* 10:427–435
- Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant cell* 23:865–72
- Masiero S, Colombo L, Grini PE, Schnittger A, et al. (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant cell* 23: 865–872
- Meyerowitz E, Bowman J, Brockman L (1991) A genetic and molecular model for flower development in Arabidopsis thaliana. *Dev. Suppl.* 1 157–161
- Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant cell* 11:949–56
- Mondragón-Palomino M, Nter Theißen G (2009) Why are orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes. *Annals of Botany* 104:583–594
- Mondragón-Palomino M, Theißen G (2008) MADS about the evolution of orchid flowers. *Trends Plant Sci.* 13:51–59
- Mondragón-Palomino M, Theißen G (2011) Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the ‘orchid code’. *Plant J.* 66:1008–1019
- 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 type II MADS-box genes in angiosperms. *Proc. Natl. Acad. Sci.* 101:1910–1915
- 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
- Ohmori S, Kimizu M, Sugita M, et al. (2009) MOSAIC FLORAL ORGANS1, an AGL6-like MADS box gene, regulates floral organ identity and meristem fate in rice. *Plant cell* 21: 3008–25
- Parenicová L, Folter S de, Kieffer M, et al. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant cell* 15:1538–51
- Passmore S, Maine GT, Elble R, Christ C, Tye B-K (1988) *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MATa cells. *J Mol. Biol.* 204:593–606
- Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405:200–203

- Pnueli L, Abu-Abeid M, Zamir D, Nacken W, Schwarz-Sommer Z, et al. (1991) The MADS box gene family in tomato: temporal expression during floral development, conserved secondary structures and homology with homeotic genes from *Antirrhinum* and *Arabidopsis*. *Plant J.* 1:255–266
- Portereiko MF, Lloyd A, Steffen JG, Punwani JA, Otsuga D, Drews GN (2006) AGL80 Is Required for Central Cell and Endosperm Development in *Arabidopsis*. *Plant Cell Online* 18:1862–1872
- Riechmann JL, Krizek BA, Meyerowitz EM (1996) Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl. Acad. Sci. U. S. A.* 93:4793–8
- Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. *Biol. Chem.* 378:1079–101
- Rouse DT, Sheldon CC, Bagnall DJ, Peacock WJ, Dennis ES (2002) FLC, a repressor of flowering, is regulated by genes in different inductive pathways. *The Plant journal: for cell and molecular biology* 29:183–91
- Sharma A, Shumayla, Tyagi S, Alok A, Singh K, Upadhyay SK (2019) Thaumatin-like protein kinases: Molecular characterization and transcriptional profiling in five cereal crops. *Plant Science* 290 <https://doi.org/10.1016/j.plantsci.2019.110317>
- Shimeld SM (1999) Gene function, gene networks and the fate of duplicated genes. *Semin. Cell Dev. Biol.* 10:549–553
- Shore P, Sharrocks AD (1995) The MADS-box family of transcription factors. *Eur. J. Biochem.* 229:1–13
- Shu Y, Yu D, Wang D, Guo D, Guo C (2013) Genome-wide survey and expression analysis of the MADS-box gene family in soybean. *Mol. Biol. Rep.* 40:3901–3911
- Skipper M, Johansen LB, Pedersen KB, Frederiksen S, Johansen BB (2006) Cloning and transcription analysis of an AGAMOUS- and SEEDSTICK ortholog in the orchid *Dendrobium thyrsiflorum* (Reichb. f.). *Gene* 366:266–274
- Sommer H, Beltrán JP, Huijser P, et al. (1990) Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* 9:605–13
- Song JJ, Nakamura T, Fukuda T, et al. (2006) Spatiotemporal expression of duplicate AGAMOUS orthologues during floral development in *Phalaenopsis*. *Dev. Genes Evol.* 216:301–313
- Southerton SG, Marshall H, Mouradov A, Teasdale RD (1998) Eucalypt MADS-Box Genes Expressed in Developing Flowers. *Plant Physiol.* 118:365–372
- Tapia-Lopez R, Garcia-Ponce B, Dubrovsky JG, et al. (2008) An AGAMOUS-Related MADS-Box Gene, XAL1 (AGL12), Regulates Root Meristem Cell Proliferation and Flowering Transition in *Arabidopsis*. *Plant physiol* 146:1182–1192
- Theissen G (2001) Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4:75–85
- Theissen G, Becker A, Rosa A Di, et al. (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.* 42:115–49
- 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, Saedler H (2001) Plant biology. Floral quartets. *Nature* 409:469–71
- Tian Y, Dong Q, Ji Z, Chi F, Cong P, Zhou Z (2015) Genome-wide identification and analysis of the MADS-box gene family in apple. *Gene* 555:277–290
- Tsai WC, Kuoh CS, Chuang MH, Chen WH, et al. (2004) Four DEF-Like MADS Box Genes Displayed Distinct Floral Morphogenetic Roles in *Phalaenopsis* Orchid. *Plant Cell Physiol.* 45:831–844
- Tsai WC, Lee PF, Chen HI, Hsiao YY, Wei WJ, et al. (2005) PeMADS6, a GLOBOSA/PISTILLATA-like gene in *Phalaenopsis equestris* involved in petaloid formation, and correlated with flower longevity and ovary development. *Plant Cell Physiol.* 46: 1125–1139
- Wang SY, Lee PF, Lee YI, et al. (2011) Duplicated C-Class MADS-Box Genes Reveal Distinct Roles in Gynostemium Development in *Cymbidium ensifolium* (Orchidaceae). *Plant Cell Physiol.* 52:563–577
- Wei B, Zhang RZ, Guo JJ, et al. (2014) Genome-Wide Analysis of the MADS-Box Gene Family in *Brachypodium distachyon*. (S Lin, Ed.). *PLoS ONE* 9:e84781
- Wells CE, Vendramin E, Jimenez Tarodo S, Verde I, Bielenberg DG (2015) A genome-wide analysis of MADS-box genes in peach [*Prunus persica* (L.) Batsch]. *BMC Plant Biol.* 15:41
- Xu Y, Teo LL, Zhou J, Kumar PP, Yu H (2006) Floral organ identity genes in the orchid *Dendrobium crumenatum*. *Plant J* 46:54–68
- Yan L, Wang X, Liu H, et al. (2015) The Genome of *Dendrobium officinale* Illuminates the Biology of the Important Traditional Chinese Orchid Herb. *Mol. Plant* 8:922–934
- 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
- Yu H, Goh CJ (2000) Identification and characterization of three orchid MADS-box genes of the AP1/AGL9 subfamily during floral transition. *Plant physiol.* 123:1325–36
- Zhang GQ, Liu KW, Li Z, et al. (2017) The *Apostasia* genome and the evolution of orchids. *Nature Publishing Group* 549
- Zhang GQ, Xu Q, Bian C, et al. (2016) The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Scientific Reports* 6:19029
- Zhang H, Forde BG (2000) Regulation of *Arabidopsis* root development by nitrate availability. *J. Exp. Bot.* 51:51–59
- Zhao T, Ni Z, Dai Y, Yao Y, Nie X, Sun Q (2006) Characterization and expression of 42 MADS-box genes in wheat (*Triticum aestivum* L.). *Mol. Genet. Genomics* 276:334–350
- Zhao Y, Li X, Chen W, et al. (2011) Whole-genome survey and characterization of MADS-box gene family in maize and sorghum. *Plant Cell, Tissue Organ Cult.* 105:159–173
- Zobell O, Faigl W, Saedler H, Munster T (2010) MIKC* MADS-Box Proteins: Conserved Regulators of the Gametophytic Generation of Land Plants. *Mol. Biol. Evol.* 27:1201–1211