

In Silico and Quantitative Analyses of MADS-Box Genes in *Coffea arabica*

Raphael Ricon de Oliveira · Antonio Chalfun-Junior · Luciano Vilela Paiva · Alan Carvalho Andrade

Published online: 26 January 2010
© Springer-Verlag 2010

Abstract MADS-box genes comprise a family of transcription factors that act as key regulators in many cellular development processes of several organisms. Members of this family have highly conserved regions and play important roles as transcription factors, activating target genes. The present work aimed at analyzing the MADS-box gene family present in a database generated by the Brazilian Coffee Genome Project (CAFEST) as well as of observing their respective sites of expression within these data. Through the use of bioinformatics tools, it was possible to identify and classify 26 expressed sequence tag contigs, 13 of them were expressed exclusively in vegetative tissues, 11 in reproductive tissues, and two in both. Later, quantitative analysis by quantitative reverse transcriptase PCR was carried out for three of them belonging to the groups of genes APETALA3 (B genes), AGAMOUS (C genes), and SEPALLATAS (E genes), which could be compared with the expression profile of in silico analysis and of its putative orthologous genes. Therefore, this study aims at a

better understanding of the development processes performed by this family of genes in *Coffea arabica*, mainly in reproductive organs, as well as to compare functions of MADS-box orthologous studied in other species.

Keywords *Coffea arabica* · Floral development · Transcriptome · MADS-box genes · qRT-PCR

Introduction

MADS-box genes consist of a family of transcription factors that play a role as key regulators in many cellular development processes of several organisms such as fungi, plants, and animals. Members of this family have highly conserved regions, around 60 amino acids, which activate the transcription processes binding to recognition elements, so-called CarG boxes (CC (A/T)6 GG), found in promoters of target genes (Riechmann et al. 1996). Alvarez-Buylla et al. (2000a) suggest that there was a duplication of the ancestral MADS gene, before the evolutionary divergence into plants and animals, giving origin to two groups of MADS-box proteins, type I (SRF-like) and type II (MEF2-like), both found in animals, fungi, and plants. The only common denominator of all MADS-box genes is the highly conserved 180 bp, codifying the DNA-binding domain of these transcription factors.

MADS type II proteins are more commonly found in plants and present modular and conservative structural organization, known as MIKC domain, due to the presence of four domains from N- to C-terminal: MADS (M) highly conserved, intervening (I) which corresponds to an internal region that connects to keratin-like domain (K) which, in turn, is responsible for the protein–protein interactions, then followed by a carboxy-terminal (C) region involved in the

Electronic supplementary material The online version of this article (doi:10.1007/s11105-009-0173-5) contains supplementary material, which is available to authorized users.

R. R. de Oliveira · A. Chalfun-Junior (✉)
Department of Biology, Federal University of Lavras,
Minas Gerais, Brazil
e-mail: chalfunjunior@dbi.ufla.br

L. V. Paiva
Department of Chemistry, Federal University of Lavras,
Minas Gerais, Brazil

A. C. Andrade
National Research Center for Genetic Resources and
Biotechnology (CENARGEN),
Brazilian Company of Agricultural Research (EMBRAPA),
Brasília, Brazil

transcription activation. On the other hand, type I proteins found in plants do not possess a well-defined K domain and are subdivided into four subfamilies, M α , M β , M γ , and M δ , this last one being currently accepted closer to the type II proteins (Parenicová et al. 2003).

The MADS-box family can be divided into subfamilies according to similarities among their sequences. These subfamilies represent phylogenetic gene clades in such a way that these genes are either extremely conservative homologous or paralogous originated from a recent duplication. The complete genome sequence of *Arabidopsis* (*Arabidopsis* Genome Initiative 2000) allowed a comprehensive description of the complexity and diversity of MADS genes in plant species. In 18 of these genes, the function is known and well defined by phenotypic analysis of single mutants, or multiple mutants, in the case of redundant genes. All of these genes are involved in development processes and, without exception, are all of MIKC type. Generally, these genes represent monophyletic groups and can be subdivided into ancestral classes (Theißen et al. 2000) which, in eudicotyledon angiosperms, form the following paraphyletic groups: AG, AGL2, AGL6, AGL12, AGL15, AGL17, DEF, FLC, GGM13, GLO, SQUA, STMADS11, and TM3 (Becker and Theißen 2003). The analysis of the *Arabidopsis* genome demonstrates that these classes probably comprise all of the MIKC-type genes.

The great majority of the already identified flowering genes, involving several plant species, encode transcription factors of MADS-box domain. Studies on *Arabidopsis* mutants confirm the ABCDE model proposed for the organization of floral organ, which describes the gene interaction for the floral components (sepals, petals, stamens, and carpels), since the number of combinations among them is enough to characterize the floral organ (Coen and Meyerowitz 1991; Honma and Goto 2001). It was discovered that proteins of MADS domain can be highly associated to form protein complexes, permitting the elaboration of biochemical models (Krizek and Fletcher 2005), for example the quaternary model which proposes tetrameric complexes of MADS proteins for the organ characterization in each flower component (Theißen and Saedler 2001). Recent studies indicate that the association of MADS proteins related to flowering can be the main approach of combinatorial control for the floral organ specification (Krizek and Fletcher 2005).

Other experiments using mutants with loss of function for determined MADS genes presented homeotic transformations, indicating that these genes functions as organ specifiers (homeotic selector genes) during flower ontogeny. Except for some A-class genes, like *Arabidopsis APETALA2*, all of the floral organ identity genes that have so far been molecularly characterized belong to the MADS-

box gene family (Ng and Yanofsky 2001; Theißen 2001). Corroborating the idea that type-MIKC MADS genes are homeotic for several functions of plant development, many studies have reported that genes of A, B, C, D, and E functions, when submitted to phylogenetic reconstructions, fall into separate classes called SQUAMOSA (class A), DEFICIENS and GLOBOSA (class B), AGAMOUS (classes C and D), and AGL2 genes (class E; Theißen et al. 1996, 2000), respectively.

The existence of MADS-box genes in gymnosperms, ferns, and moss, which do not form either flowers or fruits, demonstrates that the role of these genes in plants is not restricted to the development of reproductive organs (Münster et al. 2002b). Studies have demonstrated that MADS-box genes are involved in many aspects of plant development such as meristem identity, flowering time, floral organ characterization, pollen fertility, ovule development, fruit characterization and development, and lateral root elongation (Zhang and Forde 1998). MADS-box genes expressed in leaves, guard cells, and in trichomes were also found (Alvarez-Buylla et al. 2000b).

Currently, all the complexity of flowering network is mainly based on the information obtained from the *Arabidopsis* data (Izawa et al. 2003). However, extensive efforts have been made to describe the process in other related species (Mouradov et al. 2002; Izawa et al. 2003). Recent advances in molecular biology, including the complete sequencing of plant genomes, have made possible detailed comparison of gene sequences allowing the identification of orthologous genes and inferring their likely roles in different species.

It is known that the initiation of flowering depends on the balanced expression of a gene complex network that is regulated by endogenous and environmental factors. However, in the coffee plant, some additional features are associated—such as biennial cycle, dormant buds, induction of flowering by water stress, sequential maturation of flower buds, etc. (Camargo 1985; Rena and Maestri 1985; Pezzopane et al. 2008)—which complicate the understanding of this process. In this way, environmental factors have been extensively studied; however, little is known to create a flowering model in coffee, highlighting the need for a better understanding of genetic factors.

Bearing in mind the importance of coffee cultivation as an international commodity and its peculiar characteristics in the flowering and fruit set, the present study aimed at identifying and analyzing the MADS-box gene family present in the databank of expressed sequence tags (ESTs) generated by the Brazilian Coffee Genome Project, CAF-EST (Vieira et al. 2006). Bioinformatics tools and analysis of quantitative expression were used to identify these genes, determining their probable orthologous and comparison of their expression profiles in different tissues. Therefore, this

work paves the way for understanding the development processes in *Coffea arabica*, also identifying genes of interest for further applied studies that may improve the productive chain of coffee, whether enabling control of flowering time and harvesting or increasing productivity, for example.

Materials and Methods

In Silico Analyses

MADS-Box Sequence Obtainment

The Brazilian Coffee Genome Project (Vieira et al. 2006), which generated ESTs for the CAFEST database and allowed its manipulation through the GeneProject interface (<http://www.lge.ibi.unicamp.br/cafe/>), provided the data source for the collection of MADS sequences. For the selection of reads, two search forms were carried out, one by key word and the other by BLAST tool (Altschul et al. 1997), using the band consensus of MADS region generated by the COBBLER program (Consensus Biasing By Locally Embedding Residues, <http://blocks.fhcr.org/blocks/cobbler.html>) of the MADS-box sequences already published. Thus, all of the reads that presented significant similarity (e value $> 10^{-5}$) were selected.

The more than 100 reads found were clustered by the CAP3 program (Huang and Madan 1999) integrated to the system, forming the EST contigs. For the verification of the MADS-conserved domain, the sequences were inspected by the InterProScan program (<http://www.ebi.ac.uk/InterProScan/>); all of the validated EST contigs are used for a second search, aiming at finding new reads, as well as to remount incomplete clusters. This process was repeated until no more new significant reads were found.

Finally, the most representative clone of each EST contig formed was re-sequenced, confirming the data obtained previously in silico, allowing in addition to that the correction of the incorrect bases or mounting errors.

Phylogenetic Analysis

Classification of MADS sequences found in CAFEST was achieved by comparison with *Arabidopsis* MADS-box genes published by Parenicová et al. (2003). The sequences were aligned using the ClustalW program (Thompson et al. 1994) with the standard patterns, being visually inspected and manually corrected. Segments whose homology could not be adjusted were removed. The final tree was drawn using the MEGA 4 program (Tamura et al. 2007), with neighbor-joining comparison model (Saitou and Nei 1987), and the validity of the tree, regarding the phylogenetic

distance of the clusters, was tested by the bootstrap probability test (Sitnikova et al. 1995).

Digital Northern

For the in silico expression analysis, frequencies of reads that form each EST contig in the libraries in which they were expressed were calculated. This procedure required that the data were previously normalized to give a more accurate idea of the expression degree of the sequences in each treatment and plant organ when all of the libraries were considered in this work.

The normalization consisted in multiplying each read by the quotient between the number of reads from the library where it was expressed and the sum of reads of all libraries where expression was found. The sequences were reorganized according to their classification group, and their data were cast into a matrix relating genes to libraries. The EST contigs and the libraries were grouped into a hierarchical clustering and visualized by the associated programs, Cluster and TreeView (Eisen et al. 1998). Results of expression were presented in a grayscale spectrum where no expression or negative was represented by the lightest (gray) region, being gradually increased until reaching the black, which represented the maximal degree of positive expression.

Identification of Common Grouping Motifs

In order to discover cluster motifs between the MADS sequences selected from CAFEST, the MEME program (Multiple Expectation Minimization for Motif Elicitation, <http://meme.sdsc.edu/meme/meme.html/>) version 3.5.4 (Bailey and Elkan 1994) was used. The total sequence of protein amino acids was compared according to the cluster order observed in the phylogenetic tree, this way facilitating the observation of cluster motifs in each subclass. The parameters used were: any number of replications, maximum number of motifs set to 20, and optimum width between 6 and 200 (Parenicová et al. 2003).

The motif annotations regarding the functional domains of the MADS-box proteins were carried out using the SMART software (Simple Motif Architecture Research Tool, <http://smart.embl-heidelberg.de/>) version 5.0 (Schultz et al. 1998; Letunic et al. 2002), capable of recognizing more than 500 domains present in different protein types.

Quantitative Analyses (qRT-PCR)

Plant Material

Plant material used in this experiment was obtained from cultivar Rubi of coffee plants in reproductive stage,

cultivated without shade protection and in moist tropical climate. Four different tissues were collected: young leaves, roots, inflorescence, and fruit, the latter two being collected each in three different development stages G4, G6, FL and F2, F4, M2, respectively (Morais et al. 2008).

The collection was formed by the material harvested from three different plants for each of the aforementioned tissue, constituting our biological replicates, with the exception of the roots for which three particular plants, cultivated in plastic bags with common soil in greenhouse and measuring about 0.5 m in height, were used. Harvesting occurred between May and October 2007, and the tissues were dipped immediately into liquid nitrogen.

RNA Extraction and cDNA Synthesis

Total RNA extraction of the treatments was carried out by the methodology described by Birtic and Kraner (2006) used for plants with high phenol content, polysaccharides, and RNases like the coffee plants, with small changes to adapt to the laboratory procedures.

Samples were treated with 1 μ l RNase inhibitor (Invitrogen), constituting our RNA stock solutions, and aliquots of these solutions containing 10 μ g RNA were treated with 1 μ l DNase (Invitrogen), as suggested by the manufacturer's, without EDTA addition. The integrity and purity of the samples were assessed by agarose gel electrophoresis and spectrophotometer quantification. All of them presented high degree of integrity and purity ($1.8 < A_{260}/A_{280} < 2.0$); thus, they were used later for the synthesis of cDNA single strand.

Synthesis of cDNA for the collected tissues was performed according to the first-strand cDNA synthesis protocol using Superscript II (Invitrogen) and 1 μ g total RNA (treated with DNase I) used as template in a reaction of 20 μ l mixture reagent.

Tests with two cDNAs dilutions, 1:50 and 1:100, showed that the concentrations did not affect the amplification levels for the primers; therefore, 2 μ l of a more diluted template was used for the SYBR Green PCR (final dilution 1:500).

Primers

The respective clones containing the more representative EST contig fragment (read) selected from CAFEST were sequenced again for confirmation. For that, the Primer3 program found at <http://fokker.wi.mit.edu/primer3/input.htm> was used, determining the parameters for amplification of PCR products between 80 and 140 bp, the annealing temperature (TM) around 60°C, and a G/C content between 45% and 55%.

The specificity of the PCR products generated for each primer was tested in several ways: through the alignments between the primers (BLASTn) and different databases (CAFEST and NCBI); analysis of dissociation curve, gradient of annealing temperature in which the fluorescence decreases at a discrete temperature indicating the separation of the double-stranded DNA; and agarose gel electrophoresis of PCR products to verify the size of the band and presence of nonspecific products.

Quantitative Reverse Transcriptase PCR

The quantitative reverse transcriptase PCRs (qRT-PCRs) were carried out on an Applied Biosystem 7500 Fast Real-Time PCR System using SYBR® Green. Reactions contained 2 μ l cDNA, 7.6 μ l SYBR Green PCR Master mix reagent, and 0.2 μ l for each of the forward and reverse primers, totalling 10 μ l. Programming of the cycles for amplification included 10 min at 95°C for polymerase activation, followed by 45 cycles at 95°C for 15 min, and at 60°C for 1 min.

Data were analyzed using Applied Biosystems SDS 2.1 software to determine C_t (cycle threshold) values. PCR efficacy was estimated for each gene by LinReg program, which uses the captured data of absolute fluorescence during the exponential phase of the amplification of each reaction, using the equation $(1 + E) = 10^{(-1/\text{slope})}$ (Ramakers et al. 2003). All primer pairs showed efficacy higher than 0.80. Efficacy values were taken into consideration in all subsequent calculations.

The cDNA levels in the samples were normalized by amplification of Ubiquitin gene since their expression is considered very stable for being constitutive in the tissues (Warrington et al. 2000). The reaction followed the same procedures of the primer specificity test (dissociation curve) and could be accessed by geNorm program, which was employed in the normalization of the expression data, taking into consideration the differences in the RNA quality and in the efficacy of qRT-PCR within the samples. Expression values were estimated by the equation $(1 + E)^{-\Delta(\Delta C_t)}$ (Livak and Schmittgen 2001). Relative standard deviation was calculated using C_t values; the values above 25% were discharged and the analyses were repeated.

Results and Discussion

After eliminating the false and incomplete sequences, a group of 26 EST contigs were selected, which presented significant similarity to the MADS-box domain (e value $> 10^{-5}$). Some sequences that appeared incomplete in direction to the C-terminal region (CaC02, CaS01, CaS10, CaS14, CaS21, CaS22), but presented a perfect MADS

domain, were kept because they were not differentiated in the classification when only MADS domain was considered (data not shown).

Analyzing the phylogenetic relations found in this work (Fig. 1), all of the sequences could be classified as transcript of MADS-box type II proteins. The great majority of them belong to MIKC group in their different subfamilies and only one (CaC13) was similar to M δ group, in the present accepted as type II genes (Parenicová et al. 2003). The sequences could be visualized by the analyses of motif grouping (Fig. 2) where common domains are observed among them together with MADS-box genes published from *Arabidopsis* used for classification in the phylogenetic tree. This procedure allowed, after the motif annotation, the confirmation of the presence of MADS-box domain, highly conserved and common to all of the sequences, and also the visualization of the regions that differed within each subgroup.

From the 26 probable coffee MADS-box genes considered, 11 of them were expressed in reproductive organs, 13 in vegetative organs, and two in both (Fig. 3). In the contigs expressed in the vegetative organs, the expression profile had a very wide occurrence, appearing in several libraries, while it was very restricted to contigs expressed in libraries of reproductive organs. This corroborates the importance of this family for the development of the plant as a whole, as well as and especially during reproduction phase (flowering and fructification) where they appear as key elements in the activation or repression of genic cascades.

qRT-PCR analysis of the contigs CaC03, CaC12, and CaC14—putative orthologous to the *Arabidopsis* genes AGAMOUS, APETALA3, and SEPALLATA3, respectively—indicates their levels of expression in different tissues (Fig. 4). This result can be compared to the in silico expression analysis (Fig. 3), in addition to validate it, giving more evidence to prove their homology hypothesis. Although, the precise qRT-PCR expression levels of these transcripts were different when compared to the in silico analyses, the results are similar regarding the expression sites. These changes in expression levels can be explained by differences in methods used as RNA source in the CAFEST, such as different collected tissue phases, RNA extraction, and cDNA synthesis methods. In addition to that, most likely, the normalization was not sufficiently representative for the ESTs numbers that should be really found in the determined tissues.

Groups Expressed in Vegetative Organs

TM3-Like

Members of this subfamily are preferably expressed in vegetative parts of the plant, either in angiosperms or in

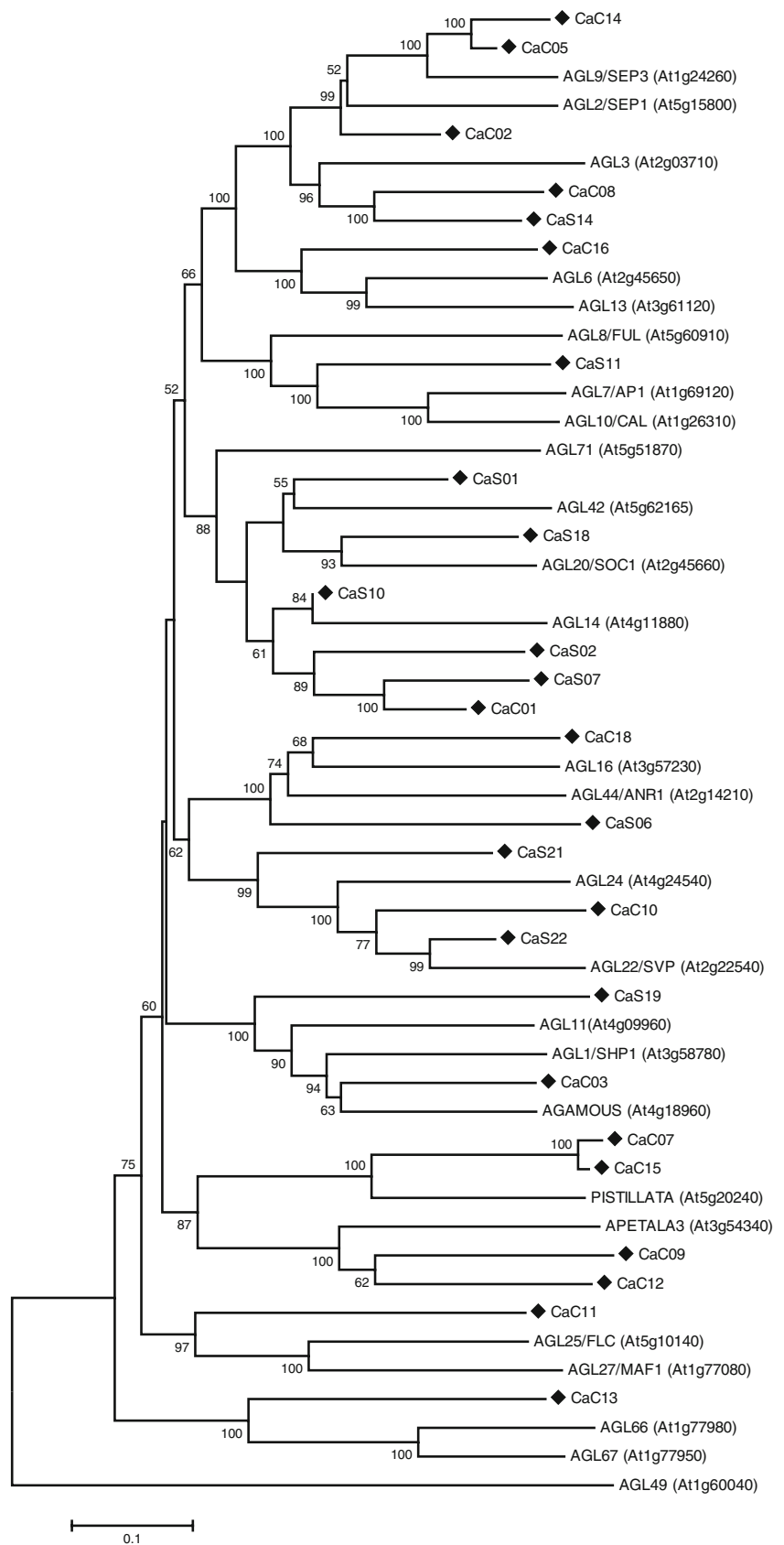
gymnosperms (Winter et al. 1999), even though cases of expression in reproductive organs were reported (Münster et al. 2002a; Heuer et al. 2001). *Arabidopsis* presents six genes related to this class, *AGL14* and *AGL19*, exclusively expressed in root (Rounsley et al. 1995), *AGL20* or *SOC1* (*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*) involved in the mechanism required to promote flowering, vernalization, and autonomous pathways of floral induction and to integrate photoperiod signals (Borner et al. 2000), being expressed in some root extensions but more abundantly in leaves and flowers (Onouchi et al. 2000; Samach et al. 2000), besides the less studied *AGL42*, *AGL71*, and *AGL72*. Likewise, STMADS11 class, TM3 genes can constitute an ancient class with ancestral function in the vegetative development (Becker and Theißen 2003).

In our analyses, this was the class with higher number of classified contigs, with a total of six. They all presented expression in non-reproductive organs in accordance with the majority of the reports published for this class. CaS01 and CaS18 contigs showed being homologue to *AGL42* and *SOC1* genes, respectively (Fig. 1), the latter being differentiated for not presenting expression in flowers, but highly expressed in the non-embryogenic callus library (Fig. 3). The four other contigs, CaS02, CaS07, CaS10, and CaC01, joined to *AGL14* gene (Fig. 1) expressed in roots, but presented expression profile very diversified, appearing in several vegetative libraries and suggesting probable participation in the plant development.

AGL17-Like

AGL17 class is reported as being preferably expressed in roots, and in *Arabidopsis*, four genes belonging to this class were found: *AGL16*, *AGL17*, and *AGL21*, besides the *ANRI* gene. The only reported mutant of this class has precisely the *ANRI* gene blocked and as a phenotype does not respond with the lateral root proliferation toward the more nitrate-rich regions present in the soil, suggesting that this gene is a key element in the signal transduction chain through which the roots are stimulated by nitrate (Zhang and Forde 1998). The expression of *ANRI* gene has been detected only in roots, which is also true for *AGL17* and *AGL21* genes (Burgeff et al. 2002; Rounsley et al. 1995). On the other hand, there are studies that demonstrate additional patterns of expressions besides roots for other constituents of the class, as occurs to *DEFH125* genes of *Antirrhinum* which are detected not only in the vegetative cells mainly involved in the maturation of pollen in stamens but also in the trait of transmission of carpels (Zachgo et al. 1997), *ZmMADS2* gene of maize, restricted to both pollen and root (Heuer et al. 2000), and *AGL16* gene, which presents high levels of transcript accumulation in guard

Fig. 1 Phylogenetic tree relating MADS sequences found in CAFEST (*diamond*; sequences submitted to GenBank and accessions numbers are as [Electronic supplementary material](#)) and published MADS MIKC-type genes. Neighbor-joining comparison model was used with the distance p method and pairwise suppression. Bootstrap values smaller than 50% were omitted



cells and trichomes and can be also expressed in root epidermic cells (Alvarez-Buylla 2000a).

Two EST contigs joined this class, sharing sequence homology to *AGL16* and *ANRI* or *AGL44* (Fig. 1). The CaC18 contig, close to *AGL16* gene, is very similar regarding the motifs found (Fig. 2) and shows varied expression profile, always related to treatment of stressed cells, suggesting a hydric relation when compared to the expression observed in the *Arabidopsis* homologue. The CaS06 contig shares homology to *ANRI* gene, but because of its expression in the library of cells treated with acibenzolar-*S*-methyl (BHT), a powerful activator of systemic resistance induced against several diseases (Silva et al. 2000), doubt remains regarding its function since it is reported to be expressed only in roots in *Arabidopsis*.

STMADS11-Like (SVP)

Three sequences belonging to this class were found in coffee plants when compared to *Arabidopsis* genes (Fig. 1). CaS21 contig close to *AGL24* gene and CaC10 and CaS22 contigs close to *AGL22* or *SVP (SHORT VEGETATIVE PHASE)* and *AGL24* genes are found in *Arabidopsis* and expressed in young leaves and throughout the apical meristem, being abolished in the inflorescence meristem. The mutant for *SVP* gene shows precocious flowering, seeming to undergo rapidly through the vegetative development stages (Hartmann et al. 2000). This gene probably codes a flowering repressor which prolongs all vegetative stages in wild plants, independently of photoperiod or vernalization. During the flower development, expression of this gene first disappears and then the sepals appear (Hartmann et al. 2000). *AGL24* gene is strongly expressed in apical meristem and in primordial leaves. During the floral transition, the expression is also detected in the floral meristem and later on in the whole inflorescence meristem, stamens, and carpels (Yu et al. 2002).

All of the EST contigs grouped in this subfamily presented expression pattern similar to their published homologues, since they were found expressed in vegetative tissues (Fig. 3), without the expression in floral buds, as occurs with *AGL24*. This can be explained by the fact that this gene is expressed in initial phases during the transition of the vegetative to the reproductive meristem, a phase which was not monitored in the CAFEST Project.

FLC-Like

The only EST contig found in CAFEST belonging to this subfamily was CaC11, which showed to be orthologous to *FLC (AGL25)* and *MAF1 (AGL27)* genes of *Arabidopsis* (Fig. 1). *FLC (Flowering Locus C)* gene is reported as a

repressor of flowering since its null mutant presents precocious flowering (Michaels and Amasino 1999), and it has been interpreted as a convergence point for environmental factors and endogenous pathways that regulate the flowering time. It is expressed in all parts of the plants except in inflorescences, being regulated mainly by vernalization, but also by an autonomous pathway and by the *FRI (Frigida1)* allele expression (Sheldon et al. 2000). *MAF1 (MADS AFFECTING FLOWERING 1)* gene also functions as flowering inhibitor, but contrary to *FLC*, its expression is not affected by the same factors. Its transcript is detected in shoots, later restricted to roots and apical meristem.

Furthermore, after the transition to flowering, *MAF1* is found expressed in primordial floral organ (Ratcliffe et al. 2001). Becker and Theißen (2003) have already reported that genes of *FLC* subfamily are not found outside Brassicaceae family, and because many other species respond to vernalization, they suggest that these genes constitute a relatively recent evolutionary response to winter, indicating its end and the beginning of a more favorable season for development and propagation.

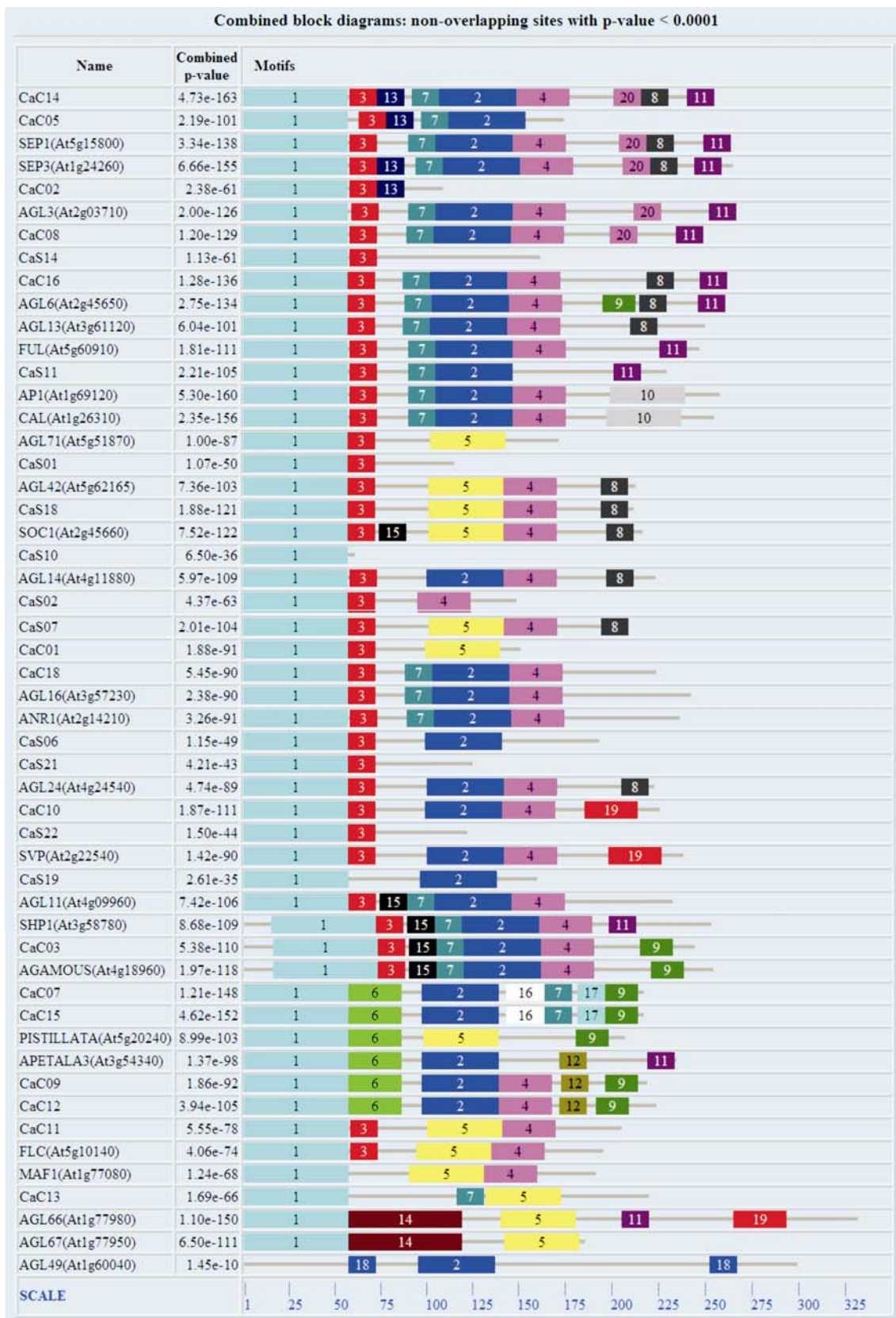
CaC11 EST contig showed expression pattern in libraries of vegetative tissues submitted to these stresses, but in disagreement with their possible homologues in *Arabidopsis*, this profile was not widely spread (Fig. 3). In spite of this, the analyses of motifs among the sequences show high similarity to *FLC* gene, evidencing their phylogenetic proximity (Fig. 1). In coffee plants, expression of the homologue to *FLC* gene seems to have been influenced by the condition of water stress. This could be an adaptive characteristic of tropical plants since in regions of tropical climate, the longest periods of drought are associated to the lowest temperature throughout the year (dry winter and moist summer). That is, drought conditions could be acting in a direct form in the activation of *FLC* gene expression or indirectly, concerning the pathway relative to vernalization, for example.

Groups Expressed in Reproductive Tissues

AGL6-Like

This subfamily appears grouped together to *SQUA* and *AGL2* genes in all relevant phylogenetic reconstructions

Fig. 2 Graphic representation showing the complete grouping motifs found in CAFEST using MEME program (*Multiple Expectation Minimization for Motif Elicitation*, <http://meme.sdsc.edu/meme/meme.html>). The parameters used were: number of replications—any maximum number of motifs—20 and optimum amplitude between 6 and 200. Motifs were annotated by SMART (*Simple Motif Architecture Research Tool*, <http://smart.embl-heidelberg.de/>): Motif 1—Mads, Motif 3—Intervening, Motif 3 C-terminal until Motif 2 C-terminal - K-box, Motif 2 C-terminal until the end of the sequences—C-domain



(Lawton-Rauh et al. 2000; Becker and Theißen 2003). There are reports where the expression of *AGL6* genes under CaMV35S promoter control has drastic effects in *Arabidopsis*, such as the extremely reduced size of the plant, very early flowering, and formation of terminal flowers (Becker and Theißen 2003). However, its function is still unknown since no mutant for loss of this function has been documented.

In *Arabidopsis*, two representatives of this class are found. Even though they are phylogenetically very close and considered as having a recent duplication, their expression sites are very different. While *AGL6* gene is expressed in the four floral whorls (Mouradov et al. 1998), *AGL13* gene is restricted to ovules (Rounsley et al. 1995). Other homologues described, like *ZAG3* gene of maize, are expressed in carpels and floral organs, but not in stamens (Mena et al. 1995), and in *Gnetum*, *GGM9* and *GGM11* genes are expressed in the reproductive cones of male and female parts (Winter et al. 1999).

In the coffee plants, only CaC16 contig was clustered to this subfamily, which is intimately related to the *AGL6* gene and also shows similar motifs (Fig. 1), for this reason being considered its probable homologue. However, lack of data for each isolated flower part is an obstacle to finding a more exact approach of its homology to the *Arabidopsis* gene.

SQUA-Like

Several studies revealed that the genes documented for this clade are essential to the initial activation of floral development mechanism and later in the perianth formation (Mandel et al. 1992; Bowman et al. 1993; Theißen et al. 1996). According to Litt and Irish (2003), the participation of *SQUA* genes in perianth formation is recent, occurring only in more recent angiosperms. The genes belonging to this clade are known as A-function because in the classical ABC model described by Coen and Meyerowitz (1991) for floral organ identity, they participate in the formation of the first and the second whorls, one not always distinct from another.

Parenicová et al. (2003) described four genes in *Arabidopsis* clustered to the SQUA subfamily: *APETALA1* (*API* or *AGL 7*), *FRUITFULL* (*FUL* or *AGL 8*), *CAULIFLOWER* (*CAL* or *AGL 10*), and *AGL79*, the first three being functionally characterized through studies on mutants. *API* and *CAL* mutants are reported as partially redundant genes (Becker and Theißen 2003) since homozygote plants with mutations in the *CAL* gene present wild phenotype, but together with mutants of the *API* gene produce the typical *CAULIFLOWER* phenotype, characterized by an extensive meristem proliferation in each position that would originate a single flower (Kempin et al. 1995). *FUL* gene, on the other hand, seems to be a valve identity

gene because the main defect of its mutant occurs in the elongation and differentiation of valve cells in the fruits, giving them an aspect of a compact fruit (Gu et al. 1998). The major part of silique is constituted by carpel valves, and even though the mutant does not present specific inflorescence phenotype, it is one of the first genes expressed after transition to flowering. According to that, triple mutant presents the same phenotype as the *API/CAL* mutant, but extremely accentuated, indicating functions partially redundant among these three genes during floral initiation (Férrandiz et al. 2000).

In our analyses, only one sequence, CaS11, found in the coffee plants was positioned into this subfamily (Fig. 1), which by the analyses of common grouping motifs showed higher similarity to the *FUL* gene (Fig. 3). In addition to

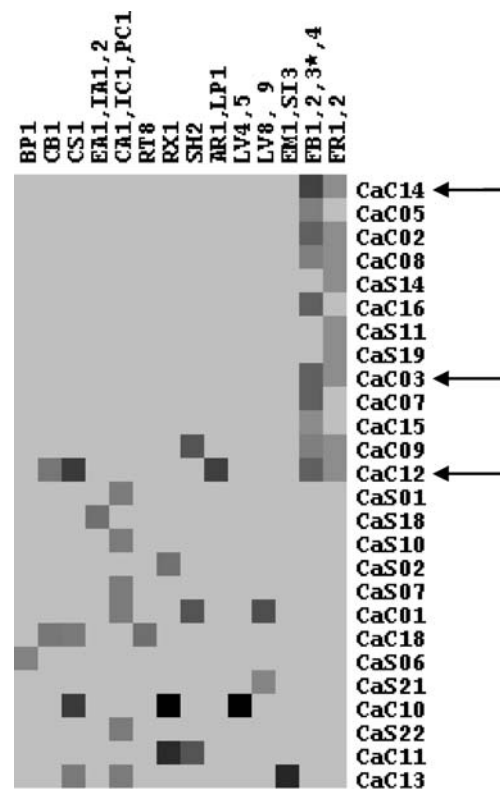
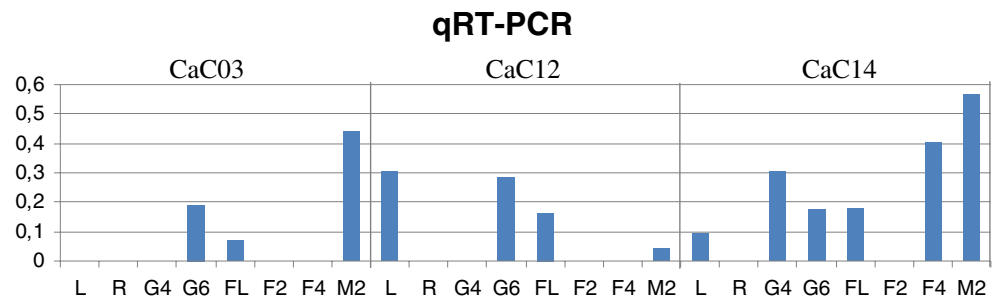


Fig. 3 Digital Northern represented by grayscale expression levels of EST contigs in different libraries. The darker the shade, the higher the expression. BP1—Suspension of cells treated with acibenzolar-S-methyl; CB1—Suspension of cells treated with acibenzolar-S-methyl and brassinosteroids; CS1—Suspension of cells treated with NaCl; EA1/IA1,2—Embryogenic callus; CA1, IC1, PC1—Non-embryogenic callus with and without 2,4-D; RT8—Suspension of cells treated stressed with aluminum; RX1—Thallus infected with *Xylella* spp.; SH2—Plants stressed by hydric deficit (tissue pool); AR1,LP1—Plantlets and leaves treated with araquidonic acid; LV4,5—Young leaves of ortotropic branches; LV8,9—Mature leaves of plagiotropic branches; EM1,SI3—Seeds germinating; FB1,2,3*,4—Floral buds in different stages; FR1,2—Fruit in different stages (Vieira et al. 2006). For comparisons, arrows indicate EST contigs selected for qRT-PCR. *FB3 library was included after the reference article was published

Fig. 4 qRT-PCR for EST contigs CaC03 (class C), CaC12 (class B), and CaC14 (class E). Tissues: leaf (L), root (R), floral bud stages (G4, G6, FL), fruit stages (F2, F4, M2). The latter two being collected each in three different development stages according Morais et al. 2008



that, reinforcing this orthologous hypothesis, this sequence was expressed only in fruits, the same local reported by another authors of high influence and activity of the FUL gene necessary for a normal fruit development (Gu et al. 1998).

AGL2-Like

AGL2 group, more known as *SEPALLATAs*, includes genes called E-function genes because they do not play or interfere in the route of B and C genes but are also required for the differentiation of the four flower whorls (Pelaz et al. 2000; Honma and Goto 2001). The discovery that the SEP proteins can form highly organized complexes with A-, B-, or C-function proteins and that certain versions of these complexes are enough for the transformation of leaves into floral organs suggests that AGL2 genes play a role as mediators among genes of floral meristem and identity of floral organ (Savidge et al. 1995).

Five EST contigs were classified into the AGL2 subfamily (Fig. 1), all presenting a pattern of expression in reproductive organs (Fig. 3). CaC02, CaC05, and CaC14 contigs appear as homologues to the first described *SEPALLATAs*, while CaC08 and CaS14 contigs are homologues to the *AGL3* gene, currently called *SEP4*. Studies on grouping motifs (Fig. 2) indicated that the CaC14 transcript gene is the probable orthologue of *SEP3* gene of *Arabidopsis*, also in agreement with its profile of expression in flowers (Flanagan and Ma 1994). For this sequence, a quantitative qRT-PCR analysis was performed, and the results will be discussed later in this manuscript. For the *AGL3* gene, CaC08 EST contig was found to be the probable orthologue because it possesses a very similar sequence, but its expression profile appears restricted to reproductive organs and is not expressed in other vegetative tissues, as proposed by Huang et al. (1995).

AG-Like

The *AGAMOUS* (*AG*) gene was the first homeotic gene cloned (Yanofsky et al. 1990) given name to the group. The genes of the AG subfamily are involved in the specification of reproductive organs, both male and female, differentiat-

ing them from the vegetative organs (Winter et al. 1999), and as a result, they are called C-function genes. Within the ABC model, C function is known as antagonistic to B function and involved in specification of stamens and carpels, besides promoting floral determination. In *Arabidopsis*, three more representatives are found, *AGL1* and *AGL5*, more known as *SHATTERPROOF1 e 2* (*SHP1 e 2*), and *AGL11* or *SEEDSTICK* (*SDK*).

Wide genomic analyses comparing expression profiles in inflorescences of wild *Arabidopsis* and mutants for *AG* showed that more than 1,000 genes are regulated by *AG*, especially in stamens (Wellmer et al. 2004). Other studies involving microarray analysis and using lines induced by *AG* activity permitted to examine the immediate targets during the initial organogenesis (Ito et al. 2004; Gómez-Mena et al. 2005). Furthermore, the *AG* gene was also described as controlling microsporogenesis (pollen induction) by the *SPOROCTELESS* (*SPL* or *NOZZLE*) gene regulation, a gene that codifies a probable transcription factor (Ito et al. 2004). Other genes related to this subfamily seem to be linked to fruit development, as happens with the already mentioned *SHP1 e 2*, and *SDK*, more specific for ovule development and mistakenly understood as a D-function gene (Favaro et al. 2003; Pinyopich et al. 2003).

In this work, were found two transcripts probably belonging to the AG subfamily, CaC03 and CaS19 (Fig. 1). The expression pattern of the AG genes found in coffee plants (Fig. 3) agrees with the expression sites characteristics of this subfamily since they appear expressed only in reproductive organs, either in flowers or in fruits. CaC03 sequence appears phylogenetically close to the *AG* gene of *Arabidopsis* also expressed in floral organs and fruits. CaS19 EST contig seems to have specific function in female parts of the flower, expressing only in fruits, which agrees with its phylogenetic proximity to *SDK* and *SHP*, both expressed in the fruit and playing a role in its development.

DEF/GLO-Like

Several studies have demonstrated that DEF/GLO subfamilies originated from the duplication of an ancestral gene in the basal group of angiosperms (Winter et al. 2002a) and

therefore are related as paralogous and constantly clustered as a group. To this class, two *Arabidopsis* genes are included, which are intimately related, *APETALA3* (*AP3*) and *PISTILLATA* (*PI*; Bowman et al. 1989), and in *Antirrhinum* by the respective homologues, *DEFICIENS* (*DEF*) and *GLOBOSA* (*GLO*; Sommer et al. 1990; Tröbner et al. 1992). In both species, these genes have presented functions of heterodimer formation for the establishment of petal and stamen identity (Riechmann et al. 1996; Zachgo et al. 1997).

Genes belonging to this class, formed by the two subfamilies DEF and GLO-like genes, are denominated B genes with probable ancestral function of differentiation of male reproductive organs from the female organs (Winter et al. 2002a). After the establishment of the distinct genes during angiosperm evolution, they were also recruited to specify petal identity (Winter et al. 2002b).

In the coffee plants, four sequences clustered into this class were found (Fig. 1). In the groups of GLO-like subfamily genes, CaC07 and CaC15 EST contigs showed up, being orthologous to the *PI* gene, a hypothesis supported by the presented expression pattern (Fig. 3); that is, they are expressed only in the floral organ library (FBs). In the DEF-like subfamily gene, on the other hand, two other contigs were grouped along, CaC09 and CaC12, probable orthologues of the *AP3* gene, but both presenting differentiated expression patterns as besides the expression in floral organs, they were also expressed in vegetative organs (Fig. 3).

Quantitative Analyses

Aiming at validating our *in silico* obtained data and mainly to obtain more precise data of expression sites of some genes of our interest, analyses in qRT-PCR was performed for the representative genes of our phylogenetic tree. The contigs tested were CaC12, CaC03, and CaC14, respectively, belonging to the B-, C-, and E-functional groups.

From the analysis of Fig. 4, we can see that the three EST contigs had differentiated expression occurring mainly in reproductive organs but also appearing in vegetative organs. The transcript relating to CaC03 EST contig appears quantitatively expressed in different phases of floral organ and fruit development, but mainly expressed in the late stage of fruit development. For the CaC03 EST contig, its respective transcript is expressed besides the reproductive organs, also in leaf tissues, a fact not observed by *in silico* analysis visualized by Digital Northern (Fig. 3). Expression levels were high in the intermediate and final stages of floral organ development (G6 and FL stages according to Morais et al. 2008) and also in leaves. Finally, the transcript representative of CaC14 EST contig showed a wider expression, appearing in almost all of the phases of

floral organ and fruit development, mainly in the initial stages of the first and in the final stages in the last, and a small expression in leaf tissue.

These data, even though not precisely confirming the result obtained by Digital Northern, show that the libraries where these sequences are expressed have real expression in that library, but not with the accuracy obtained by qRT-PCR. This way, the results seem to be absent or inverted because they are not specific, which added to the fact that global analyses, using bioinformatics tools, are important in the identification of widely expressed genes, but fail at the expression level that they present in the tissues.

Conclusion

The high number of reads in the many libraries of different tissues and the low number of redundancy levels of these libraries in the transcriptome CAFEST project contributed for an efficient *in silico* identification approach and validation of MADS-box family genes in different tissues and treatments. Computational analyses of ESTs of probable MADS genes, found in vegetative and reproductive tissues, revealed interesting expression profiles which can be related in their routes in the gene transcription control during the plant development, tissue differentiation, and other stress conditions. From 26 MADS-box transcript identified in the analyses, 13 were expressed exclusively in vegetative tissues, 11 in reproductive tissues, and two in both.

The fact that transcript of MADS type I proteins were not found can be explained by the hypothesis of low expression level or because they were expressed under conditions not monitored in transcriptome projects (Dias et al. 2005). Parenicová et al. (2003) reported that it would be possible that members of MIKC group controlled many phenotypic effects obvious and redundant in different studied mutants (Férandiz et al. 2000; Pelaz et al. 2000; Liljegren et al. 2000), but the M α , M β , M γ , and M δ groups would control much more subtle functions, probably not always related to development.

Recent evidences suggest that MADS-box members, besides being fundamental in the development of reproductive organs, are also decisive in regulatory routes during vegetative development, such as the embryo, root, and leaf development. This increases the importance of understanding the role of genes of this family in the plant development, opening many possibilities for the increment and improvement of coffee industry. The knowledge of molecular mechanisms performed by MADS-box proteins responsible for the flowering time, fruit ripening, development, and response to stresses, for example, would promote important understanding on the plant development and

could reveal means through which coffee plant characteristics could be modified for the improvement of industrial properties.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST, PSI BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, de Pouplana LR, Martinez-Castilla L, Yanofsk MF (2000a) An ancestral MADS-box gene duplication occurred before the divergence of plants, animals. *PNAS* 97:5328–5333
- Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsk MF (2000b) MADS-box gene evolution beyond flower: expression in pollen endosperm grand cells roots, trichomes. *Plant J* 24:457–466
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plants *Arabidopsis thaliana*. *Nature* 408:796–815
- Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceeding of the Second International Conference on Intelligent Systems for Molecular Biology*. AAAI, Menlo Park, pp 28–36
- Becker A, Theißen G (2003) The major clades of MADS-box genes, their role in the development, evolution of flowering plants. *Mol Phylogenet Evol* 29:464–489
- Birtic S, Kranner I (2006) Isolation of high-quality RNA from polyphenol- polysaccharide-, lipid-rich seeds. *Phytochem Anal* 17:144–148
- Borner R, Kampmann G, Chandler J, Gleißner R, Wisman E, Apel K, Melzer S (2000) A MADS domain gene involved in the transition to flowering in *Arabidopsis*. *Plant J* 24:591–599
- Bowman JL, Smyth DR, Meyerowitz EM (1989) Genes directing flower development in *Arabidopsis*. *Plant Cell* 1:37–52
- Bowman JL, Alvarez J, Weigel D, Meyerowitz EM, Smyth DR (1993) Control of flower development in *Arabidopsis thaliana* by *APETALA1*, interacting genes. *Development* 119:721–743
- Burgeff C, Liljegren SJ, Tapia-López R, Yanofsky MF, Alvarez-Buylla ER (2002) MADS-box gene expression in lateral primordia meristems, differentiated tissues of *Arabidopsis thaliana* roots. *Planta* 214:365–372
- Camargo AP (1985) Florescimento e frutificação do café arábica nas diferentes regiões cafeeiras do Brasil. *Pesquisa Agropecuária Brasileira* (Brasília) 20(7):831–839
- Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353:31–37
- Dias BFO, Simões-Araújo JL, Russo CAM, Margis R, Alves-Ferreira M (2005) Unravelling MADS-box gene family in *Eucalyptus* spp.: a starting point to an understanding of their developmental role in trees. *Genetics* 28(3):501–510
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis, display of genome-wide expression patterns. *PNAS* 95(25):14863–14868
- Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, Yanofsky MF, Kater MM, Colombo L (2003) MADS-box protein complexes control carpel, ovule development in *Arabidopsis*. *Plant Cell* 15:2603–2611
- Férrandiz C, Gu Q, Martienssen R, Yanofsky MF (2000) Redundant regulation of meristem identity, plant architecture by *FRUITFULL APETALA1*, *CAULIFLOWER*. *Development* 127:725–734
- Flanagan CA, Ma H (1994) Spatially, temporally regulated expression of the MADS-box gene *AGL2* in wild-type, mutant *Arabidopsis* flowers. *Plant Mol Biol* 26:581–595
- Gómez-Mena C, de Folter S, Costa MM, Angenent GC, Sablowski R (2005) Transcriptional program controlled by the floral homeotic gene *AGAMOUS* during early organogenesis. *Development* 132(3):429–438
- Gu Q, Férrandiz C, Yanofsky MF, Martienssen R (1998) The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125:1509–1517
- Hartmann U, Höhmann S, Nettesheim K, Wisman E, Saedler H, Huijser P (2000) Molecular cloning of SVP: a negative regulator of the floral transition in *Arabidopsis*. *Plant J* 21:351–360
- Heuer S, Lörz H, Dresselhaus T (2000) The MADS box gene *ZmMADS2* is specifically expressed in maize pollen, during maize pollen tube growth. *Sex Plant Reprod* 13:21–27
- Heuer S, Hansen S, Bantin J, Brettschneider R, Kranz E, Lörz H, Dresselhaus T (2001) The maize MADS box gene *ZmMADS3* affects node number, spikelet development, is co-expressed with *ZmMADS1* during flower development in egg cells, early embryogenesis. *Plant Physiol* 127:33–45
- Honma T, Goto K (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409:525–529
- Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Res* 9:868–877
- Huang H, Tudor M, Weiss C, Hu Y, Ma H (1995) The *Arabidopsis* MADS-box gene *AGL3* is widely expressed, encodes a sequence-specific DNA-binding protein. *Plant Mol Biol* 28:549–567
- Ito T, Wellmer F, Yu H, Das P, Ito N, Alves-Ferreira M, Riechmann JL, Meyerowitz EM (2004) The homeotic protein *AGAMOUS* controls microsporogenesis by regulation of *SPOROCTELESS*. *Nature* 430:356–360
- Izawa T, Takahashi Y, Yano M (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. *Curr Opin Plant Biol* 6:113–120
- Kempin SA, Savidge B, Yanofsky MF (1995) Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science* 267:522–525
- Krizek BA, Fletcher JC (2005) Molecular mechanisms of flower development: an armchair guide. *Nature* 6:688–698
- Lawton-Rauh AL, Alvarez-Buylla ER, Purugganan MD (2000) Molecular evolution of flower development. *Trends Ecol Evol* 15:144–149
- Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P (2002) Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* 30:242–244
- Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404:766–770
- Litt A, Irish VF (2003) Duplication, diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165:821–833
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR, the 2^{-ΔΔCT} method. *Methods* 25:402–408
- Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* 360:273–277
- Mena M, Mandel MA, Lerner DR, Yanofsky MF, Schmidt RJ (1995) A characterization of the MADS-box gene family in maize. *Plant J* 8:845–854
- Michaels SD, Amasino RM (1999) *Flowering Locus C* encodes a novel MADS domain protein that act as a repressor of flowering. *Plant Cell* 11:949–956

- Morais H, Caramori PH, Koguish MS, Ribeiro AMA (2008) Escala fenológica detalhada da fase reprodutiva de *Coffea arabica*. *Bragantia* 67(1):257–260
- Mouradov A, Glassick TV, Hamdorf BA, Murphy LC, Marla SS, Yang Y, Teasdale R (1998) Family of MADS-box genes expressed in early male, female reproductive structures of Monterey pine. *Plant Physiol* 117:55–61
- Mouradov A, Cremer F, Coupland G (2002) Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell Supplement*, pp S11–S130
- Münster T, Deleu W, Wingen LU, Ouzunova M, Cacháron J, Faigl W, Werth S, Kim JTT, Saedler H, Theißen G (2002a) Maize MADS-box genes galore. *Maydica* 47:287–301
- Münster T, Faigl W, Saedler H, Theißen G (2002b) Evolutionary aspects of MADS-box genes in the eusporangiate fern *Ophioglossum*. *Plant Biol* 4:474–483
- Ng M, Yanofsky M (2001) Function, evolution of the plant MADS-box gene family. *Nat Rev Genet* 2:186–185
- Onouchi H, Igeno MI, Périlleux C, Graves K, Coupland G (2000) Mutagenesis of plants overexpressing *CONSTANS* demonstrates novel interactions among *Arabidopsis* flowering-times genes. *Plant Cell* 12:885–900
- Parenticová L, Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L (2003) Molecular, phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15:1538–1551
- Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* 405:200–203
- Pezzopane JRM, Pedro Junior MJ, de Camargo MBP, Fazuoli LC (2008) Heat requirements of Mundo Novo coffee for the flowering-harvest phenological stage. *Ciência e Agrotecnologia* 32:1781–1786
- Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF (2003) Assessing the redundancy of MADS-box genes during carpel, ovule development. *Nature* 424:85–88
- Ramakers C, Ruijter JM, Deprez RH, Moorman AF (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 13:62–66
- Ratcliffe OJ, Nadzan GC, Reuber TL, Riechmann JL (2001) Regulation of flowering in *Arabidopsis* by an FLC homologue. *Plant Physiol* 126(1):122–132
- Rena AB, Maestri M (1985) Fisiologia do cafeeiro. Informe Agropecuário (Belo Horizonte) 11(126):26–40
- Riechmann JL, Krizek BA, Meyerowitz EM (1996) Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1 APETALA3 PISTILLATA, AGAMOUS. *PNAS* 93:4793–4798
- Rounsley SD, Ditta GS, Yanofsky MF (1995) Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell* 7:1259–1269
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* 288:1613–1616
- Savidge B, Rounsley SD, Yanofsky MF (1995) Temporal relationships between the transcription of two *Arabidopsis* MADS box genes, the floral organ identity genes. *Plant Cell* 7:721–733
- Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART a simple modular architecture research tool: identification of signaling domains. *PNAS* 95:5857–5864
- Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES (2000) The molecular basis of vernalization: the central role of *FLOWERING LOCUS C (FLC)*. *PNAS* 97:3753–3758
- Silva LHCP, Resende MLV, Martins H Jr, Campos JR, Souza RM, Castro RM (2000) Épocas e modo de aplicação do ativador de plantas benzothiadiazole (BHT) na proteção contra a mancha bacteriana do tomateiro. *Horticultura Brasileira* (Brasília) 18:375–376
- Sitnikova T, Rzhetsky A, Nei M (1995) Interior-branch, bootstrap tests of phylogenetics trees. *Mol Biol Evol* 12:319–333
- Sommer H, Beltrán JP, Huijser P, Pape H, Lönning WE, Saedler H, Schwarz-Sommer Z (1990) Deficiens a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J* 9:605–613
- Tamura K, Dudley J, Nei M, Kumar S (2007) *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Theißen G (2001) Development of floral organ identity: stories from the MADS house. *Curr Opin Plant Biol* 4:75–85
- Theißen G, Saedler H (2001) Floral quartets. *Nature* 409:469–471
- Theißen G, Kim JT, Saedler H (1996) Classification, phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J Mol Evol* 43:484–516
- Theißen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H (2000) A short history of MADS-box genes in plants. *Plant Molecular Biology* 42:115–149
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties, weight matrix choice. *Nucleic Acids Research* 22:4673–4680
- Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönning WE, Saedler H, Sommer H, Schwarz-Sommer Z (1992) GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of *Antirrhinum* floral organogenesis. *EMBO J* 11(13):4693–4704
- Vieira LGE et al (2006) Brazilian coffee genome project: an EST-based genomic resource. *Brazilian J Plant Physiol* 18(1):95–108
- Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M (2000) Comparison of human adult, fetal expression, identification of 535 housekeeping/maintenance genes. *Physiological Genomics* 2:143–147
- Wellmer F, Riechmann JL, Alves-Ferreira M, Meyerowitz EM (2004) Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers. *Plant Cell* 16(5):1314–1326
- Winter K-U, Becker A, Münster T, Kim JT, Saedler H, Theißen G (1999) MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *PNAS* 96:7342–7347
- Winter K-U, Saedler H, Theißen G (2002a) On the origin of class B floral homeotic genes: functional substitution, dominant inhibition in *Arabidopsis* by expression of an ortholog from the gymnosperm *Gnetum*. *Plant Journal* 31:457–475
- Winter K-U, Weiser C, Kaufmann K, Bohne A, Kirchner C, Kanno A, Saedler H, Theißen G (2002b) Evolution of class B floral homeotic proteins: obligate heterodimerization originated from homodimerization. *Mol Biol Evol* 19:587–596
- Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. *Nature* 346:35–39
- Yu H, Xu Y, Tan EL, Kumar PP (2002) *AGAMOUS*-like 24 dosage-dependent mediator of the flowering signals. *PNAS* 99:16336–16341
- Zachgo S, Saedler H, Schwarz-Sommer Z (1997) Pollen-specific expression of *DEFH125* a MADS-box transcription factor in *Antirrhinum* with unusual features. *Plant Journal* 11:1043–1050
- Zhang H, Forde BG (1998) An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture. *Science* 279:407–409