

# Cloning, Structural Characterization, and Phylogenetic Analysis of Flower MADS-Box Genes from Crocus (*Crocus sativus* L.)

Athanasios S. Tsaftaris<sup>1,2,\*</sup>, Alexios N. Polidoros<sup>1</sup>, Konstantinos Pasentsis<sup>1</sup>, and Apostolos Kalivas<sup>2</sup>

<sup>1</sup>Institute of Agrobiotechnology, Center for Research and Technology Hellas, Themi, Greece, GR-570 01; <sup>2</sup>Department of Genetics and Plant Breeding, Aristotle University of Thessaloniki, Thessaloniki, Greece, GR-541 24

E-mail: [tsaft@certh.gr](mailto:tsaft@certh.gr), [alexios@certh.gr](mailto:alexios@certh.gr), [kpasents@certh.gr](mailto:kpasents@certh.gr), [kalivasapostolis@yahoo.gr](mailto:kalivasapostolis@yahoo.gr)

Received November 15, 2006; Revised April 3, 2007; Accepted April 18, 2007; Published June 22, 2007

**Crocus (*Crocus sativus* L.) is a crop species cultivated for its flowers and, more specifically, for its red stigmas. The flower of crocus is bisexual and sterile, since crocus is a triploid species. Its perianth consists of six petaloid tepals: three tepals in whorl 1 (outer tepals) and three tepals in whorl 2 (inner tepals). The androecium consists of three distinct stamens and the gynoecium consists of a single compound pistil with three carpels, a single three-branched style, and an inferior ovary. The dry form of the stigmas constitutes the commercial saffron used as a food additive, in the coloring industry, and in medicine. In order to uncover and understand the molecular mechanisms controlling flower development in cultivated crocus and its relative wild progenitor species, and characterize a number of crocus flower mutants, we have cloned and characterized different, full-length, cDNA sequences encoding MADS-box transcription factor proteins involved in flower formation.**

Here we review the different methods followed or developed for obtaining these sequences involving conventional 5' 3' RACE, as well as newly developed methods from our group, named Rolling Circle Amplification – RACE (RCA-RACE) and its modification named familyRCA-RACE (famRCA-RACE). Furthermore, the characteristics of the protein structure and their common and specific domains for each type of MADS-box transcription factors in this lower nongrass monocot belonging to the Iridaceae family are described. Finally, a phylogenetic tree of all the MADS-box sequences available in our lab is presented and discussed in relation to other data from studies of species of the Iridaceae group and closely related families from an evolutionary perspective. The structural and phylogenetic analyses are based on both published and unpublished data.

**KEYWORDS:** *Crocus sativus* L., monocots, MADS-box genes

## INTRODUCTION

Flowering plants represent one of the most successful and diverse groups of organisms on the planet, with more than 250,000 extant species in the wild and many more varieties generated by horticulturists through hybridization and other breeding efforts. The majority of flowers possess four types of floral organs: two outer whorls of sterile organs, the sepals and petals (also known as the perianth), and two inner whorls of fertile organs, the male stamens and female carpels, with the carpels positioned centrally. Although the fundamental characteristics of angiosperm flowers are generally conserved, the enormous morphological diversity suggests a high degree of plasticity in the genetic control of floral development. Variation is observed in every aspect of floral architecture, including phyllotaxy, merosity, floral symmetry, and floral organ identity. In-depth analyses of model species, such as *Arabidopsis thaliana* and *Antirrhinum majus*, have contributed significantly to our understanding of the genetic pathways that control these morphological components. By using this work as a foundation for comparative studies, a picture is gradually coming into focus of how alterations in floral genetic programs have contributed to the evolution of floral architecture.

Forward mutagenesis studies of these two model species uncovered an intriguing series of homeotic floral mutants[1,2,3]. In both taxa, the mutants appeared to fall into similar classes: mutations that affected sepal and petal identity were placed into what was termed the “A” class; those that affected petal and stamen identity, the “B” class; and those that affected stamen and carpel identity, the “C” class. For instance, B mutants exhibited the transformation of petals into sepals and stamens in carpels[2,4]. Analysis of double and triple mutants[5] led to the proposition of a simple and elegant model that explained the major aspects of genetic interactions among the loci. This became known as the ABC model[3]. Fundamentally, the ABC model holds that the overlapping domains of three classes of gene activity, referred to as A, B, and C, produce a combinatorial code that determines floral organ identity in successive whorls of the developing flower. Another critical component of the ABC program is that A and C functions are mutually exclusive[5], such that elimination of C gene activity causes the A domain to expand and vice versa[6,7]. Additional studies of mutants and overexpressing lines have largely confirmed the model and demonstrated the completely homeotic nature of this developmental program[5,8,9]. As our understanding of the genes involved in the program has grown, the model has expanded as well to what is now referred to as the ABCDE model[10]. “D” class genes were proposed as ovule-identity genes based on work done in *Petunia hybrida*[11], while “E” class genes function broadly across the floral meristem to facilitate the function of many of the original ABC loci[12,13]. With the exception of AP2, all of the organ-identity genes identified to date are members of the eukaryotic MADS transcription factor family[14,15]. MADS-box genes are characterized by the highly conserved MADS-box domain and can be divided into the type I and type II main lineages that are present in plants, animals, and fungi[16,17]. These lineages differ in the amino acid sequence of the MADS-box as well as in the domain structure of the predicted protein. Most type II proteins exhibit a typical MIKC structure, where the MADS domain is followed by a short I (intervening) domain, a well-conserved K (keratin-like) domain, and a variable C-terminal region, while type I proteins lack the K domain, forming a structure of a MADS-box followed by a rather undefined and length-variable C domain[17].

Almost all the MIKC-type MADS-box genes of plants can be subdivided into 12 major clades, each of which includes one to six paralogs from *Arabidopsis* and putative orthologs from other seed plants[14]. Genes of the different A, B, C, D, and E classes related to flower development fall into separate phylogenetic clades that are known as APETALA1/FRUITFULL (AP1/FUL) or SQUAMOSA (SQUA) (class A), APETALA3 (AP3) and PISTILLATA (PI) or DEFICIENS (DEF) and GLOBOSA[18] (class B), AGAMOUS (AG) (classes C and D), and AGAMOUS-like6 (AGL6) and SEPALLATA (SEP) (class E)[14]. Not all the genes in the above-mentioned clades determine floral organ identity. Genes from other clades may be related to flowering as well through their function in determining flower initiation and meristem identity. These are the FLOWERING LOCUS C-like (FLC-like) clade with the *Arabidopsis* genes FLC and AGL27 that are inhibitors of flowering, the STMADS11-like clade with the genes AGL24 and SHORT VEGETATIVE PHASE (SVP) that are repressors of flowering in vegetative tissues, and the Tomato MADS-box gene 3-like (TM3-like) clade with SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) that is involved in regulation of flowering time[14].

Although monocot flowers contain stamens and carpels, they differ from eudicot flowers in the type of organs that are present in the outer whorls. Liliaceae family members often have two outer whorls of showy petal-like tepal organs, whereas grass flowers have paleas, lemmas, and lodicules instead of sepals and petals. A modified ABC model in which a B function is present in whorls 1, 2, and 3 has been proposed to explain the presence of tepals in Liliaceae flowers[19].

Studying flower development is not only significant for improving our understanding of basic regulatory mechanisms of flower initiation and organ identity, but could have practical applications in crops cultivated for their flowers. Crocus is an example of such a crop with flowers of economic importance, and we have set up a study to understand and possibly improve the crocus flower. Crocus is a monocot triploid sterile species belonging to the Iridaceae family, whose red stigmatic styles constitute saffron, a popular food additive with a delicate aroma and attractive color. Saffron also has medicinal properties and is used in the coloring industry. Different wild-type species like *Crocus cartwrightianus*, *C. hadriaticus*, and *C. oreoreticus* have been proposed as progenitor species, giving rise to the triploid sterile cultivated *C. sativus*, but no conclusive molecular data have been obtained, yet, to identify the two progenitor species with certainty.

The flower of crocus is bisexual and sterile. The perianth consists of six petaloid tepals: three tepals in whorl 1 (outer tepals) and three tepals in whorl 2 (inner tepals). The androecium consists of three distinct stamens and the gynoecium consists of a single compound pistil with three carpels, a single three-branched style, and an inferior ovary. Several phenotypic flower mutants have been described, such as flowers with larger numbers of styles and stamens as well as flowers without stamens[20,21]. Crocus blooms only once a year and is hand harvested. After mechanical separation of the tepals, the stigmas are hand separated from the carpels and dried. The size and the amount of individual stigmas collected from each flower influence the total yield and quality of saffron. Between 70,000 and 200,000 flowers are needed to produce 1 kg of dried saffron, which equates to 370–470 h of work. Consequently, the cultivation of this crop for its flowers and specifically its stigmas is very labor intensive, leading to high costs[22]. Thus, understanding flower development in crocus could reveal ways to increase yield and lower production costs since flowers and, more specifically, isolated stigmas comprise the valuable commercial part of the plant. With this goal, we have cloned, characterized, and studied the expression of all types of MADS-box genes from crocus involved in flower formation. Obtaining a similar type of molecular information for the three putative progenitor species, *C. cartwrightianus*, *C. hadriaticus*, and *C. oreoreticus*, and comparing them with the cultivated *C. sativus* genes, could contribute to the identification of its diploid progenitor species.

The different methods used to clone the sequences of MADS-box genes expressed in the flower of crocus are reviewed and their comparative structural and phylogenetic relationships are described.

## CLONING THE MADS-BOX SEQUENCES FROM CROCUS

### The Use of Conventional 5' 3' RACE

Table 1 displays homologues of the major A, B, C, D, and E types of MADS-box genes from crocus; the two dicot model species, *Arabidopsis* and *Antirrhinum*; and the monocot grass model, crop rice. The number of homologues for each type of MADS-box sequences obtained from crocus and the methods followed to obtain each sequence are also indicated.

In their recent analysis of all the MADS-box genes of *Arabidopsis* flower, Wellmer et al.[23] found that 17 out of 39 MIKC-type MADS-box genes were shown to be up-regulated in early flower development. We do not know how many MADS-box genes exist in the crocus genome, but as it is an allotriploid species[24], we expect this number to be much higher in crocus. Isolation of full-length gene transcripts is important to determine the protein-coding region and study gene structure. However, isolation of novel gene sequences is often limited to expressed sequence tags (ESTs) (i.e., short cDNA fragments that predominantly represent the 3' end of the transcript). Rapid amplification of cDNA ends

(RACE) is by far the most popular approach for obtaining full-length cDNAs when only part of the transcript's sequence is known. Since its original description[25,26], numerous modifications and

**TABLE 1**  
**Homologues of the Major A, B, C, D, E Types of MADS-box Genes from *Crocus*, the Two Dicots Model Species *Arabidopsis* and *Antirrhinum* and the Monocot Model, Rice\*.**

Gene class	<i>Arabidopsis</i>	<i>Antirrhinum</i>	<i>Oryza</i>	<i>Crocus</i>	Number of <i>Crocus</i> homologues sequenced	Cloning method ( <i>Crocus</i> )	Reference ( <i>Crocus</i> )
A	<i>APETALA1</i>	<i>SQUAMOSA</i>	<i>RAP1</i>	<i>CsatAP1/FUL</i>	3 2	5' 3' RACE and famRCA- RACE	[22] [39]
B	<i>APETALA3</i> <i>PISTILLATA</i>	<i>DEFICIENS</i> <i>GLOBOSA</i>	<i>SUPERWOMAN</i> 1 <i>OsMADS2</i> , <i>OsMADS4</i> <i>OsMADS3</i>	<i>CsatAP3</i> <i>CsatPI</i>	2 5	5' 3' RACE 5' 3' RACE	[27] [29]
C	<i>AGAMOUS</i>	<i>FARINELI</i>	<i>OsMADS3</i>	<i>CsatAG</i>	1	5' 3' RACE	[28]
D	<i>SEEDSTICK</i>	<i>Nk**</i>	<i>OsMADS13</i>	<i>nk</i>	-	-	-
E	<i>SEPALLATA1</i> , 2, 3  <i>AGAMOUS- LIKE6</i>	<i>DEFH49</i> , 72	<i>LHS1</i> , <i>OsMADS5</i> , 34	<i>CsatSEP3</i>  <i>CsatAGL6</i>	4  2	famRCA- RACE  5' 3' RACE and famRCA- RACE	[39]  [21]

\*The number of *Crocus* homologues for each type of MADS-box sequence obtained from *Crocus* and method followed to obtain each *Crocus* sequence is also indicated.

\*\*nk: not known.

improvements of the method have been developed that consist of a collection of PCR-based cloning procedures that extend a known cDNA fragment toward the 3' (3' RACE) or the 5' (5' RACE) cDNA end. The original method is based on attachment of an anchor sequence to one end of the cDNA that can be used as a primer-binding template in a PCR with a second gene-specific primer from the known part of the gene.

As shown in Table 1, most of the original *Crocus* sequences were obtained using the typical 5' and 3' RACE method. Since no ESTs are yet available from *Crocus* flower for use as primers for the 5' 3' RACE, degenerate primers for conserved domains, specific for each type of *Crocus* MADS-box type genes, were designed. For example, using 5' 3' RACE for obtaining *CsatAP3* sequences[27], a degenerate primer, corresponding to conserved amino acid sequence of the MADS-box genes, was used in 3' RACE experiments with an dT Adaptor Primer. The synthesized cDNA was used as template in a touchdown PCR reaction with the degenerate primer and an Abridged Universal Amplification primer. Several products between 500 and 900 bp were cloned and sequenced, and based on the sequence information obtained by the 3' RACE experiments, two gene-specific primers were designed from the 3-UTR and used to isolate the cDNA 5' ends following the recommendations of the manufacturer[27]. For more details as to how the type of sequences using 5' 3' RACE were obtained, see the original publications for *CsatAP3*[27], *CsatAPI/FUL*[22], *CsatAG*[28], *CsatPI*[29], and *CsatAGL-6*[21]. For *CsatAPI/FUL*[22] and *CsatAG*[28], we used the abbreviation Cs as prefix to characterize the *Crocus* genes, but since this caused some confusion because MADS-box genes from *Chloranthus spicatus* have been reported using the Cs prefix, we decided to use the Csat prefix for the *Crocus* genes from now on and will update the GeneBank records, accordingly.

## The Development and Use of RCA-RACE Methods

The 5' 3' RACE method is time consuming and will take several years of work to identify the anticipated large number of MADS-box genes operating during crocus flowering and flower formation. Furthermore, 5' 3' RACE is technically difficult and usually requires substantial optimization and several repetitions before satisfactory results can be obtained[30]. Using a universal primer corresponding to the anchor sequence present in all cDNAs results in a high background of nonspecific products even after a nested PCR with a gene-specific primer internal to the first gene-specific primer is performed. Another drawback of the method is the difficulty of obtaining the full-length 5' end of the transcript due to the presence of many truncated transcripts in the messenger RNA (mRNA) pool. Several strategies aimed at eliminating these problems have been developed, some of which require the generation of double-stranded cDNA, including the use of template-switching reverse transcription[31] or a postreverse transcription adaptor ligation step[32]. Methods that are performed directly on first-strand cDNA are complicated by the low efficiency of RNA ligase for the circularization reaction[33] or the need for bridging oligonucleotides for this step[34]. Furthermore, existing inverse-RACE methods typically require nested PCR to amplify the transcript of interest, and only a limited number of transcripts can be isolated from a single reverse transcription reaction, making it difficult to analyze rare transcripts from scarce tissue.

We have recently described an improved inverse-RACE method, named Rolling Circle Amplification – RACE (RCA-RACE)[35] (Fig. 1). The process takes advantage of the properties of CircLigase™ (Epicentre Biotechnologies, Madison, WI) to circularize single-stranded cDNA molecules via an intramolecular link for cDNA circularization, followed by rolling circle amplification[36] of the circular cDNA with  $\phi$ 29 DNA polymerase. In this way, a large amount of the PCR template is produced, allowing the simultaneous isolation of the 3' and 5' unknown ends of a virtually unlimited number of transcripts after a single reverse transcription reaction. To prove the concept, the method was used to isolate a previously characterized transcript: the crocus *API/FUL*-like MADS-box gene, which represents transcript coding for a rare transcript[27]. Subsequently, the method was also used to clone MADS-box genes from the peach (*Prunus persica*) fruit[37] and a number of genes from other crops.

We have also developed a modification of the RCA-RACE method named familyRCA-RACE (famRCA-RACE) that allows the simultaneous isolation of members of a family of homologous genes in one RCA-RACE reaction by using degenerate primers corresponding to a conserved amino acid domain in antisense orientation. To prove the concept of the proposed methodology, we used it to isolate members of the plant-specific family of NAC transcription factors sequences expressed in crocus flower[38]. The method was also used to obtain members of the MADS-box family of sequences from crocus. For obtaining the sequences of *CsatSEP3* (*CsatSEP3A,B,C,D*), for example[39], and *CsatAPI/FUL* (*CsatAPIFULd,e*) using famRCA-RACE, first-strand cDNA synthesis on total RNA extracted from crocus flowers was purified, circularized, and amplified with  $\phi$ 29 DNA polymerase.

For obtaining MADS-box sequences out of this circular cDNA library, the  $\phi$ 29 amplified library was used as template in a PCR reaction with two degenerate primers, corresponding to conserved amino acid sequences of the MADS-box genes and designed in antisense orientation. Several individual clones were screened for the presence of an insert and sequenced giving MADS-box-containing inserts. Finally, based on the sequence information obtained from this famRCA-RACE experiment, gene-specific primers were designed and used for PCR to obtain the full-length coding sequences of the four *CsatSEP3*-like and *CsatAPI/FUL*-like genes mentioned in Table 1. Five additional clones with sequences identical to previously obtained sequences of *CsatAGL6a* and *CsatAGL6b* by 5' 3' RACE were also obtained, whereas experiments are underway to complete the sequence of one more *CsatAG*-like gene. The famRCA-RACE looks promising for eventually obtaining the full range of MADS-box sequences present in cultivated *Crocus sativus*. Following the same method and using the same degenerate primers, we can compare them with those of the putative progenitor diploid species.



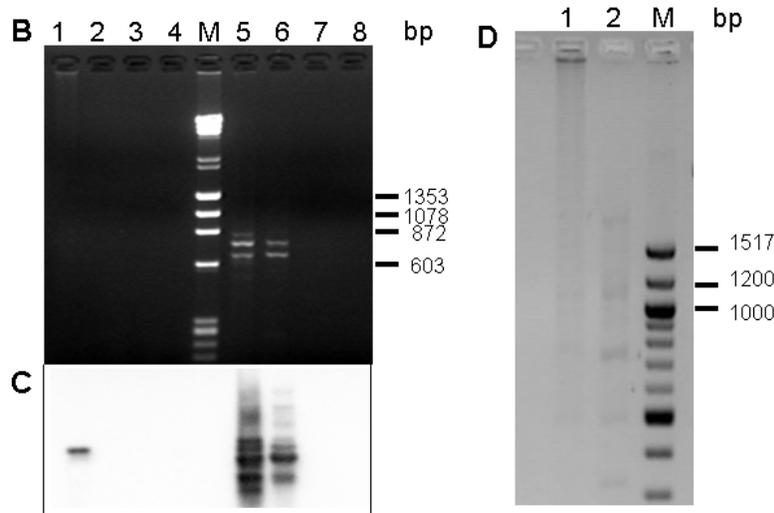
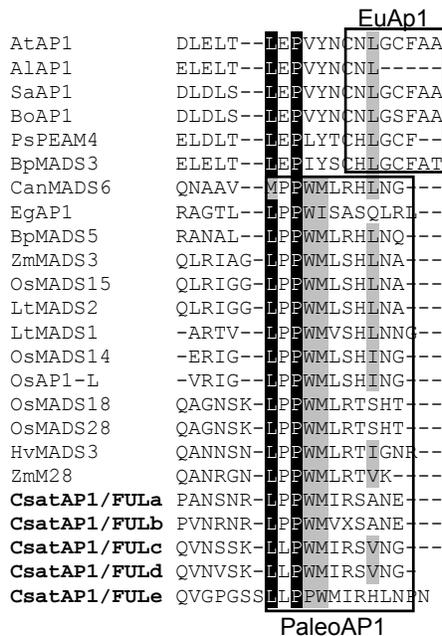


FIGURE 1B-D

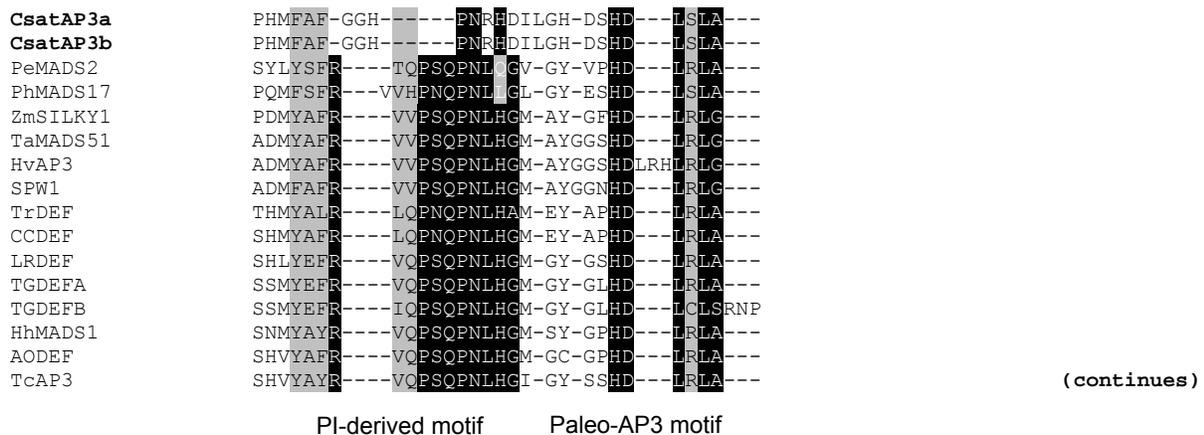
only when the MADS transcription factor proteins are dimerized, which is primarily mediated by the adjacent I and K dimerization domains, also preserved in all *crocus* MADS proteins isolated. On the contrary, the C domain on the carboxyl protein terminal is relatively more variable in both the length and the sequences involved, but lineage-specific, short, highly conserved sequence motifs are also present in the C domain indicative of its significance functional specificity[41]. Alignments constructed using *Clustal W*[42] of the C-terminal motifs that are characteristic for deep branching of the MADS proteins in each lineage are collectively shown in Fig. 2 for all the *crocus* MADS proteins isolated, together with homologues plant MADS proteins. For each lineage of the MADS proteins, characteristic C-terminus motifs are boxed according to Vandebussche et al.[43] and Kramer et al.[44], while the *crocus* sequences are in bold.

More specifically, as shown in Fig. 2, the C-terminal domain of AP1/FUL predicted proteins is highly variable, as is characteristic of plant MADS-domain-containing proteins. Nonetheless, there is a strongly conserved hydrophobic six-amino-acid motif (consensus LPPWML) at the end of all FUL-like and euFUL proteins. This FUL-like motif can be seen in SEP and AGL6 sequences, although the exact residue composition is not strictly conserved. The high degree of conservation of this motif is a strong indication that it is functionally important and suggests that its loss and replacement with a different motif in euAP1 proteins may result in altered functional capabilities of the euAP1 proteins[45]. Published proteins belonging to the AP1/FUL lineage, which showed a high degree of homology to the *crocus* *CsatFUL/AP1*, were selected for the multiple alignment process as previously described[27]. In this comparison, the three *crocus* homologues A, B, C presented 87, 82, 84% identity and 92, 92, 91% similarity to the consensus FUL/AP1-like proteins MADS-box domain, and 64, 59, 51% identity and 79, 77, 76% similarity to the consensus K-box of the AP1/FUL -like proteins, respectively. High identity and similarity percentages were calculated between all the three homologues and the maize *ZmM28* and the rice *OsMADS15* proteins. The amino acid sequence of *API* in dicots has a characteristic EuAP1 C-terminal motif and many dicot homologues terminate in CFAA, a typical CaaX box recognition motif for farnesyltransferase (FTase). *API* is a target of FTase and farnesylation alters the function and perhaps specificity of the transcription factor[46]. The five *crocus* *API/FUL*-like genes lack a CaaX box at the C-terminus, since monocot genes in this lineage terminate with a paleoAP1 motif characteristic for dicot *FRUITFULL (FUL)* genes (previously called *AGL8* in *Arabidopsis*). For this reason, we also decided to rename the previously published *CsatAPI/FUL*[27] to *CsatAPI/FUL*. *Arabidopsis FUL* is weakly expressed in rosette leaves during vegetative development and is subsequently strongly up-regulated in the shoot apex on the transition to flowering. Experimental evidence suggests that *FUL* regulates the transcription of genes required for cellular differentiation during fruit and leaf development[16].

APETALA1 – FRUITFULL



APETALA3



**FIGURE 2** Alignment of the deduced amino acid (aa) sequences of the C-terminus of crocus MADS-box genes and other homologous plant MADS proteins constructed with *Clustal W*[42]. Identical amino acids are highlighted in black and similar amino acids in gray with a threshold for shading of 70% sequence similarity. For each lineage of the MADS-box proteins, characteristic motifs are boxed. Motifs are described in Vandenbussche et al.[43] and Kramer et al.[44]. Crocus sequences are indicated in bold (see text for details).

*AP3*-like genes in higher plants have conserved C-terminal motifs, which in most higher eudicots is called the eu*AP3* motif, and in lower eudicots and magnoliids is called the paleo*AP3* motif[41]. The ancestral paleo*AP3* motif was retained in the so-called TM lineage of paralogous core eudicot lineages[47]. Detailed comparison between the paleo*AP3* and the eu*AP3* motif revealed that the latter could have been derived by an 8-bp insertion at the C-terminus of the former, causing a frame shift mutation beyond the insertion site in eu*AP3* genes[43]. The paleo*AP3* motif is also conserved in many B-class *AP3*-like proteins in monocots, so far examined including lily, asparagus, and tulip[48,49,50]. This led to the suggestion that all monocot *AP3*-like proteins have a paleo*AP3* motif[50], which is in agreement with recent hypotheses

PISTILLATA

TaPI1	RDF	T...	S	QMP	FT	FRL	QPS	H	PNL	Q	E	D	K	208
AoGLOB	GAT	A...	T	HMP	MA	FRV	Q	E	I	Q	P	N	L	Q
BvMADS2	.	DYN	...	S	QMP	F	FRV	Q	E	I	Q	P	N	L
LrGLOA	GDF	S...	S	QMP	MA	FRV	Q	E	I	Q	P	N	L	Q
LrGLOB	GEY	E...	G	H	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
MdMADS	GN	YNN	N	Q	Q	P	F	A	FRV	Q	E	I	Q	P
OsMADS2	RDF	A...	A	QMP	IT	FRV	Q	P	S	H	P	N	L	Q
OsMADS4	RDF	A...	A	SMP	FT	FRV	Q	P	S	H	P	N	L	Q
Phfbp3	.	DYEY	...	Q	MP	F	AL	RV	Q	P	M	Q	P	N
PhMADS2	.	DYEY	...	Q	MP	F	AL	RV	Q	P	M	Q	P	N
SmPI	RDF	T...	A	FV	P	F	G	FRV	Q	P	M	Q	P	N
AoGLOA	TDY	A...	S	EV	Q	M	A	L	G	V	Q	P	M	Q
TaPI2	RDF	A...	A	QMP	IT	FRV	Q	P	S	H	P	N	L	Q
TgGLO	RE	N...	S	QMP	M	T	FRV	Q	E	I	Q	P	N	L
Zm16	RDL	A...	A	QMP	IT	FRV	Q	P	S	H	P	N	L	Q
Zm18	RDF	I...	S	QMP	F	T	FRV	Q	P	N	H	P	N	L
Zm29	RDF	I...	S	QMP	F	T	FRV	Q	P	N	H	P	N	L
<b>CsatPIA1</b>	RE	HA...	T	QMP	MA	F	T	V	Q	P	F	Q	P	N
<b>CsatPIA2</b>	RE	HA...	T	QMP	MA	F	T	V	Q	P	F	Q	P	N
<b>CsatPIB</b>	RAMP	...	S	QMP	I	T	FLV	Q	P	T	Q	P	N	L
<b>CsatPIC1</b>	RAMP	...	S	QMP	M	T	FRM	Q	P	T	Q	P	N	L
<b>CsatPIC2</b>	RAMP	...	S	QMP	M	T	FRM	Q	P	T	Q	P	N	L

PI- motif

AGAMOUS

VvMADS1	QP	D-S	N	L	E	V	N	L	L	D	P	N	-H	F	Y	S	R	---	H	D	O	T	A	L	Q	L	M	-----							
ZmAG1	TP	D	P	I	R	S	F	L	C	F	N	I	V	Q	P	-Q	E	Y	S	Q	---	Q	E	D	R	K	D	F	N	D	Q	G	G	R	-----
ZmM2	-P	F	D	-S	R	N	F	L	O	V	---	S	M	P	-Q	E	Y	S	H	---	Q	L	O	P	T	T	L	Q	-I	G	-----				
RrMASAKO-D1	QS	D--	R	S	F	L	E	V	-I	L	E	S	N	-H	F	Y	N	R	Q	-G	O	N	O	T	P	L	Q	L	M	-----					
PgAG2	QS	F	D	-G	R	N	I	T	Q	L	N	L	Q	P	N	-N	F	Y	S	---	R	O	D	O	T	A	L	Q	L	M	-----				
Ph-fbp6	QP	D	-A	R	N	F	L	E	V	N	L	L	D	P	N	-H	F	Y	S	R	---	Q	O	T	A	L	Q	L	M	-----					
PhAG1	QS	F	D	-A	R	N	F	L	O	V	N	L	Q	T	N	-N	F	Y	P	---	R	O	D	O	P	P	L	Q	L	M	-----				
PeMADS1	P	-F	D	-S	R	N	F	L	H	V	N	L	D	P	N	-D	H	F	Y	S	---	H	O	Q	O	T	A	L	Q	L	M	-----			
OsMADS3	NP	D	-S	R	N	F	L	O	V	N	I	M	Q	P	-Q	E	Y	A	H	---	Q	L	O	P	T	T	L	Q	-I	G	Q	P	A	F	N
NtAG1	HQ	F	D	-T	R	N	F	L	O	V	N	L	Q	T	N	-N	F	Y	T	---	R	O	D	O	P	S	L	Q	L	M	-----				
MpMADS2	G	-F	D	-S	R	N	F	L	O	V	N	L	D	S	-H	F	Y	S	---	H	O	E	O	T	A	L	Q	L	M	-----					
LeAG1	QQ	F	D	-T	R	N	F	L	O	V	N	L	Q	T	N	-N	F	Y	P	---	R	O	D	O	P	P	L	Q	L	M	-----				
HoAG1	Q	-F	D	-S	R	N	F	L	O	V	S	L	M	E	P	N	N	H	F	Y	S	R	---	Q	O	O	T	A	L	Q	L	M	-----		
AtAG1	QP	F	D	-S	R	N	F	L	O	V	A	A	L	Q	P	N	N	H	F	Y	S	S	A	G	R	O	D	O	T	A	L	Q	L	M	-----
GhGAGA2	QP	F	D	-G	R	N	F	L	O	V	N	D	L	Q	P	N	-N	N	Y	S	---	C	O	D	O	T	P	L	Q	L	M	-----			
GhGAGA1	QP	F	D	-G	R	N	F	L	Q	T	N	D	L	Q	P	N	-N	I	Y	S	---	C	O	D	O	T	P	L	Q	L	M	-----			
<b>CsatAG1a</b>	P	-F	D	-S	R	N	F	L	Q	A	N	L	V	D	P	N	-H	F	Y	S	---	H	O	Q	O	T	A	L	Q	L	M	-----			
<b>CsatAG1b</b>	P	-F	D	-S	R	N	F	L	Q	A	N	L	V	D	P	N	-H	F	Y	S	---	H	O	Q	O	T	A	L	Q	L	M	GFV	-----		
CaMADS1	QS	F	D	-S	R	N	F	L	O	V	D	A	L	Q	P	N	-H	F	Y	P	---	R	O	D	O	M	A	L	Q	L	M	-----			
BnAG1	QP	F	D	-S	R	N	F	L	O	V	A	A	L	Q	P	N	N	H	F	Y	S	S	A	G	R	O	D	O	T	A	L	Q	L	M	-----
HaMADS59	QP	F	D	-G	R	N	F	L	O	V	N	D	L	Q	P	N	-N	S	Y	S	---	C	O	D	O	T	P	L	Q	L	M	-----			

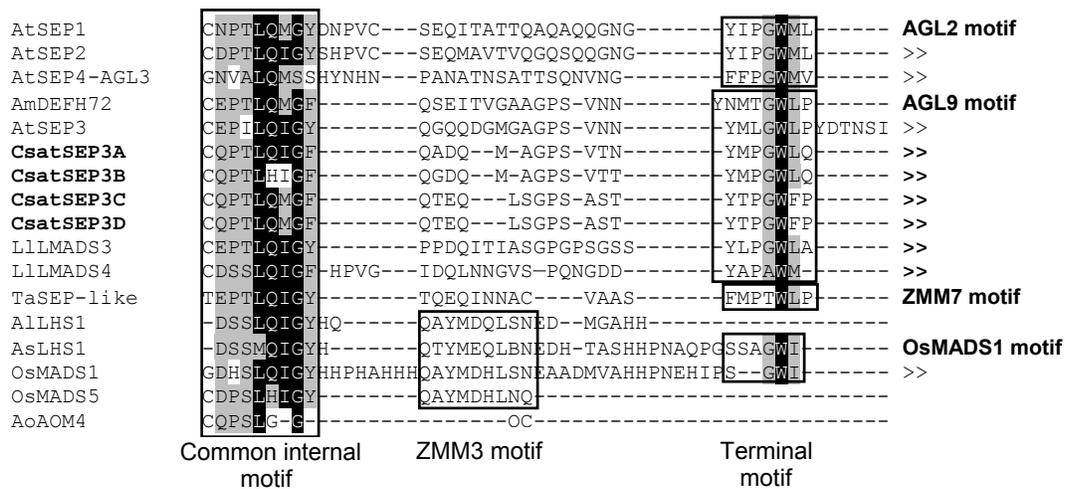
(continues)

AG motif I AG motif II

FIGURE 2 continued

on the phylogenetic position of monocots in basal angiosperms[51]. This hypothesis is supported by the presence of the paleoAP3 motif in the crocus AP3-like sequences (five out of eight identical amino acids with the consensus paleoAP3 sequence [52]). Another conserved motif in AP3-like proteins is the PI-derived motif, which is defined as a region bearing similarity with the conserved PI motif in the PI lineage of B-class proteins[41]. This region is less conserved in all AP3-like lineages including the paleoAP3 lineage of monocots, and the crocus sequence has only five out of 12 identical amino acids of the consensus PI-derived motif described previously[52]. Alignment of the predicted amino acid sequences CsatAP3a and CsatAP3b with the members of B-class MADS-box proteins used in phylogenetic analysis[27] revealed that

## SEPALATA



## AGAMOUS-LIKE 6

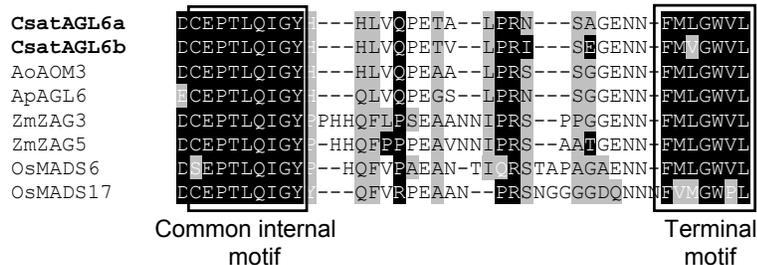


FIGURE 2 continued

the crocus sequences share high similarity with B-class genes in the conserved MIK region whereas they are more divergent in the variable C region. Within the C region, the paleoAP3 motif in the crocus AP3-like sequences (five out of eight identical amino acids with the consensus paleoAP3 sequence described in Tzeng and Yang[52]) could be identified. The PI-derived motif, which is defined as a region bearing similarity with the conserved PI motif in the PI lineage of B-class proteins[47] was less conserved (five out of 12 identical amino acids of the consensus PI-derived motif described in Tzeng and Yang[52]).

Based on the amino acid sequence similarity of the entire coding region, the five *CsatPI* homologous-deduced protein sequences can be assigned to the PI/GLO-like family of proteins[53]. Published genes that belong to the PI/GLO family of proteins and showed high degree of homology to the five *CsatPI* were selected for multiple alignment[29]. All five *CsatPI* proteins at the C-terminus include a PI-motif, which is conserved sequence in the *PI/GLO* family genes[47].

C-type MADS-box genes homologous to *AGAMOUS* (AG) from crocus revealed two transcripts designated *CsatAG1a* and *CsatAG1b*, which are different since *CsatAG1a* is missing 10 bp from 890 to 899 in comparison with *CsatAG1b*. This deletion alters the coding ORF of *CsatAG1a* in such a way that the deduced protein sequence of *CsatAG1a* is two aa shorter at the C-terminus from the deduced protein sequence of *CsatAG1b*. Alignment of C-terminal regions of predicted amino acid sequences for selected representatives of the C and D lineages and gymnosperm (Gymno) AG-like genes distinguished two highly conserved regions, AG motif I and AG motif II[44]. Both crocus transcripts contain an identical AG motif I and their AG motif II have the above described two aa difference at the 3' end.

SEP genes form a well-supported clade within the MADS-box gene phylogeny that can be further subdivided into three clades: a mixed eudicot and monocot clade containing the *Arabidopsis* *SEP3* gene matching all isolated crocus SEP sequences; a eudicot clade containing the *Arabidopsis* *SEP1*, *SEP2*, and *AGL* genes; and a clade comprised solely of monocot sequences[54]. All lineages have a conserved C-terminus internal motif in common, but diverse motifs at the 3' end. The alignment of CsatSEP3-deduced amino acid C-terminal sequences with other SEP-like proteins revealed a C-terminal motif (YMPGWLQ) typical for members of the SEP3 lineage[43] present in CsatSEP3A and CsatSEP3B, while a similar, but modified (YTPGWFP), motif was present in CsatSEP3C and CsatSEP3D sequences. Similar motifs are present in *Arabidopsis* *SEP3*, *Antirrhinum* *EFH72*, and *Lilium longiflorum* *LMADS3* and *LMADS4*.

The alignment of the two CsatAGL6-deduced amino acid sequences with other monocot AGL6-like proteins revealed a number of conserved L amino acids that play a role in protein-protein interactions as observed in the K domain[21]. There are also two conserved motifs at the C-terminus domain that are found in AGL6-like proteins: the FMLGWVL motif, typical for the AGL-6 subfamily[43] and a CEPTLQIGYH motif found in the AGL6 subfamily and in the ZMM7 lineage of the SEP subfamily of MADS-box proteins[43].

## PHYLOGENIC COMPARISONS OF CROCUS MADS-BOX SEQUENCES

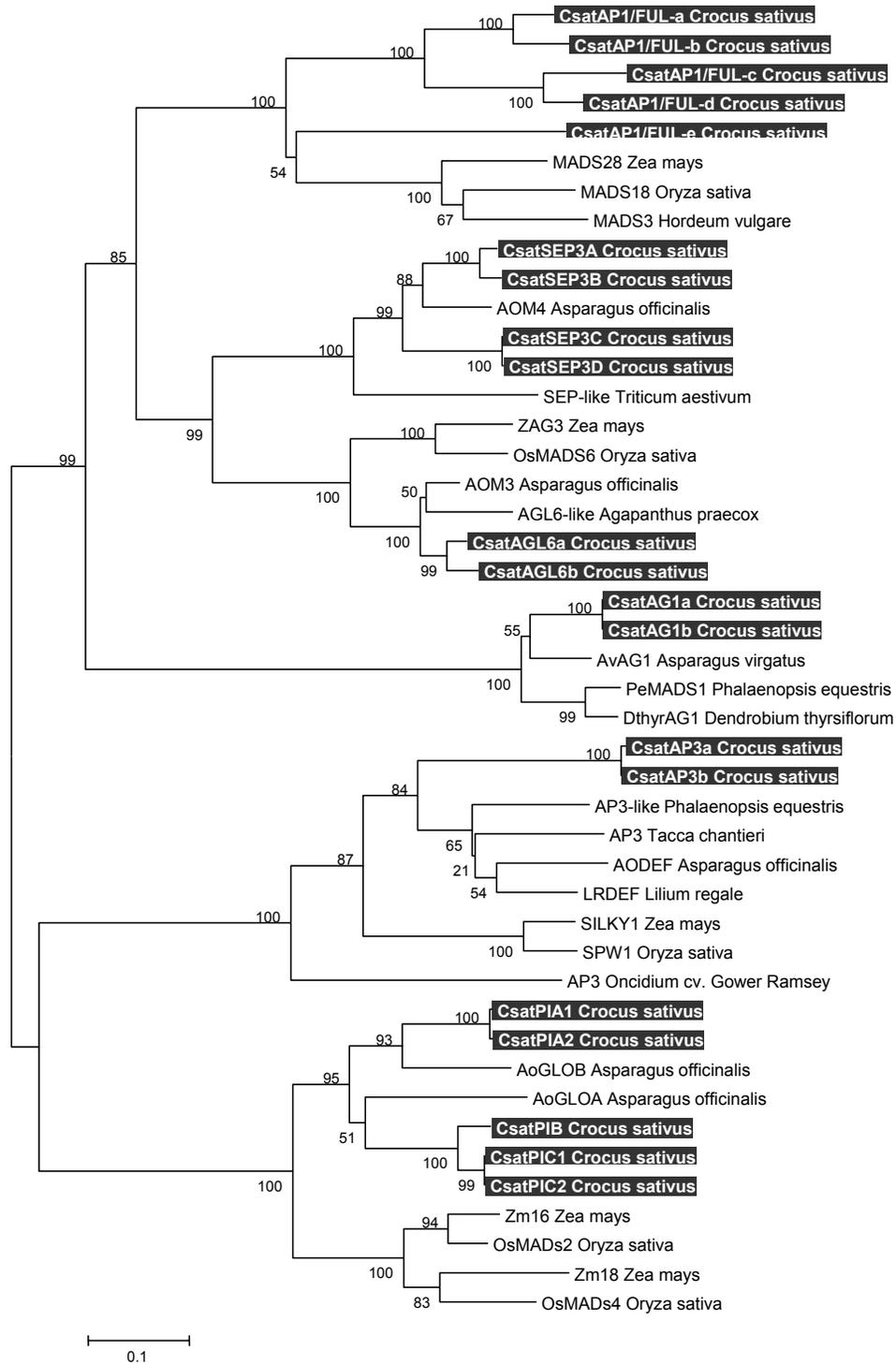
MADS-box genes in plants comprise a large family, counting 107 genes in *Arabidopsis*[55], 71 genes in rice[17], and probably numerous members in other plants. They are also present in nonflowering plants such as gymnosperms, ferns, and mosses, indicating that their role is not restricted to flower development. Phylogenetic relationships of plant MADS-box genes suggest that the different clades were probably established by gene duplication, diversification, and fixation. Genes within each clade acquired specific functions during the evolution of flowering plants[56].

The crocus sequences thus far isolated by our group sum up to 20 cDNAs, representing five major clades of the MADS phylogeny and the different A, B, C, and E functions of these genes. In Fig. 3, the phylogenetic tree showing the relationship of the crocus MADS-box sequences with related monocot members is presented. For constructing the tree, the Neighbor-Joining method was employed. The topology of the tree is in agreement with published topologies of angiosperm MADS-box proteins[14]. In this tree, the AP1/FUL, SEP, and AGL6 clades converge on a common node. This branch is always found in MADS phylogeny reconstructions[14] and is connected with the AG clade. The AP3 and PI clades also share a common node. Placement of the homologous genes in each clade is supported with high values in bootstrap resampling. The topology of the crocus MADS-box tree reflects the relationships between members of the major clades of MIKC-type genes present in angiosperms.

While core eudicots like *Arabidopsis*, *Antirrhinum*, *Petunia*, and others provided important information concerning the phylogenetic relationships of MADS proteins, comparisons with monocots like grasses and nongrass monocots like crocus represent a much longer evolutionary divergence of at least 130 mya[57]. The analysis with crocus indicated that crocus contains representatives of all ancestral gene lineages predating the core eudicot duplications[58], such as the *CsatAP1/FUL*, *CsatpaleoAP3*, *CsatAG*, and *CsatSEP3* genes. In comparison to grass monocots like rice and maize, our data indicate that crocus only contains the ancestral *SEP3*-like gene, as all four sequences obtained are highly homologous to *SEP3*, while grasses like rice and maize apparently underwent separate SEP lineage proliferation since they contain *SEP2*, *3*, and *4* homologues[54].

## FUTURE PERSPECTIVES

Further experiments are underway for isolation of the complete set of crocus MADS-box genes, and elucidation of their role in crocus flower development and tepal formation, but also possible roles in other aspects of crocus physiology like transition to flowering, corm formation, etc. For this goal, the famRCA-



**FIGURE 3** Phylogenetic relationships of MADS-box proteins of crocus and related monocots. Amino acid sequences were aligned with *Clustal W*[42]. The tree was generated by the Neighbor-Joining method using the p-distance correction. Numbers next to the nodes are bootstrap values from 1000 replications. The scale indicates amino acid substitutions.

RACE method described in this review, which allows the simultaneous isolation of members of a family of homologous genes, may prove to be a very useful and highly efficient tool. Furthermore, the availability of a complete set of MADS-box sequences from the cultivated *C. sativus* triploid species, forms the base for similar analysis of this and other important families of proteins studied[38] in the putative progenitor diploid species of cultivated crocus. Comparative structural and phylogenetic analysis of these proteins will be helpful to solve the origin of the cultivated triploid *C. sativus*. Finally, it will be possible to characterize the numerous field isolated flower mutants of *C. sativus* for further elucidating the role of the different groups of the A-, B-, C-, and E-type MADS proteins in crocus flower organ formation.

## ACKNOWLEDGMENTS

The authors are grateful to Mr. Yiannis Patsios for his invaluable help in collecting plant material in the field, and to all of our lab members for helpful comments and discussion. Work at the Institute of Agrobiotechnology is supported by the General Secretariat for the Research and Technology (Greek Ministry of Development).

## REFERENCES

1. Komaki, M., Okada, K., Nishino, E., and Shimura, Y. (1988) Isolation and characterization of novel mutants of *Arabidopsis thaliana* defective in flower development. *Development* **104**, 195–203.
2. Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989) Genes directing flower development in *Arabidopsis*. *Plant Cell* **1**, 37–52.
3. Coen, E.S. and Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31–37.
4. Carpenter, R. and Coen, E.S. (1990) Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus*. *Genes Dev.* **4**, 1483–1493.
5. Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991) Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**, 1–20.
6. Drews, G.N., Bowman, J.L., and Meyerowitz, E.M. (1991) Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* **65**, 991–1002.
7. Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1994) Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Cell* **76**, 131–143.
8. Mizukami, Y. and Ma, H. (1992) Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* **71**, 119–131.
9. Krizek, B.A. and Meyerowitz, E.M. (1996) The *Arabidopsis* homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. *Development* **122**, 11–22.
10. Theissen, G. (2001) Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* **4**, 75–85.
11. Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H.J., Angenent, G.C., and van Tunen, A.J. (1995) The *petunia* MADS box gene *FBP11* determines ovule identity. *Plant Cell* **7**, 1859–1868.
12. Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**, 200–203.
13. Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M.F. (2004) The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* **14**, 1935–1940.
14. Becker, A. and Theissen, G. (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* **29**, 464–489.
15. Messenguy, F. and Dubois, E. (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* **316**, 1–21.
16. Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Munster, T., Winter, K.U., and Saedler, H. (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115–149.
17. De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouze, P., Theissen, G., and Van de Peer, Y. (2003) Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. *J. Mol. Evol.* **56**, 573–586.
18. Szczygłowski, K. and Amyot, L. (2003) Symbiosis, inventiveness by recruitment? *Plant Physiol.* **131**, 935–940.
19. van Tunen, A.J., Eikelboom, W., and Angenent, G.C. (1993) Floral organogenesis in *Tulipa*. *Flow. Newsl.* **16**, 33–37.
20. Grilli Caiola, M., Caputo, P., and Zanier, R. (2004) RAPD analysis in *Crocus sativus* L. accessions and related crocus

- species. *Biol. Plant.* **48**, 375–380.
21. Tsaftaris, A.S., Polidoros, A.N., Pasentsis, K., Chen, M.-K., and Yang, C.-H. (2007) Characterization of *crocus* (*Crocus sativus* L.) flower specific AGL6-like MADS-box genes and the effects of ectopic expression in *Arabidopsis*. Submitted.
  22. Tsaftaris, A.S., Pasentsis, K., Iliopoulos, I., and Polidoros, A.N. (2004) Isolation of three homologous AP1-like MADS-box genes in *crocus* (*Crocus sativus* L.) and characterization of their expression. *Plant Sci.* **166**, 1235–1243.
  23. Wellmer, F., Alves-Ferreira, M., Dubois, A., Riechmann, J.L., and Meyerowitz, E.M. (2006) Genome-wide analysis of gene expression during early *Arabidopsis* flower development. *PLoS Genet.* **2**, art. no.-e117.
  24. Fernandez, J.A. (2004) Biology, biotechnology and biomedicine of saffron. *Recent Res. Dev. Plant Sci.* **2**, 127–159.
  25. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8998–9002.
  26. Loh, E.Y., Elliott, J.F., Cwirla, S., Lanier, L.L., and Davis, M.M. (1989) Polymerase chain reaction with single-sided specificity: analysis of T cell receptor delta chain. *Science* **243**, 217–220.
  27. Tsaftaris, A.S., Polidoros, A.N., Pasentsis, K., and Kalivas, A. (2006) Tepal formation and expression pattern of B-class paleoAP3-like MADS-box genes in *crocus* (*Crocus sativus* L.). *Plant Sci.* **170**, 238–246.
  28. Tsaftaris, A.S., Pasentsis, K., and Polidoros, A.N. (2005) Isolation of a differentially spliced C-type flower specific AG-like MADS-box gene from *Crocus* (*Crocus sativus*) and characterization of its expression. *Biol. Plant.* **49**, 499–504.
  29. Kalivas, A., Pasentsis, K., Polidoros, A.N., and Tsaftaris, A.S. (2007) Heterotopic expression of B-class floral homeotic genes PISTILLATA/GLOBOSA supports a modified model for *crocus* (*Crocus sativus* L.) flower formation. *DNA Seq.* **18**, 120–130.
  30. Schaefer, B.C. (1995) Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.* **227**, 255–273.
  31. Huang, J.-C. and Chen, F. (2006) Simultaneous amplification of 5' and 3' cDNA ends based on template-switching effect and inverse PCR. *BioTechniques* **40**, 187–189.
  32. Fu, G.K., Wang, J.T., Yang, J., Au-Young, J., and Stuve, L.L. (2004) Circular rapid amplification of cDNA ends for high-throughput extension cloning of partial genes. *Genomics* **84**, 205–210.
  33. Eyal, Y., Neumann, H., Or, E., and Frydman, A. (1999) Inverse single-strand RACE: an adapter-independent method of 5' RACE. *BioTechniques* **27**, 656–658.
  34. Shi, X., Karkut, T., Chahmanhkah, M., Altling-Mees, M., Hemmingsen, S.M., and Hegedus, D. (2002) 5'-RACEing across a bridging oligonucleotide. *BioTechniques* **32**, 480–482.
  35. Polidoros, A.N., Pasentsis, K., and Tsaftaris, A.S. (2006) Rolling circle amplification-RACE: a method for simultaneous isolation of 5' and 3' cDNA ends from amplified cDNA templates. *BioTechniques* **41**, 35–40.
  36. Ganguly, S.N., Sanyal, T., Sircar, P.K., and Sircar, S.M. (1970) A new gibberellin (A25) in the leaves of *Sonneratia apetala* ham. *Chem. Ind.* **25**, 832–833.
  37. Tani, E., Polidoros, A.N., and Tsaftaris, A.S. (2007) Characterization and expression analysis of FRUITFULL- and SHATTERPROOF-like genes from peach (*Prunus persica*) and their role in split-pit formation. *Tree Physiol.* **27**, 649–659.
  38. Kalivas, A., Pasentsis, K., and Tsaftaris, A.S. (2007) Simultaneous Isolation, characterization and expression analysis of full-length cDNA sequence members of the NAC-like family of transcription factor genes from *crocus* (*Crocus sativus* L.) flowers using a modified RCA-RACE method. Submitted.
  39. Tsaftaris, A.S., Polidoros, A.N., and Pasentsis, K. (2007) FamRCA-RACE Cloning, characterization and expression patterns of flower specific SEPALLATA-like MADS-box genes in *crocus* (*Crocus sativus* L.). Submitted.
  40. Riechmann, J.L., Wang, M., and Meyerowitz, E.M. (1996) DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res.* **24**, 3134–3141.
  41. Lamb, R.S. and Irish, V.F. (2003) Functional divergence within the APETALA3/PISTILLATA floral homeotic gene lineages. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6558–6563.
  42. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
  43. Vandebussche, M., Theissen, G., Van de Peer, Y., and Gerats, T. (2003) Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Res.* **31**, 4401–4409.
  44. Kramer, E.M., Jaramillo, M.A., and Di Stilio, V.S. (2004) Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* **166**, 1011–1023.
  45. Litt, A. and Irish, V.F. (2003) Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* **165**, 821–833.
  46. Yu, H. and Goh, C.J. (2000) Identification and characterization of three orchid MADS-box genes of the AP1/AGL9 subfamily during floral transition. *Plant Physiol.* **123**, 1325–1336.
  47. Kramer, E.M., Dorit, R.L., and Irish, V.F. (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages.

- Genetics* **149**, 765–783.
48. Winter, K.U., Saedler, H., and Theissen, G. (2002) On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in Arabidopsis by expression of an orthologue from the gymnosperm Gnetum. *Plant J.* **31**, 457–475.
  49. Park, J.H., Ishikawa, Y., Yoshida, R., Kanno, A., and Kameya, T. (2003) Expression of AODEF, a B-functional MADS-box gene, in stamens and inner tepals of the dioecious species *Asparagus officinalis* L. *Plant Mol. Biol.* **51**, 867–875.
  50. Kanno, A., Saeki, H., Kameya, T., Saedler, H., and Theissen, G. (2003) Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Mol. Biol.* **52**, 831–841.
  51. Friedman, W.E. and Williams, J.H. (2004) Developmental evolution of the sexual process in ancient flowering plant lineages. *Plant Cell* **16 Suppl**, S119–132.
  52. Tzeng, T.Y. and Yang, C.H. (2001) A MADS box gene from lily (*Lilium longiflorum*) is sufficient to generate dominant negative mutation by interacting with PISTILLATA (PI) in *Arabidopsis thaliana*. *Plant Cell Physiol.* **42**, 1156–1168.
  53. Riechmann, J.L. and Meyerowitz, E.M. (1997) MADS domain proteins in plant development. *Biol. Chem.* **378**, 1079–1101.
  54. Malcomber, S.T. and Kellogg, E.A. (2004) Heterogeneous expression patterns and separate roles of the SEPALLATA gene LEAFY HULL STERILE1 in grasses. *Plant Cell* **16**, 1692–1706.
  55. Parenicova, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C., and Colombo, L. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell* **15**, 1538–1551.
  56. Zahn, L.M., Kong, H., Leebens-Mack, J.H., Kim, S., Soltis, P.S., Landherr, L.L., Soltis, D.E., Depamphilis, C.W., and Ma, H. (2005) The evolution of the SEPALLATA subfamily of MADS-box genes: a pre-angiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169**, 2209–2223.
  57. Chaw, S.M., Chang, C.C., Chen, H.L., and Li, W.H. (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol. Evol.* **58**, 424–441.
  58. Kramer, E.M. (2006) Floral patterning and control of floral organ formation. In *Flowering and Its Manipulation*. Annual Plant Reviews Vol. 20. Ainsworth, C., Ed. Blackwell Publishing, Oxford. pp. 49–70.

---

**This article should be cited as follows:**

Tsaftaris, A.S., Polidoros, A.N., Pasentsis, K., and Kalivas, A. (2007) Cloning, structural characterization, and phylogenetic analysis of flower MADS-box genes from crocus (*Crocus sativus* L.). *TSW Development & Embryology* **2**, 73–87. DOI 10.1100/tswde.2007.123.

---



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

