Research Article

# Unravelling MADS-box gene family in *Eucalyptus* spp.: A starting point to an understanding of their developmental role in trees

Beatriz Fonseca de Oliveira Dias[1], Jean Luiz Simões-Araújo[2], Claudia A.M. Russo[3],
Rogério Margis[1,4] and Márcio Alves-Ferreira[1]

[1]*Universidade Federal do Rio de Janeiro, Instituto de Biologia, Departamento de Genética,*
*Laboratório de Genética Molecular Vegetal, Rio de Janeiro, RJ, Brazil.*
[2]*Embrapa Agrobiologia, Laboratório de Genética e Bioquímica, Seropédica, RJ, Brazil.*
[3]*Universidade Federal do Rio de Janeiro, Instituto de Biologia, Departamento de Genética,*
*Laboratório de Biodiversidade Molecular, Rio de Janeiro, RJ, Brazil.*
[4]*Universidade Federal do Rio Grande do Sul, Departamento de Bioquímica, Porto Alegre, RS, Brazil.*

## Abstract

MADS-box genes encode a family of transcription factors which control diverse developmental processes in flowering plants ranging from root to flower and fruit development. Members of the MADS-box gene family share a highly conserved sequence of approximately 180 nucleotides that encodes a DNA-binding domain. We used bioinformatics tools to investigate the information generated by the *Eucalyptus* Expressed Sequence Tag (FORESTs) genome project in order to identify and annotate MADS-box genes. The comparative phylogenetic analysis of the *Eucalyptus* MADS-box genes with *Arabidopsis* homologues allowed us to group them into one of the well-known subfamilies. Trends in gene expression of these putative *Eucalyptus* MADS-box genes were investigated by hierarchical clustering analysis. Among 24 MADS-box genes identified by our analysis, 12 are expressed in vegetative organs. Out of these, five are expressed predominately in wood. Understanding of the molecular mechanisms performed by MADS-box proteins underlying *Eucalyptus* growth, development and stress reactions would provide important insights into tree development and could reveal means by which tree characteristics could be modified for the improvement of industrial properties.

*Key words:* MADS-box, vegetative development, reproductive development, phylogenetic analysis, ESTs.

Received: May 31, 2004; Accepted: May 11, 2005.

## Introduction

Members of the MADS-box gene family share a highly conserved DNA-binding domain of approximately 180 nucleotides. This domain allows the protein to behave as a transcriptional regulator and is called the MADS-box domain. In flowering plants, these proteins play important roles in a wide range of important biological functions, including the control of flowering time, the determination of floral meristem identity, the establishment of floral organ identities, fruit development, seed pigmentation and endothelium development and the control of vegetative development (Michaels *et al*., 2003; Gu *et al*., 1998; Pelaz *et al*., 2000; Liljegren *et al*., 2000; Nesi *et al*., 2002; Zhang and Forde, 2000).

Send correspondence to Márcio Alves-Ferreira. Universidade Federal do Rio de Janeiro, Instituto de Biologia, Departamento de Genética, Laboratório de Genética Molecular Vegetal, Av. do Pau Brasil 211, Sala A2 76, Ilha do Fundão, 21944-970 Rio de Janeiro, RJ, Brazil. E-mail: alvesfer@biologia.ufrj.br.

Phylogenetic analyses of regulatory multigene families provided hard evidence for an ancient gene duplication event of a MADS-box precursor that took place before the divergence of plants, animals and fungi (Alvarez-Buylla *et al*., 2000a). This duplication gave rise to two main MADS-box lineages, referred to as type I and type II (Alvarez-Buylla *et al*., 2000a) that exhibit different functional domains.

Plant type II MADS-box genes, for instance, share a characteristic stereotypic organization in their functional domains. Typical type II proteins, the so-called MIKC-type, consist of two variable Carboxyl-terminal (C-) and Intervening (I-) and two well-conserved MADS-box (M-) and Keratin-like (K-) domains. The MADS-box domain (MEF2-like - MYOCYTE ENHANCER FACTOR2-like) is located at the N-terminal of the protein and it determines DNA-binding, dimerization and accessory factor binding functions (Shore and Sharrocks, 1995). The K-domain probably forms a coiled-coil structure and it is critical in

protein-protein interactions (Davies *et al*., 1996; Fan *et al*., 1997). On the other hand, I-domain constitutes a determinant for the formation of DNA-binding dimers (Riechmann and Meyerowitz, 1997). Finally, the C-terminal region is poorly conserved and may function as a trans-activation domain (Riechmann and Meyerowitz, 1997).

In contrast, type I MADS-box genes are characterized by an SRF-like MADS domain (SERUM RESPONSE FACTOR-like) and the lack of a well-defined K domain. Type I proteins classification separates the group into two main distinct classes, class M and N. This classification is based on the presence of certain conserved class specific motifs that allow alignment of longer regions in the type I than in type II MADS-box proteins. A third class (class O) does not feature the conserved motifs in the C-terminal region (De Bodt *et al*., 2003). Up until now only the single type I MADS-box gene, *PHERES1* of *Arabidopsis thaliana*, has been fully characterized. This gene is transiently expressed during embryo and endosperm development and it is currently associated with seed abortion in a specific mutant background (Köhler *et al*., 2003). Even though a rough description of the evolution of this complex family has been available, it was not until the completion of the *A. thaliana* genome sequence that a better defined picture has emerged (*A. thaliana* Genome Initiative, 2000).

For most of the well-characterized gene clades, considered in previous studies, strong correlations between membership in gene subfamilies and expression patterns and functions have been found (Theißen *et al*., 2000), such as the *GGM13*-like genes (mainly expressed in ovules), *STMADS11*- and *TM3*-like genes (mainly expressed in vegetative organs; repressors or promoters of flowering, or timers of vegetative developmental processes). In some cases, however, intriguing differences in the expression patterns of genes within one subfamily are found. AGL12-like genes, for example, may be expressed in roots, or in leaves and in inflorescences (Becker and Theißen, 2003).

Particularly well known is the importance of the MADS-box gene family in reproductive development. For example, loss-of-function of some flowering plant MADS-box genes causes homeotic transformations of floral organs, indicating that these genes work as organ identity genes (homeotic selector genes) during the ontogeny of flowers. Besides providing floral homeotic functions, MADS-box genes have many other roles within the gene networks that govern reproductive development in eudicotyledonous flowering plants. For example, some MADS-box genes are flowering time genes which, depending on internal or environmental factors such as plant age, day-length, and cold, repress or promote the floral transition (Hartmann *et al*., 2000). MADS-box genes are also involved in developmental processes that follow fertilization of the flower, *i.e.*, seed and fruit development (Gu *et al*., 1998; Liljegren *et al*., 2000). Moreover, transcription of a number of MADS-box genes outside flowers and fruits as well as an increasing number of mutant and transgenic flowering plants suggests that members of this gene family play regulatory roles also during vegetative development, such as embryo, root or leaf development (*e.g.*, Alvarez-Buylla *et al*., 2000a; Huang *et al*., 1995; Ma *et al*., 1991; Rounsley *et al*., 1995; Theißen *et al*., 2000).

Therefore, the aim of the present report was the identification of MADS-box related *Eucalyptus* expressed sequence tags (ESTs), retrieved from FORESTs data set. Moreover, by clustering genes according to their relative abundance in the various EST libraries, expression patterns of genes across various tissues were generated and genes with similar patterns could be grouped and interpreted. The combination of phylogenetic analysis and expression patterns for some of the *Eucalyptus* genes is bound to reveal various interesting aspects about some of the potential MADS-box genes and lead to a more solid understanding of *Eucalyptus* biology and the associated biotechnological applications.

## Material and Methods

The primary data source for this work was clustered gene sequences of the FORESTs project database. These sequences were assembled from ESTs obtained from the sequencing of several *Eucalyptus* spp. cDNA libraries, corresponding to different tissues and various physiological states. Complete information on cDNA libraries, sequencing, clustering and other features of FORESTs project may be found at https://forests.esalq.usp.br/. Additionally, *A. thaliana* MADS-box sequences obtained from The National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) and from The Institute for Genomic Resources (TIGR, http://www.tigr.org/tdb/e2k1/ath1/) were used for comparison.

In order to search for MADS-box sequences in the FORESTs databases, a MADS-box consensus sequence was used. This consensus was generated by the COBBLER program (COnsensus Biasing By Locally Embedding Residues, http://blocks.fhcrc.org/blocks/cobbler.html) from all identified MADS-box amino acid sequences "MGRKKIEI KRIENKTNRQVTFSKRRNGLFKKAHELSVLCDAEV ALIVFSPSGRlyeyannni". Searches were conducted using the tBLASTN module that compares the consensus amino acid sequence with a translated nucleotide sequences database (Altschul *et al.*, 1997). All sequences that exhibit a significant alignment (E-value of $10^{-5}$) with the consensus were retrieved from the FORESTs database.

All retrieved sequences were then re-inspected for occurrence of MADS conserved motif using the InterProScan (http://www.ebi.ac.uk/InterProScan/) and PRODOM (http://prodes.toulouse.inra.fr/prodom/current/html/form.php) programs. With this procedure, 23 EST-contigs were positively identified as putative MADS-box sequences.

Each identified EST-contig was then used for a second BLASTN search in the FORESTs database to search for other putative MADS-box sequences. All EST-contigs found in the database that presented an E-value of $10^{-5}$ or lower were selected and inspected for the presence of the conserved MADS-box motif. This procedure was performed using the InterProScan and PRODOM programs. At this point, six additional EST-contigs were identified (EGJMFB1092E09.g, EGCCFB1224D05.g, EGUTFB 1098B01.g, EGBMLV2270D12.g, EGBFFB1042G10.g and EGQHRT6253G06.g). Nevertheless, a single new EST-contig (EGJMFB1092E09.G) was added to the previous group of 23 EST-contigs, increasing the total to 24 EST-contigs included in the following analyses. The other five EST-contigs coded for proteins with incomplete MADS-box were excluded from the analysis. To conclude the identification analysis, a third round of BLAST (tBLASTN) search was performed using consensus amino acid sequences from MADS-box present in *A. thaliana* MADS-box proteins; however, no additional EST-contigs were identified.

Sequences alignment was performed at the CLUSTALW website (www.ebi.ac.uk/index/clustalw.html) using default parameters. Amino acid sequences were used for the phylogenetic analysis due to the high variability of the sequences and the unreliability of the nucleotide alignment. Final alignment was visually inspected and manually corrected. Also, wherever homology could not be ascertained, segments were removed from the subsequent phylogenetic analyses. The Molecular Evolutionary Genetics Analysis (MEGA) software, version 2.0 (Kumar *et al.*, 2000) was used for the phylogenetic analysis. Average p-distances were high so the Poisson model was used to provide unbiased estimates of the number of substitutions between sequences. Phylogenetic trees were obtained from the Neighbor-joining method with Poisson distances (Saitou & Nei, 1987), and the pair-wise deletion option. The confidence probability test was performed to evaluate interior branch stability. This test is known to be a reliable estimator of the confidence of clusters in a distance-based tree (Sitnikova *et al.*, 1995).

For each EST-contig, the frequency of reads in the selected libraries was calculated. This procedure requires a normalization that is accomplished by dividing the total number of reads in the specific library by the total number of reads in all libraries and then dividing the number of reads of each EST-contig by the ratio found for each library. The results were cast in a matrix relating contigs and libraries, forming the so-called 'digital northern blot.' The EST-contigs and libraries were grouped by hierarchical clustering, using the Cluster and Tree View programs (Eisen *et al.*, 1998). Aggregation of both putative MADS-box genes and expression libraries was based on Spearman Rank correlation matrix, with previously formed clusters being substituted by their average pattern. The data matrix

was reordered, according to similarities in the pattern of gene expression and displayed as black and white arrays of EST-contigs, using a gray scale representing the number of reads from a specific library in each EST-contig.

## Results and Discussion

After the elimination of incomplete or artificial sequences, a final data set of 24 *Eucalyptus* EST-contigs sharing significant sequence similarity to MADS-box domain were selected for further analysis. We found only one representative of type I MADS-box gene family in *Eucaliptus* spp. The phylogenetic relationship between members of MIKC-type MADS-box gene type II present in the FORESTs databank and *A. thaliana* genes from public databases is shown in the Figure 1. Among the type II MADS-box genes, the representatives were distributed mainly among three subfamilies: *TM3*-like (five EST-contigs), *AGL2*-like (four EST-contigs) and *DEF+GLOB*-like (four EST-contigs) (see Figure 1). One *Eucalyptus* EST-contig (EGU TST6222D07.g) was not included in any of the well-defined groups (Figure 1). Analysis of gene expression *in silico* of the 24 putative MADS-box *Eucalyptus* EST-contigs was performed based on the frequency of sequence tags in cDNA libraries (Figure 2; Ewing *et al.*, 1999). We were able to identify two distinct clusters, cluster I and II, representing clusters of EST-contigs preferentially expressed in reproductive and vegetative organs, respectively.

Apart from the vast number of described reproductive development functions, MADS-box genes might play many other roles within the gene networks that govern vegetative development. In fact, the transcription of a number of MADS-box genes outside flowers and fruits strongly suggests that MADS-box members are critical also in regulatory roles during vegetative development, such as embryo, root or leaf development (*e.g.*, Alvarez-Buylla *et al.*, 2000a; Huang *et al.*, 1995; Ma *et al.*, 1991; Rounsley *et al.*, 1995; Theiβen *et al.*, 2000).

### MADS-box of *Eucalyptus* with vegetative-specific expression

### Vegetative-specific expression type I MADS-box genes

The single EST-contig of MADS-box type I was found to be expressed exclusively at the ST2 library (plantlets under drought stress). Unfortunately, the type I MADS-box genes constitute a largely unexplored subfamily and their function remains completely unknown. Even though, in *A. thaliana*, 47 MADS-box type I genes were identified in the genome (De Bodt *et al.*, 2003; Alvarez-Buylla *et al.*, 2000b), other plant species that have been surveyed show a similar number of type I genes (De Bodt *et al.*, 2003).
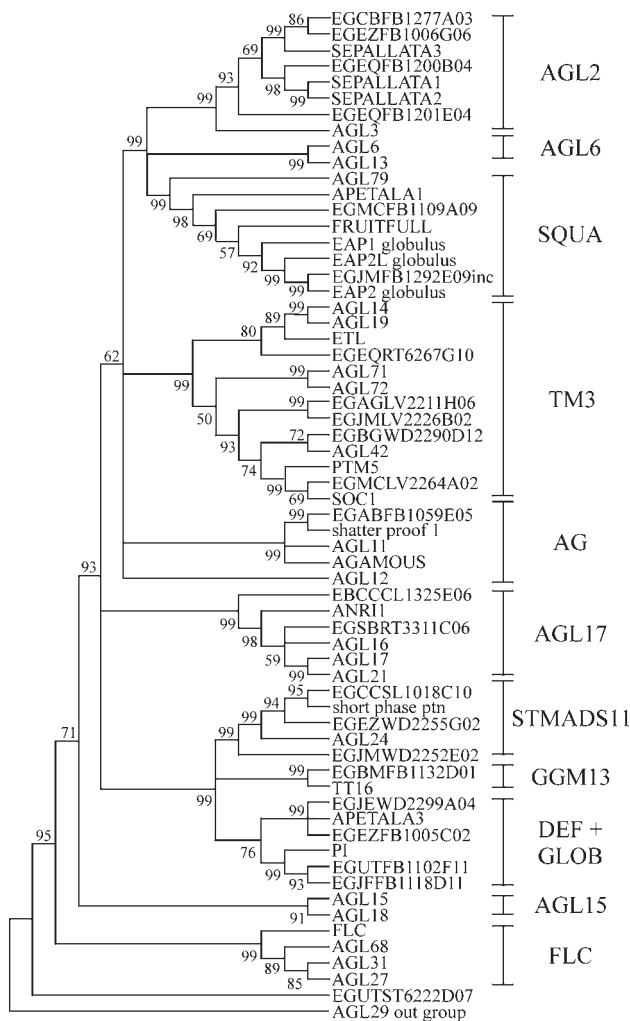
**Figure 1** - Phylogenetic relationship between members of MIKC-type MADS-box gene present in the FORESTs databank and *A. thaliana* genes from public databases. Clustering was performed using the deduced amino acid sequence from the *Eucalyptus* EST-contigs consensus nucleotide sequences and amino acid sequences in public databases. *Eucalyptus* EST-contigs encoding for putative MADS-box proteins were selected based on the procedures described in material and methods. Globulus stands for a *E. globulus* sequences. (O22456) SEP3/At1g24260, (NP_18 6880) SEP2/At3g02310, (NP_568322) SEP1/At5g15800, (NP_973411) AGL3/At2g03710, (NP_182089) AGL6/At2g45650, (NP_191671) AG L13/At3g61120, (NP_189645) AGL79/At3g30270, (Q8RYD9) TT16/ At5g23260, (NP_177074) AP1/At1g69120, (Q38876) FUL/At5g60910, (NP_849351) AGL11/At4g09960, (NP_565022) AGL12/At1g71692, (P17839) AGAMOUS/At4g18960, (NP_191437) SHP1/At3g58780, (N P_191002) AP3/At3g54340, (NP_197524) PI/At5g20240, (NP_196883) AGL15/At5g13760, (NP_974450) AGL18/At3g57390, (AAN52783.1) AGL68/At5g65060, (AAN52775.1) AGL31/At5g65050, (Q9S7Q7) FLC/At5g10140, (NP_974161.1) AGL27/At1g77080, (CAB79364.1) AGL24/At4g24540, (AAG24508.1) SVP/At2g22540, (CAB09793.1) ANR1/At2g14210, (NP_191282.2) AGL16/At2g57230, (AAL73213.1) AGL21/At4g37940, (NP_179848.1) AGL17/At2g22630, (NP_192925) AGL14/At4g11880, (NP_194026) AGL19/At4g22950, (AAN52777.1) AGL42/At5g62615, (AAN52784.1) AGL71/At5g51870, (AAN52785.1) AGL72/At5g51860, (O64645) SOC1/At2g42660 from *A. thaliana,* (AF 377868.1) PTM5 from *P. tremulodies*, (AAG30923) EAP2L, and (AAG 27459) EAP2S, and (AAG24909) EAP1, and (AAD16052.1) ETL from *E. globulus.*

The low number of ESTs may be attributed to a particularly low level of expression or, alternatively, that the type I genes are only expressed under conditions not yet monitored in EST sequencing projects. EGEZST2245 G11.g belongs to class M, subclass M3 genes. This EST-contig sequence possesses a single motif "YSFGHPSV DAV." A comparative analysis showed a strikingly high sequence similarity of this EST-contig with *A. thaliana* AGL29 (data not shown).

## Vegetative-specific expression type II MADS-box genes

### TM3 subfamily

Among the 11 EST-contigs expressed in non-reproductive organs, five belong to the TM3 MADS-box subfamily. Members of the TM3 MADS-box subfamily in angiosperms are preferentially expressed in the vegetative parts of the plant, as are their gymnosperm counterparts (Tandre *et al.*, 1995; Walden *et al.*, 1998; Winter *et al.*, 1999). Nonetheless, instances of *TM3*-like gene expressed in reproductive organs have been reported (such as *ZMM5* from maize; Münster *et al.*, 2002; *ZmMADS1* from maize by Heuer *et al.*, 2001).

In terms of mutants, a single mutant has been described for the TM3 subfamily up to this point (Becker and Theiβein, 2003). The mutant is deficient for the gene *SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1)*. This gene is supposed to be an early target of the flowering time gene *CONSTANS (CO)*, as *CO* promotes flowering in part through the activation of *SOC1* (Yu *et al.*, 2002). Other members of subfamily TM3, however, have been also associated with vascular tissue formation (Alvarez-Buylla *et al.*, 2000a; Decroocq *et al.*, 1999). More recently, Cseke and collaborators beautifully characterized a *TM3*-like gene involved with the vascular cambium or expanding xylem during wood formation (Cseke *et al.*, 2003). This association may motivate a better characterization of *Eucalyptus TM3*-like genes that are potentially related with wood formation since wood is produced from vascular cambium. Sadly, little is known about how differentiation of the cambial derivatives is controlled at the molecular level.

In order to have a more detailed account of the evolutionary events that took place in the history of this important family, we also analyzed them separately (Figure 3). In the *A. thaliana* genome, there is a high number of different *TM3*-like genes, some of which have not even been fully annotated yet (Figure 1) (Becker and Theiβen, 2003). Of these, only four, *SOC1* (also known as *AGL20*), *AGL14*, *AGL19* and *AGL42* have been named so far (Becker and Theiβen, 2003). Of these, *AGL14* and *AGL19* are exclusively expressed in roots (Rounsley *et al.*, 1995; Alvarez-Buylla *et al.*, 2000a); *AGL19* is especially expressed in the columella, lateral root cap, and epidermal
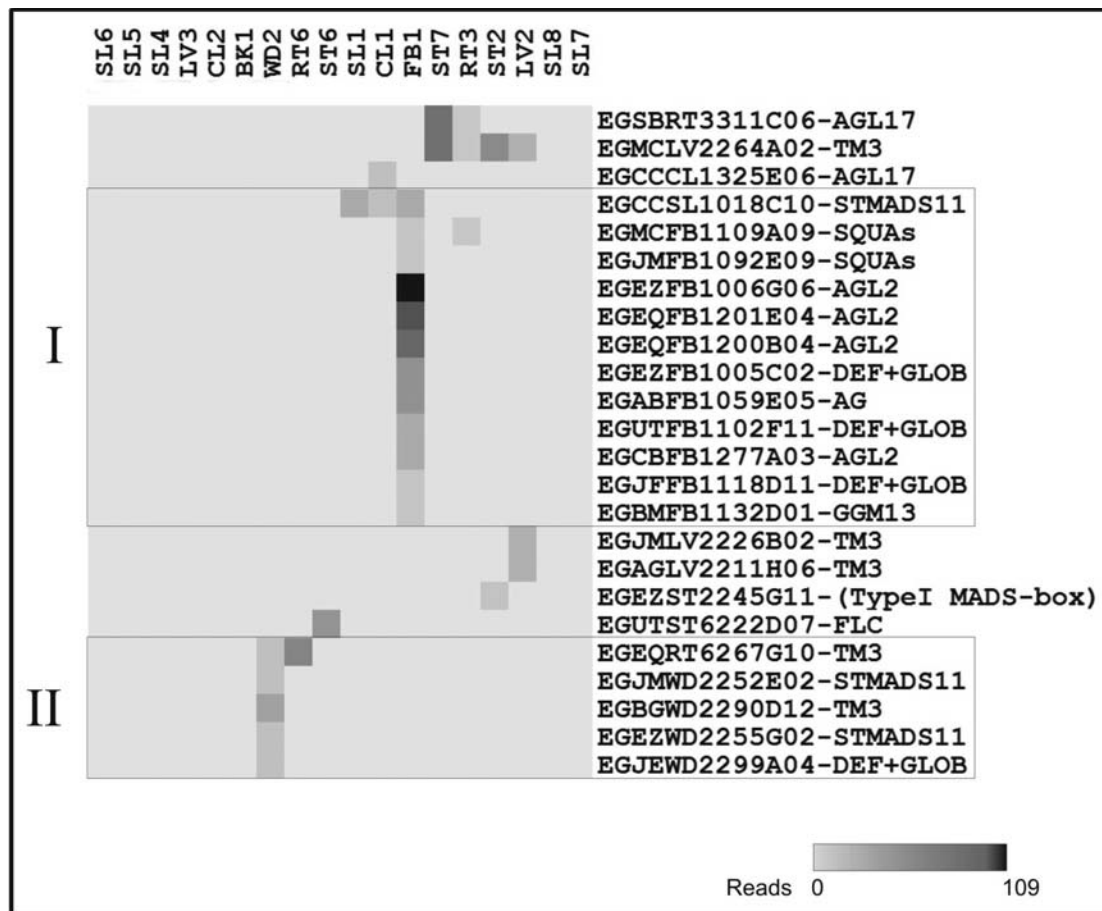
**Figure 2** - Expression profiles of the 24 putative MADS-box *Eucalyptus* EST contigs in selected cDNA libraries from the FORESTs database. I and II represent clusters of EST-contigs preferentially expressed in flower buds/flower and fruits (FB1) and wood (WD2), respectively. Data represent the relative number of reads from a specific library in each EST-contigs after normalization. Each EST-contig is represented by a single row, and each library is represented by a simple column. *O*uter bark, inner bark, cortex and the pith of eight-year-old *E. grandis* (BK1); *E. grandis* dark formed calli (CL1); *E. grandis* light formed calli (CL2); Flower buds, flowers and fruits (FB1); leaves from developing plants (LV1); Phosphate and Boron deficient leaves and with susceptibility to canker and rust (LV2); leaves damaged with *Thyrinteina* for seven days (LV3); *E. grandis* dark growing seedlings with three hours of light exposition (SL1); *E. globulus* dark growing seedlings (SL4); *E. saligna* dark growing seedlings (SL5); *E. urophylla* dark growing seedlings (SL6); *E. grandis* dark growing seedlings (SL7); *E. camaldulensis* dark growing seedlings (SL8); stem from drought stress susceptible seedlings prepared with 0.6 to 2.0 kb DNA fragments (ST2); stem from drought stress susceptible seedlings prepared with 0.8 to 3.0 kb DNA fragments (ST6); stem from frost resistant and susceptible plants (ST7); roots from developing plants (RT3); roots from frost resistant and susceptible trees (RT6) and *E. grandis* wood (WD2).

cells of the meristematic region and in the central cylinder of mature roots (Alvarez-Buylla *et al*., 2000a); and *SOC1* (Onouchi *et al*., 2000; Samach *et al*., 2000), much like its orthologous gene from *Sinapis alba* (*SaMADSA*, Menzel *et al*., 1996) is abundantly expressed in the apical meristem and responds to long photoperiods.

In our analysis, we have carefully examined the phylogeny of the TM3 subfamily from different plant species with known spatial expression patterns (Figure 3). Two *Eucalyptus* EST-contigs (EGEQRT6267G10.g and EGBGW D2290D12.g) are expressed in the WD2 (wood of *E. grandis*) library. EGEQRT6267G10.g is expressed in WD2, but also in the RT6 (roots from frost resistant and susceptible trees) library for *Eucalyptus*. The phylogenetic analysis showed it to be part of a group formed by AGL14, AGL19 of *A*. *thaliana* and ETL of *E. globulus,* all of which

are expressed in vegetative organs (Alvarez-Buylla *et al*., 2000a; Decroocq *et al*., 1999).

The phylogenetic analysis also demonstrated a close relationship between EGEQRT6267G10.g and ETL of *E. globulus*. It is important to bear in mind that, unlike *AGL14* and *AGL19*, the *ETL* gene is expressed in both vegetative and reproductive organs, predominantly in root and shoots meristems and organ primordia (Decroocq *et al*., 1999). In fact, *ETL* transcripts are found in the vasculature of young roots, yet no transcripts are detectable in wood formation. This characterizes fundamental differences in terms of functions of EGEQRT6267G10.g and ETL. Although the comparison between *A. thaliana* and *Eucalyptus* orthologues conveys a broader spatial expression domain that may even include wood forming tissues in trees, an *in situ*
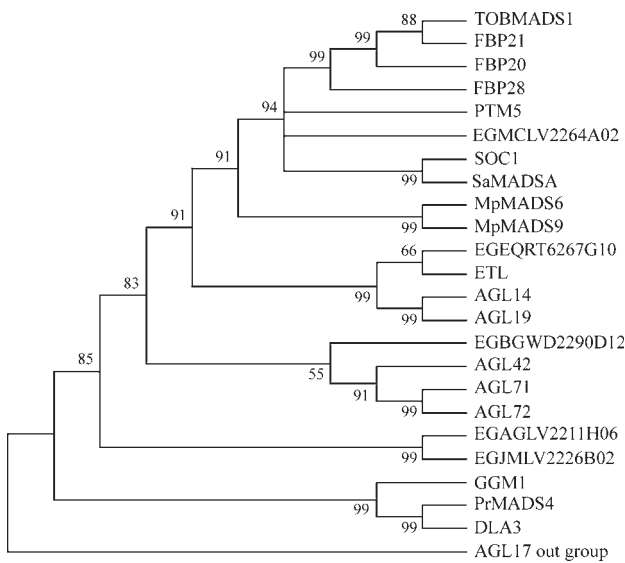
**Figure 3** - Phylogeny of TM3 MADS-box genes. An informative subset of all *TM3*-like genes known has been used in a phylogeny reconstruction. The tree was generated as that of Figure 1, except that conceptual full-length amino acid sequences were used. (AAN52777.1) AGL42/At5g 62615, and (AAN52784.1) AGL71/At5g51870, and (AAN52785.1) AGL 72/At5g51860, and (NP_192925) AGL14/At4g11880, and (NP_194026) AGL19/At4g22950, and (O64645) SOC1/At2g42660 from *A. thaliana*, (AF377868.1) PTM5 from *P. tremuloides*, (AAD16052.1) ETL from *E. globules*, (BAB70741.1) MpMADS6 and (BAB70744.1) MpMADS9 from *Magnolia*, (AAB41526.1) SaMADS from *S. alba*, (CAA53782.1) TOBMADS1 from *Nicotiana*, (S51936) DAL3 from *Picea abies*, (AAQ72502) FBP28, and (AAK21251) FBP20, and (AAK21252) FBP21 from *P. hybrida*, (CAB44447) GGM1 from *Gnetum gnemon*, and (AAB 80807) PrMADS4 from *Pinus radiata*.

hybridization experiment with EGEQRT6267G10.g probes is clearly needed to confirm this hypothesis.

Reads from the other *Eucalyptus* EST-contig (EGB GWD2290D12.g) are exclusively found in the WD2 library. Our phylogenetic analysis showed EGBGWD2290 D12.g to be closely related to AGL42, AGL71 and AGL72 of *A. thaliana*. The expression pattern of these genes in *A. thaliana* has not been described yet.

Another member of the TM3 subfamily found in *Eucalyptus* were assigned to the EST-contigs: EGMCLV2264 A02.g, EGAGLV2211H06.g and EGJMLV2226B02.g. EGMCLV2264A02.g has a broader expression pattern than the other TM3 members. Also, EGMCLV2264A02.g exhibits a high similarity sequence with two known genes: *SOC1* of *A. thaliana* and *PTM5* (*POPULUS TREMULOIDES MADS-BOX 5*) of aspen trees (*Populus tremuloides*). *SOC1* is involved in the mechanism required to promote flowering (Samach *et al.*, 2000) and is expressed to some extent in roots but more abundantly in leaves and flowers. Temporal expression analysis of *PTM5* in staged vascular cambium and other tissues indicated that *PTM5* expression is seasonal and is limited to spring wood formation and rapidly expanding floral catkins (Cseke *et al.*, 2003). Spatial expression analysis using *in situ* hybridization revealed that *PTM5* expression is localized within a few layers of differentiating vascular cambium and xylem tissues as well as the vascular bundles of expanding catkins. The *Eucalyptus* putative ortologous of *SOC1* and *PTM5* is also expressed in roots, as SOC1; in addition, it has an expression related with abiotic and biotic stress in *Eucalyptus*: ST7 library (plantlets under cold stress), ST2 library and LV2 library (leaves under boron and phosphate deficiency rust and canker).

Our phylogenetics analysis has shown two remaining members of the TM3 subfamily in FORESTs databank: EGAGLV2211H06.g and EGJMLV2226B02.g. The phylogenetic analysis suggests a recent duplication event that gave rise to these genes. This is also substantiated by the expression pattern, confined to library LV2. It also suggests that they may act redundantly. This expression pattern suggests a role of these genes in the response to boron/phosphate deficiency or biotic stress. A single MADS-box gene has been related with response to nutrient deficiency, the *ANR1* from *A. thaliana* (Zhang and Forde, 1998).

### AGL17 subfamily

A second subfamily was also analyzed separately, the AGL17 subfamily (data not shown). *ANR1* with three other genes (*AGL16*, *AGL17* and *AGL21*) comprise the AGL17 subfamily in the *A. thaliana* genome (Figure 1). Transcripts of *AGL17, AGL21* and *ANR1* have been detected exclusively in roots. Despite their close relationship, *AGL17* and *AGL21* have shown contrasting mRNA expression patterns in root tissues, suggesting that they are not redundant (Burgeff *et al.*, 2002). On the other hand, *AGL16* mRNA accumulates at high level in leaves and moderate levels in roots and stems (Alvarez-Buylla *et al.*, 2000a). *ANR1* is the only *AGL17*-like gene for which a mutant phenotype is known. Transgenic plants in which *ANR1* expression was blocked failed to respond to nitrate-rich zones in the soil by lateral root proliferation. This revealed that ANR1 is a key component of the signal transduction chain by which nitrate stimulates lateral root proliferation (Zhang and Forde, 1998).

In our analysis, the EST-contigs EGSBRT3311C06.g and EGCCCL1325E06.g were recognized by our phylogenetic analysis as members of the subfamily AGL17. EGSBRT3311C06.g is tightly clustered with *ANR1*, *AGL21* and *AGL17* and is also expressed in root. These associations, however, are tentative and should be regarded with caution. The expression patterns of member of AGL17 subfamily, even in the same family, from other species have been found to vary to a large extent. For instance, transcripts of putative *Antirrhinum* orthologue of *ANR1* (namely DEFH125) have been detected not only in stamens, mainly in the vegetative cell within maturing pollen, but also in the transmitting tract of the carpel (Zachgo *et al.*, 1997). EGCCCL1325E06.g, the other member of AGL17 subfamily, where it is expressed exclusively in the CL1 (*E. grandis* dark formed calli) library.

*STMADS11 subfamily*

Among the three EST-contigs of *Eucalyptus* STMADS11 subfamily found in FORESTs databank, one (EGC CSL1018C10.g) is expressed at the CL1, FB1 (flower buds, flower and fruits), and SL1 (*E. grandis* dark growing seedlings with three hours of light exposition) libraries, but two of them (EGEZWD2255G02.g and EGJMWD2252E02.g) are expressed exclusively in the WD2 library. The *A. thaliana* genome contains two genes in this subfamily, *SHORT VEGETATIVE PHASE* (*SVP*) and *AGL24*. The *SVP* gene probably encodes a dosage-dependent repressor of flowering. The repression prolongs all vegetative stages in wild-type plants independently of photoperiodic control and vernalization (Becker and Theiβen, 2003). Conversely, the *AGL24* seems to be a mediator acting in the genetic pathways from *SOC1* to the floral meristem identity gene *LEAFY* (*LFY*) (Yu *et al*.; 2004). *JOINTLESS* from tomato is a third *STMADS11*-like gene for which a loss-of-function phenotype is known. In jointless mutants, abscission zones on the pedicels fail to develop and accordingly, abscission of flowers and fruits does not occur normally (Mao *et al*., 2000).

Our phylogenetic analysis (Figure 4) shows a group formed by EGCCSL1018C10.g, *SVP*, *JOINTLESS*, *PkMADS1* (*PAULOWNIA KAWAKAMII MADS1*) and *IbMADS3* (*IPOMOEA BATATAS MADS3*). *PkMADS1* was reported to be expressed in the vegetative shoot apex and leaf primordial of *Paulownia kawakamii* (a wood species) (Prakash and Kumar, 2002). Alterations in the antisense transgenic plants indicate that this gene is involved in the regulation of vegetative shoot morphogenesis in *P. kawakamii*. The expression of *IbMADS3* in vascular cambium indicates a role in facilitating cell division and expansion of vegetative tissue during tuber organogenesis (Kim *et al*., 2002). Taking together, this data suggest that EGCCSL1018C10.g may also be involved in regulation of vegetative shoot morphogenesis. On the other hand, *JOINTLESS* functions do not seem to be very similar, which is in remarkable contrast to the close relationship and similarity of the expression patterns among genes of this group.

Of the two STMADS11 subfamilies expressed in the WD2 library, EGEZWD2255G02.g and EGJMWD2252 E02.g, the former has a higher sequence similarity with *AGL24*, *FBP13* (from *Petunia hybrida*), *IbMADS4* (from *I. batatas)* and *STMADS16* (from *Solanum tuberosum*). *IbMADS4* and *STMADS16* genes are expressed in stems and they seem to promote vegetative development in these species (Garcia-Maroto *et al*., 2000; Kim *et al*., 2002). Our phylogenetic analysis was unable to associate EGJMWD 2252E02.g to any other gene with known expression pattern or function described up to now in the STMADS11 subfamily.
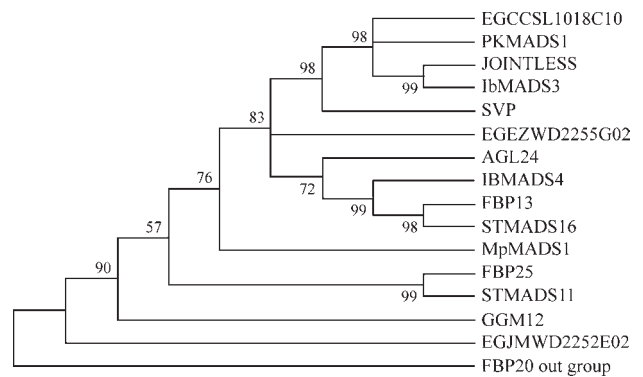


**Figure 4** - Phylogeny of STMADS11 MADS-box genes. An informative subset of all *STMADS11*-like genes known has been used in a phylogeny reconstruction. The tree was generated as in Figure 1, except that conceptual full-length amino acid sequences were used. (AAF22455.1) PkMADS1 from *P. kawakamii*, (CAB79364.1) AGL24/At4g24540, and (AAG24508.1) SVP/At2g22540 from *A thaliana,* (BAC15562.1) IbMADS4, and (BAC15561.1) IbMADS3 from *I. batatas,* (AAB94006.1) STMADS11, and (AAB94005.1) STMADS16 from *S. tuberosum,* (BAB 70736.1) MpMADS1, and (Q9FUY6) JOINTLESS from *Lycopersicon esculentum,* (AAK21250) FBP13, and (AAK21256) FBP25 from *P. hybrida*, and (CAB44458) GGM12 from *G. gnemon*.

## MADS-box of *Eucalyptus* with reproductive-specific expression

*Type II MADS-box genes*

No type I MADS-box EST-contig was identified in the FB1 library. As for the type II MADS-box EST-contigs, a total of 12 were recovered from FORESTs cDNA libraries. The analysis of expression pattern shown in 12 out of 24 type II MADS-box ESTs-contigs are expressed in the flower buds, flower and fruits (FB1 hereafter) library (Figure 2). Only two of these MADS-box EST-contigs are also expressed in other libraries, EGMCFB1109A09.g (RT3, roots of developing plants) and EGCCSL1018C10.g (CL1 and SL1). These two EST-contigs with predominant reproductive expression will be considered as reproductive-specific for the sake of simplicity. This demonstrates the utmost importance of MADS-box genes in the reproductive development of *Eucalyptus*. Remarkably, we were able to identify homologues for almost all *Arabidopsis* "ABCE" (Becker and Theiβen, 2003) class genes among *Eucalyptus* EST-contigs.

*AGL2 subfamily*

Among the four EST-contigs of *Eucalyptus* AGL2 subfamily found in FORESTs databank (Figure 1), EGCBFB1277A03.g and EGEZFB1006G06.g are putative orthologues to *SEPALLATA3* (*SEP3*), while EGEQFB 1200B04.g has high sequence similarity with *SEP1* and *SEP2*. The *SEPALLATA* genes are *AGL2*-like (class E) genes that are known to be directly involved in floral organogenesis. There are four *AGL2*-like genes in the *Arabidopsis* genome. The SEP1; SEP2 and SEP3 proteins possess redundant functions and the expression patterns are

restricted to flower organs, while *AGL3* is expressed in all major plant organs above ground (Huang *et al.*, 1995). In spite of having a higher sequence similarity with *AGL3*, the third EST-contig member of the AGL2 subfamily of *Eucalyptus* (EGEQFB1201E04.g) does not present a similar expression pattern of its putative *Arabidopsis* orthologue, since EGEQFB1201E04.g is expressed exclusively in the FB1 library.

### SQUA subfamily

EGJMFB1092E09.g is expressed in FB1 whereas EGMCFB1109A09.g EST-contig is expressed in both FB1 and RT3 libraries (Figure 2). These EST-contigs correspond to members of the SQUA subfamily that, in *A. thaliana*, are involved with flower initiation and fruit development and are typically expressed in inflorescence or floral meristems. EGJMFB1092E09.g, of *E. grandis*, shows a close phylogenetic relationship with the EAP2 protein of *E. globulus* (Kyozuka *et al.*, 1997) (Figure 1). This protein is a functional equivalent of AP1 *Arabidopsis* protein. Naturally, our results suggest that the EGJMFB1092E09.g EST-contig might also be a functional orthologue of AP1 in *E. grandis*.

### AG subfamily

The members of the AG subfamily are involved in specifying male and female reproductive organs. Alternatively, they are required for the proper development of fruit and ovule identity (Pinyopich *et al.*, 2003). The single AG-like EST-contig of *Eucalyptus* found, EGABFB1059E05.g, presents a higher sequence similarity with SHP1 (Figure 1) and the orthologue status of these proteins is also corroborated by the expression pattern.

### STMADS11 subfamily

Among the EST-contigs of the *Eucalyptus* STMADS11 subfamily, only EGCCSL1018C10.g is expressed at the FB1 library. Reads for this EST-contig were also found in the CL1 and SL1 libraries. Its relationship to other STMADS11 subfamily as well its expression pattern were described in detail before in the vegetative-specific expression genes.

### GGM13 subfamily

There is only one regular *GGM13*-like gene in the *Arabidopsis* genome, termed *ABS*, *TT16* or *AGL32* (Becker *et al.*, 2002). *GGM13*-like genes are assumed to represent the sister group of the B genes (including the *DEF*- and *GLO*-like genes, and their gymnosperm orthologues) and are hence also termed *B sister* (*Bs*) genes (Becker *et al.*, 2002). In contrast to *DEF*- and *GLO*-like genes, which are mainly expressed in male reproductive organs (and angiosperm petals), *Bs* genes are mainly expressed in female reproductive organs, especially in ovules, in both gymnosperms and angiosperms (Becker *et al.*, 2002). EGBMFB1132D01.g is a newly identified member of the subfamily in angiosperms and it is likely to be the single orthologue of TT16 in *E. grandis*.

### DEF and GLO subfamily

Phylogeny reconstructions and evaluation of exon-intron structures indicate that the gene duplication, which led to distinct subfamilies of *DEF*- and *GLO*-like genes, occurred in the lineage that gave rise to extant angiosperms after the gymnosperms split. Within eudicots, further duplications of *DEF*- and *GLO*-like genes occurred several times independently in different lineages (Kramer *et al.*, 1998; Kramer and Irish, 1999; 2000), including *Eucalyptus* (Figure 1). Two orthologous EST-contigs of *PI* (EGUTFB 1102F11.g and EGJFFB1118D11.g) and *AP3* (EGJEWD 2299A04.g and EGEZFB1005C02.g) were identified in the FORESTs databank. Intriguingly, one *AP3 Eucalyptus* paralogous gene is expressed in the WD2 library. The association of *DEF*- and *GLO*-like genes with vascular development has been suggested before in herbaceous plants such as *Eranthis hyemalis* and *Solanum tuberosum* (Skipper, 2002; Garcia-Maroto *et al.*, 1993). Nonetheless, no record has linked the expression of members of this subfamily with wood formation.

## Concluding Remarks

The elevated number of reads in several libraries from different tissues and the low redundancy level of these libraries in the FORESTs transcriptome project contributed to an efficient *in silico* approach for the identification and the evaluation of gene expression patterns in different tissues and conditions in *Eucalyptus*. Computer analysis of the ESTs of vegetative and reproductive MADS-box genes revealed some interesting expression patterns, which may be related to their roles in controlling gene transcription during plant development, tissue differentiation and many stress conditions. Among the 24 MADS-box genes identified in our analysis, 11 are expressed exclusively in vegetative tissues, among them 5 were found in the wood library. Recent evidence strongly suggests that MADS-box members are critical also in regulatory roles during vegetative development, such as embryo, root, or leaf development. It is likely that complex regulatory networks involving several MADS-box genes, similar to those that control flower development, underlie development of vegetative structures, for instance, the wood formation in trees. Understanding of the molecular mechanisms performed by MADS-box proteins underlying *Eucalyptus* growth, development and stress reactions would provide important insights into tree development and could reveal means by which tree characteristics could be modified for the improvement of industrial properties. Our genomic analysis, however, should be regarded with caution, as it represents a preliminary screening of the library that provided an important start for future biochemical investigation of these molecules.

## Acknowledgments

## References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI BLAS: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F and Yanofsky MF (2000a) MADS-box gene evolution beyond flowers: Expression in pollen, endosperm, guard cells, roots and trichomes. Plant J 24:457-466.

Alvarez-Buylla ER, Pelaz Z, Liljegren SJ, Gold SE, Burgeff C and Ditta GS (2000b) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sc USA 97:5328-5333

*Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.

Becker A, Kaufmann K, Freialdenhonven A, Vicent C, Li M-A, Saedler H and Theißen G (2002) A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. Mol Genet Genomics 266:942-950.

Becker A and Theißein G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Physiolog Evol 29:464-489.

Burgeff C, Liljegren SJ, Tapia-Lopez R, Yanofsky MF and Alvarez-Buylla ER (2002) MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots. Planta 214:365-72.

Cseke LJ, Zheng J and Podila GK (2003) Characterization of PTM5 in aspen trees: A MADS-box gene expressed during woody vascular development. Gene 318:55-67.

Davies B, Egea-Cortines M, de Andrade Silva E, Saedler H and Sommer H (1996) Multiple interactions amongst floral homeotic MADS box proteins. EMBO J 15:4330-43.

De Bodt S, Raes J, Florquin K, Rombauts S, Rouze P, Theißen G and Van de Peer Y (2003) Genome-wide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J Mol Evol 56:573-586.

Decroocq V, Zhu X, KauVman M, Kyozuka J, Peacock WJ, Dennis ES and Llewellyn DJ (1999) A *TM3*-like MADS-box gene from *Eucalyptus* expressed in both vegetative and reproductive tissues. Gene 228:155-160.

Eisen MB, Spellman PT, Brown PO and Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863-8.

Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S and Claverie J-M (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res 9:950-959.

Fan HY, Hu Y, Tudor M and Ma, H (1997) Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. Plant J 12:999-1010.

Garcia-Maroto F, Salamini F and Rohde W (1993) Molecular cloning and expression patterns of three alleles of the Deficiens-homologous gene St-Deficiens from *Solanum tuberosum*. Plant J 4:771-80.

Garcia-Maroto F, Ortega N, Lozano R and Carmona MJ (2000) Characterization of the potato MADS-box gene STMADS16 and expression analysis in tobacco transgenic plants. Plant Mol Biol 42:499-513.

Gu Q, Ferrándiz C, Yanofsky MF and Martienssen R (1998) The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development 125:1509-1517.

Hartmann U, Hohmann S, Nettesheim K, Wisman E, Saedler H and Huijser P (2000) Molecular cloning of SVP: A negative regulator of the floral transition in *Arabidopsi*s. Plant J 21:351-360.

Heuer S, Hansen S, Bantin J, Brettschneider R, Kranz E, Lorz H and Dresselhaus T (2001) The maize MADS box gene ZmMADS3 affects node number and spikelet development and is co-expressed with ZmMADS1 during flower development, in egg cells, and early embryogenesis. Plant Physiol 127:33-45.

Huang H, Tudor M, Weiss CA, Hu Y and Ma H (1995) The *Arabidopsis* MADS-box gene AGL3 is widely expressed and encodes a sequence-specific DNA-binding protein. Plant Mol Biol 28:549-567.

Kim SH, Mizuno K and Fujimura T (2002) Isolation of MADS-box genes from sweet potato (*Ipomoea batatas* (L.) Lam.) expressed specifically in vegetative tissues. Plant Cell Physiol 43:314-22.

Köhler C, Hennig L, Spillane C, Pien S, Gruissem W and Grossniklaus U (2003) The Polycom-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. Genes Dev 17:1540-1553.

Kramer EM, Dorit RL and 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 149:765-783.

Kramer EM and Irish VF (1999) Evolution of genetic mechanisms controlling petal development. Nature 399:144-148.

Kramer EM and Irish VF (2000) Evolution of the petal and stamen developmental programs: Evidence from comparative studies of lower eudicots and basal angiosperms. Int Plant Sc 161:S29-S40.

Kumar S, Tamura K, Jacobsen I and Nei N (2000) MEGA: Molecular Evolutionary Genetics Analysis, version 2.0. Pennsylvania and Arizona State University, University Park, Pennsylvania and Tempe, Arizona.

Kyozuka J, Harcourt R, Peacock WJ and Dennis ES (1997) *Eucalyptus* has functional equivalents of the *Arabidopsis* AP1 gene. Plant Mol Biol 35:573-84.

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL and Yanofsky MF (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature 404:766-770.

Ma H, Yanofsky MF and Meyerowitz EM (1991) AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral ho-

meotic and transcription factor genes. Genes Dev 5:484-495.

Mao L, Begum D, Chuang HW, Budiman MA, Szymkowiak EJ, Irish EE and Wing RA (2000) JOINTLESS is a MADS-box gene controlling tomato flower abscission zone development. Nature 406:910-913.

Menzel G, Apel K and Melzer S (1996) Identification of two MADS box genes that are expressed in the apical meristem of the long-day plant *Sinapis alba* in transition to flowering. Plant J 9:399-408.

Michaels SD, Ditta G, Gustafson-Brown C, Yanofsky M and Amasino RM (2003) AGL24 acts as a promoter of flowering in *Arabidopsis* and is positively regulated by vernalization. Plant J 33:867-874.

Münster T, Deleu W, Wingen LU, Ouzunova M, Cacharrón J, Faigl W, Werth S, Kim JTT, Saedler H and Teiβein G (2002) Maize MADS-box genes galore. Maydica 47:287-301.

Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M and Lepiniec L (2002) The *TRANSPARENT TESTA16* locus encodes the *Arabidopsis* BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell 14:2463-2479.

Onouchi H, Igeno MI, Perilleux C, Graves K and Coupland G (2000) Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among *Arabidopsis* flowering-time genes. Plant Cell 12:885-900.

Pelaz S, Ditta GS, Baumann E, Wisman E and Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. Nature 405:200-203.

Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E and Yanofsky MF (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature 424:85-88.

Prakash AP and Kumar PP (2002) PkMADS1 is a novel MADS box gene regulating adventitious shoot induction and vegetative shoot development in *Paulownia kawakamii*. Plant J 29:141-51.

Riechmann JL and Meyerowitz EM (1997) MADS domain proteins in plant development. Biol Chem 378:1079-1101.

Rounsley SD, Ditta GS and Yanofsky MF (1995) Diverse roles for MADS box genes in *Arabidopsis* development. Plant Cell 7:1259-1269.

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-25.

Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF and Coupland G (2000) Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. Science 288:1613-1616.

Shore P and Sharrocks AD (1995) The MADS-box family on transcriptional factors. Eur J Biochem 229:1-13.

Sitnikova T, Rzhetsky A and Nei M (1995) Interior-branch and bootstrap tests of phylogenetic trees. Mol Biol Evol 12:319-33.

Skipper M (2002) Genes from the APETALA3 and PISTILLATA lineages are expressed in developing vascular bundles of the tuberous rhizome, flowering stem and flower primordia of *Eranthis hyemalis*. Annals of Botany 89:83-88.

Tandre K, Albert VA, Sundas A and Engstrom P (1995) Conifer homologues to genes that control floral development in angiosperms. Plant Mol Biol 27:69-78.

Theiβen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter K-U and Saedler H (2000) A short history of MADS-genes in plants. Plant Mol Biol 42:115-149.

Yu H, Xu Y, Tan EL and Kumar PP (2002) AGAMOUS-like 24, a dosage-dependent mediator of the flowering signals. Proc Natl Acad Sci USA 99:16336-41.

Yu H, Ito T, Wellmer F and Meyerowitz EM (2004) Repression of AGAMOUS-like 24 is a crucial step in promoting flower development. Nat Genet 2004 36:157-61.

Walden AR, Wang DY, Walter C and Gardner RC (1998) A large family of TM3 MADS-box cDNAs in *Pinus radiate* includes two members with deletions of the conserved K domain. Plant Sci 138:167-176.

Winter KU, Becker A, Münster T, Kim JT, Saedler H and Theissen G (1999) MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc Natl Acad Sci USA 96:7342-7.

Zachgo S, Saedler H and Schwarz-Sommer Z (1997) Pollen-specific expression of DEFH125, a MADS-box transcription factor in Antirrhinum with unusual features. Plant J 11:1043-50.

Zhang H and Forde BG (2000) Regulation of *Arabidopsis* root development by nitrate availability. J Exp Bot 51:51-59.

Zhang H and Forde BG (1998) An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture. Science 279:407-409.

*Associate Editor: Claudia Monteiro-Vitorello*