Plant Cell Reports
https://doi.org/10.1007/s00299-018-2310-9

ORIGINAL ARTICLE

Identification of a seed maturation protein gene from *Coffea arabica* (*CaSMP*) and analysis of its promoter activity in tomato

Fabíola OCampo Quintero[1] · Layra G. Pinto[1] · Carla F. Barsalobres-Cavallari[1] · Mariana de Lara Campos Arcuri[1] · Lilian Ellen Pino[2] · Lázaro Eustáquio Pereira Peres[2] · Mirian P. Maluf[3] · Ivan G. Maia[1]

Received: 15 January 2018 / Accepted: 4 June 2018

Abstract

Key message **A seed maturation protein gene (*CaSMP*) from *Coffea arabica* is expressed in the endosperm of yellow/green fruits. The *CaSMP* promoter drives reporter expression in the seeds of immature tomato fruits.**

Abstract In this report, an expressed sequence tag-based approach was used to identify a seed-specific candidate gene for promoter isolation in *Coffea arabica*. The tissue-specific expression of the cognate gene (*CaSMP*), which encodes a yet uncharacterized coffee seed maturation protein, was validated by RT-qPCR. Additional expression analysis during coffee fruit development revealed higher levels of *CaSMP* transcript accumulation in the yellow/green phenological stage. Moreover, *CaSMP* was preferentially expressed in the endosperm and was down-regulated during water imbibition of the seeds. The presence of regulatory *cis*-elements known to be involved in seed- and endosperm-specific expression was observed in the *CaSMP* 5′-upstream region amplified by genome walking (GW). Additional histochemical analysis of transgenic tomato (cv. Micro-Tom) lines harboring the GW-amplified fragment (~ 1.4 kb) fused to *uidA* reporter gene confirmed promoter activity in the ovule of immature tomato fruits, while no activity was observed in the seeds of ripening fruits and in the other organs/tissues examined. These results indicate that the *CaSMP* promoter can be used to drive transgene expression in coffee beans and tomato seeds, thus representing a promising biotechnological tool.

Keywords Promoter · Seed maturation protein · Seed-specific expression · Coffee · Tomato

Introduction

Coffee is one of the most important agricultural commodities in the world, being widely cultivated in tropical countries for both internal consumption and export. Coffee trees belong to the genus *Coffea*, within the Rubiaceae family, which comprises more than 100 species. The most cultivated species in Brazil and worldwide is *C. arabica*, an autogamous allotetraploid with very narrow genetic diversity (Anthony et al. 2002; Maluf et al. 2005), a feature that impairs the transfer of desired traits to elite materials through conventional breeding.

To date, several research efforts in genetics and biotechnology have been performed aiming to obtain improved coffee plants with greater agronomical traits (reviewed in Tran et al. 2016). However, despite the molecular tools currently available, progresses in classical coffee breeding programs are slow and generally limited. An effective alternative to advance in coffee breeding is the use of genetic engineering, which has the potential to accelerate the improvement of

Communicated by Marcelo Menossi.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00299-018-2310-9) contains supplementary material, which is available to authorized users.

✉ Ivan G. Maia
igmaia@ibb.unesp.br

[1] Department of Genetics, Institute of Biosciences, Sao Paulo State University (UNESP), Botucatu, Sao Paulo 18618-689, Brazil

[2] Department of Biological Sciences (LCB), Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), University of Sao Paulo (USP), Piracicaba, Sao Paulo 13418-900, Brazil

[3] Embrapa Coffee and Coffee Center Alcides Carvalho, Agronomic Institute of Campinas, Campinas, Sao Paulo 13012-970, Brazil

Published online: 13 June 2018

elite materials. In this regard, among the tools often required to generate genetically modified coffee plants, the availability of promoters with desired expression patterns is of paramount importance (Mishra and Slater 2012).

Promoters are crucial to fine-tune the transcriptional control of the inserted transgenes. Among the promoters employed in biotechnology, tissue-specific promoters represent a valuable tool since they allow confined transgene expression (reviewed in Hernandez-Garcia and Finer 2014), and avoid the constitutive expression that is usually achieved by conventional promoters such as, for example, the 35S promoter of *Cauliflower mosaic virus* (CaMV). In coffee, most studies evaluating organ/tissue-specific promoters have focused on seed- and fruit-specific genes (Marraccini et al. 1999; Cotta et al. 2014). Other examples include the promoter regions of genes involved in stress and defense responses (Brandalise et al. 2009; Severino et al. 2012; Petitot et al. 2013; Nobres et al. 2016; Alves et al. 2017), in light-regulated carbon fixation (Marraccini et al. 2003) and in caffeine biosynthesis (Satyanarayana et al. 2005). In practice, however, the 35S promoter is still the preferred choice for the construction of expression cassettes employed in coffee biotechnology (Mishra and Slater 2012).

Despite representing a desirable strategy, the achievement of restricted transgene expression in coffee requires a larger number of promoters with appropriate features of organ/tissue specificity. In this respect, the availability of a wide range of seed-specific promoters is of special interest, since they represent ideal tools for bioengineering important traits in coffee beans such as cup quality and pest resistance. Although promoters from genes showing specific expression in seeds have been identified in *C. arabica* (Marraccini et al. 1999; Cotta et al. 2014) and *C. canephora* (Hinniger et al. 2006; Kumar et al. 2007), adding novel seed-specific promoters to the existing repertoire of coffee promoters is strategic for the advancement of coffee biotechnology. The biotechnological application of such promoters will expand our possibilities to drive transgene expression in a specific seed tissue or at a particular developmental stage of seed development.

Apart from coffee, seed-specific promoters have been cloned in model and cultivated plant species. Major examples include the promoter regions of genes encoding seed storage proteins from dicotyledonous and monocotyledonous species (reviewed in Hernandez-Garcia and Finer 2014). According to Fauteux and Strömvik (2009), the regulatory motifs implicated in seed-specific expression are generally located in the promoter-proximal region lying within 500 base pairs upstream of the transcriptional start site. Experimentally validated seed- and endosperm-specific *cis*-regulatory elements include the RY motif (CATGCA; Dickinson et al. 1988), the so-called G- (CACGTG; Ezcurra et al. 1999) and E-boxes (CANNTG; Stalberg et al. 1996), the prolamin-box (TGTAAAGT; Forde et al. 1985) and the Skn-1 motif (GTCAT; Washida et al. 1999). Mutational analyses have demonstrated that these elements are redundant and operate in a combinatorial manner to confer seed-specific expression (Chandrasekharan et al. 2003).

In this study, with the aim to increase the number of tissue-specific promoters for application in coffee biotechnology, we employed an expressed sequence tag (EST)-based approach to identify a coffee gene with fruit/seed-specific expression for promoter isolation. In a first step, the tissue-specific expression of the selected gene (referred hereafter to as *CaSMP*) that encodes a yet uncharacterized seed maturation protein (SMP) from *C. arabica* was validated using different coffee organs/tissues. Upon confirmation of its seed specificity, the spatial and temporal expression patterns driven by the *CaSMP* 5′-flanking region (~ 1.4 kb) were examined in transgenic tomato (*Solanum lycopersicum* cv. Micro-Tom) plants. We anticipate that the cloned promoter represents an additional tool for directing seed-specific expression.

Materials and methods

Plant material

The samples from five different coffee organs/tissues (root, stem, leaf, flower and pool of fruits at different stages of development) used for gene validation were harvested as described (Barsalobres-Cavallari et al. 2009). Whole fruits of *C. arabica* cv. Obatã IAC 1669-20 at different stages of development and ripening, as defined by the phenological scale proposed by Pezzopane et al. (2003), were harvested from different branches of 10 randomly selected mature coffee trees (15 years) grown under field conditions at the experimental station of Instituto Agronômico de Campinas (Campinas, SP, Brazil). Investigated fruit stages were: (1) pinhead; (2) expansion; (3) green; (4) yellow/green; (5) green/cherry and (6) cherry. Fruit tissues (pulp and endosperm) from *C. arabica* cv. Catuaí Vermelho were kindly provided by Dr. Douglas S. Domingues from UNESP—Rio Claro. These fruits were collected at 90, 180 and 240 days after flowering (DAF) from six mature trees (10 years) kept under field conditions. Coffee seeds from *C. arabica* cv. Catuaí Vermelho were provided by Dr. Edivaldo Amaral from UNESP—Botucatu. These seeds were submitted to imbibition in the dark and sampled at different time points (0, 1, 6, 12, 24, 48 and 72 h). All material collected was stored at − 80 °C until used.

All stable transformation assays were performed using tomato plants (*Solanum lycopersicum* cv. Micro-Tom) grown in a greenhouse under environmental conditions. Wild-type (WT) and a transgenic tomato cv. Micro-Tom line bearing

the DR5 synthetic auxin response promoter fused to the β-glucuronidase (GUS) reporter (Ulmasov et al. 1997) were used as negative and positive controls, respectively. Fruits were collected at different developmental stages ranging from 10 to 50 DAF.

In silico selection of the coffee EST candidate

The EST candidate (CA00-XX-FR2-061-H09-JE) used in this study was selected from the Brazilian Coffee Genome Database (http://www.lge.ibi.unicamp.br/cafe/; Vieira et al. 2006) during in silico searches for the identification of fruit-specific or fruit-abundant ESTs. The selected EST, deposited in GenBank under the accession number GW467278, was exclusively found in a *C. arabica* library (FR2) prepared from flower buds, pinhead fruits and fruits at different developmental stages (Vieira et al. 2006).

Phylogenetic analysis

A phylogenetic analysis was performed including the predicted protein encoded by the identified EST, its homolog protein from *C. canephora*, three seed maturation proteins identified in *S. lycopersicum* (Cao and Li 2015) and 51 well-characterized late embryogenesis abundant (LEA) proteins from *Arabidopsis thaliana* (Hundertmark and Hincha 2008) (listed in Table S1). The deduced amino acid (aa) sequences were aligned using the Muscle algorithm (Edgar 2004) with default parameters as implemented in molecular evolutionary genetics analysis (MEGA), version 7.0 (Kumar et al. 2016). Model testing was performed using the tool "Find best DNA/protein models (ML)" in MEGA 7.0. All available models were tested with four categories in rate variation. The best model was configured with PhyML through the SeaView user interface (Gouy et al. 2010) to obtain the maximum likelihood tree. The accuracy of branching was obtained by the approximate likelihood ratio test method. The nodes supported by a-LRT < 50% were hidden.

Expression analysis by RT-qPCR

Frozen coffee organ/tissue samples were ground in liquid nitrogen for total RNA extraction using the NucleoSpin RNA Plant kit according to the manufacturer's instructions (Macherey-Nagel). One microgram of the extracted RNA samples was treated with RNase-free DNase I (Promega) and then reverse-transcribed (RT) using the High-Capacity cDNA Reverse Transcription kit following the manufacturer's instructions (Invitrogen).

Relative expression analysis was carried out using a StepOnePlus Real-Time PCR System (Applied Biosystems). Specific primers for the amplification of the *CaSMP* gene (5′-CGTCTGTCATGCAATCTGCT-3′ and 5′-AATGTTGACGCCCTGTTCTC-3′) were designed based on the nucleotide sequence of the starting EST using Primer Express 2.0 software (Applied Biosystems). Each reaction sample contained 270 ng of cDNA, 1.0 µl of each primer (~200 ng) and 5 µl of the SYBR Green Master Mix (Applied Biosystems) to a total volume of 10 µl. The cycling conditions were as follows: 10 min at 95 °C, followed by 40 cycles each of 15 s at 95 °C, 1 min at 60 °C. Each reaction was performed in triplicate. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). A coffee polyubiquitin coding gene (5′-AAAAGGAATCCACCCTCCAC-3′ and 5′-ACCCTCCTTGTCCTGGATCT-3′) was used as endogenous control (Barsalobres-Cavallari et al. 2009). Amplification efficiencies were derived from the amplification plots using the program LinRegPCR, and a value of two was used in the calculations. The Relative Expression Software Tool (REST) software was employed to analyze the relative expression data, which was tested for significance by the Taylor test. Differences with p values < 0.05 were considered statistically significant.

Amplification by genome walking (GW) and analysis of the *CaSMP* promoter

The 5′-flanking region of the target gene was amplified using genomic DNA extracted from coffee leaves and the GenomeWalker Universal kit (Clontech) essentially as described by the manufacturer. For this, the following primers were employed in the first and second rounds of PCR: GSP1 (5′-AGAACCCTGTTTTCAGCGGCTTGCATGGC-3′) and AP1 (5′-GTAATACGACTCACTATAGGGC-3′); GSP2 (5′-CCCTTGTGGCGTTCTTGGCTGGTCTTGG-3′) and AP2 (5′-ACTATAGGGCACGCGTGGT-3′). The major GW amplification product (1400 bp) was gel-purified, cloned into pGEM-Teasy (Promega) and sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). The presence of putative *cis*-regulatory elements within the *CaSMP* promoter was determined using the Plantcare (Lescot et al 2002; http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), PLACE (http://www.dna.affrc.go.jp/PLACE/) and PlantPAN 2.0 (http://plantpan2.itps.ncku.edu.tw/) databases.

Construction of the *CaSMP* promoter expression cassette

To perform the functional analysis of the investigated promoter in stably transformed tomato plants, an expression cassette was constructed using the binary vector pBI121 (Chen et al. 2003). For this, the GW-amplified fragment (~1400 bp from the start codon), initially inserted into pGEM-Teasy, was re-amplified by PCR using primers CaSMP-*Hin*dIII (5′-CCCAAGCTTGAAAGAAGTATCTACAACTCAGAA

GTTG-3′) and CaSMP-*Bam*HI (5′-CCGGATCCCTTGAT TTCTTCTCTTTT-3′), which contain *Hin*dIII and *Bam*HI recognition sites (underlined), respectively, and then cloned into a *Hin*dIII- and *Bam*HI-digested pBI121. The resulting expression cassette containing the GUS reporter gene under the control of the *CaSMP* promoter (*CaSMP:GUS*) was verified by sequencing, and the recombinant binary vector introduced into *Agrobacterium tumefaciens* LBA4404 strain for subsequent transformation of tomato.

Transformation of tomato cv. Micro-Tom and molecular analysis

Stable transformation of tomato plants cv. Micro-Tom was carried out essentially as described previously (Pino et al. 2010). For this, cotyledons were isolated from 8-day-old tomato seedlings and used as explants. After the removal of the proximal and distal portions, the collected cotyledons were divided transversely into two parts. The resulting explants were placed on Petri dishes (20 per dish) containing solid MS medium (Murashige and Skoog 1962) supplemented with 4 ml/l of vitamin B5 (250×), 30 g/l of sucrose, 7 g/l of agar, 0.4 µM of naphthalene acetic acid (NAA) and 100 µM of acetosyringone. For transformation, *A. tumefaciens* LBA4404 was inoculated into 3 ml of liquid Luria–Bertani (LB) medium supplemented with rifampicin (50 mg/l) and kanamycin (100 mg/l). After incubation at 28 °C for 48 h (120 rpm), 500 µl of this pre-culture were transferred to 50 ml of fresh LB media containing the same antibiotics and incubated overnight at 28 °C. The bacterial suspension was centrifuged (2000*g* for 15 min) and the pellet was resuspended in 50 ml of MS medium supplemented with 30 g/l of sucrose and B5 vitamins to a final $OD_{600\,nm}$ of 0.2–0.3. Ten minutes before explants transformation, sterile acetosyringone was added to the bacterial suspension to a final concentration of 100 µM. Two drops of this *Agrobacterium* suspension were applied per explants using a micropipette. After 10 min, *Agrobacterium* was removed and the dishes were incubated in the dark for 2 days. After that, the explants were transferred to new dishes containing solid MS medium supplemented with 4 ml/l of vitamin B5 (250×), 30 g/l of sucrose, 7 g/l of agar, 1 mg/l of zeatin, 300 mg/l of Timentin and 100 mg/l of kanamycin, and kept in a growth room at 25 ± 1 °C and 16 h photoperiod. To allow root elongation, well-developed shoots (2–4 mm) were separated from the explants and transferred to flasks containing fresh MS medium supplemented with kanamycin (100 mg/l) and Timentin (300 mg/l). After 15 days, the regenerated plantlets were acclimated in a greenhouse. Transgenic plants were allowed to self-pollinate.

T-DNA integration in the regenerated and acclimated T_0 tomato lines was confirmed by PCR using genomic DNA extracted from leaves and primers complementary to the *CaSMP* promoter and to the GUS coding region (5′-GTC TGCCAGTTCAGTTCGTTGTTC-3′), respectively. Only self-pollinated T_1 progenies were employed in subsequent assays.

Histochemical analysis

Promoter activity was assessed by histochemical GUS staining essentially as described by Jefferson et al. (1987). Different organs/tissues (root, stem, leaf, flower and fruits at different stages of development) were collected from representative transgenic tomato cv. Micro-Tom lines ($n = 4$). Southern blot analysis revealed that these lines harbored single- ($n = 3$) or double-copy ($n = 1$) insertions of the transgene (data not shown).

For GUS staining, the sampled organs/tissues were incubated at 37 °C for 16 h in a reaction buffer containing 100 mM of $Na_2PO_4 \cdot H_2O$, 0.5 mM of $K_4Fe(CN)_6 \cdot 3H_2O$, 10 mM of $Na_2EDTA \cdot 2H_2O$, 0.1% of Triton X-100, 1 mM of X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronide; 50 mg/ml), with an additional supplement of 20% methanol (v/v). Thereafter, samples were immersed in ethanol 70% for 24 h for complete chlorophyll removal. After this period, the stained samples were photographed.

Results

Identification of the EST candidate and validation of its organ/tissue-specific expression

Aiming to discover novel organ/tissue-specific promoters for use in coffee biotechnology, we searched for fruit-specific gene candidates within a set of coffee ESTs with categorized organ/tissue expression profiles. As a result, an EST (CA00-XX-FR2-061-H09-JE; 664 bp) predicted to be exclusively present in a library (FR2) made from floral buds, pinhead fruits, and fruits at developmental stages 1 and 2 (Vieira et al. 2006) was identified. To validate this predicted in silico expression profile, the organ/tissue-specific expression of the selected EST candidate was examined using RT-qPCR. As can be seen in Fig. 1, the corresponding transcripts were exclusively detected in a sample composed of a pool of coffee fruits at different developmental stages, while no transcripts were detected in the other coffee organs/tissues investigated, thus confirming restricted EST expression in coffee fruit tissues. Overall, these results corroborate the corresponding gene as a good tissue-specific candidate for promoter isolation.

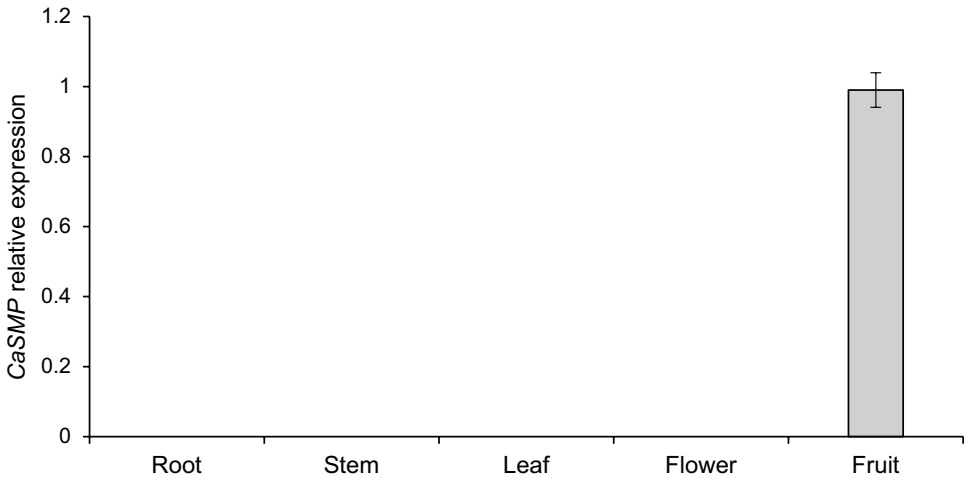

Fig. 1 Validation of *CaSMP* expression in different coffee (*C. arabica* cv. Mundo Novo) organs/tissues using RT-qPCR. The fruit sample corresponds to a pool of fruits at different developmental stages. Average data with standard errors from three replicates is presented

Phylogenetic analysis of the predicted protein encoded by the identified EST

The aforementioned EST candidate belongs to contig 2823, which contains an open reading frame (ORF) predicted to encode a 205-aa polypeptide sharing similarity to proteins belonging to the seed maturation proteins (SMP) group of the LEA protein family (Hundertmark and Hincha 2008). The so-called SMPs (also known as LEA D-34) contain the Pfam conserved domain PF04927 (Bateman et al. 2002), which is present in two copies within the putative polypeptide. BLAST searches revealed aa identity of 55% with the LEA D-34 proteins of *Cucumis melo* and *Cucumis sativus*, respectively, suggesting that the cognate gene (termed *CaSMP*) encodes a putative SMP specifically expressed in coffee fruit tissues.

Noteworthy, additional searches in the *C. canephora* genome (http://coffee-genome.org/) allowed the identification of a gene (Cc04_g16680) sharing 98% nucleotide identity with the *C. arabica* EST sequence. Likewise, the encoded product (274 aa in length) shares 97% aa sequence identity with the predicted peptide encoded by the *C. arabica* EST ORF. This orthologous gene (*CcSMP*), located on chromosome 4 of *C. canephora*, contains three exons and two introns. RNA-Seq expression data derived from different organs/tissues libraries available at the Coffee Genome Hub (http://coffee-genome.org/coffeacanephora) indicate moderate expression of this gene in stamens.

To confirm the phylogenetic relationship of CaSMP with other members of the LEA family, a phylogenetic tree was constructed using the deduced aa sequences of 51 LEA proteins from *A. thaliana* (Hundertmark and Hincha 2008) and three SMPs from tomato (Cao and Li 2015). The homologue of CaSMP from *C. canephora* was also included. The presence of nine distinct LEA groups was observed in the generated tree (Fig. 2), which is consistent with the results previously described by Hundertmark and Hincha (2008). In general, based on their Pfam domains and other features, plant LEA can be classified into eight different groups (LEA1, LEA2, LEA3, LEA4, LEA5, LEA6, dehydrin and SMP). In the generated tree, an additional group encompassing the AtM proteins, which are exclusively present in Brassicaceae species (Hundertmark and Hincha 2008), was also observed. As expected, CaSMP was included in the group that encompasses SMPs (Fig. 2), being closely related to two SMPs from Arabidopsis (At3g22490 and At3g22500) and one from tomato (SlSMP2/SlLEA21; Solyc09g082110.2). Interestingly, the corresponding Arabidopsis genes are described as displaying seed-specific expression (Illing et al. 2005).

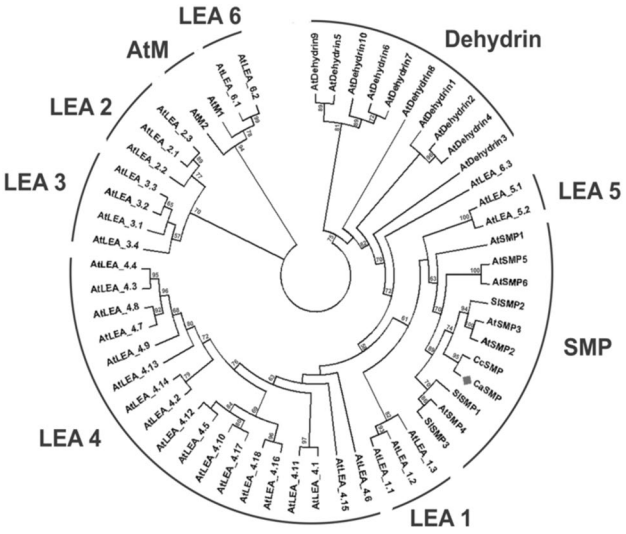

Fig. 2 Unrooted phylogenetic tree showing the relationship among CaSMP (depicted with a black diamond), its homolog from *Coffea canephora*, three seed maturation proteins from tomato and different LEA proteins identified in *A. thaliana*. Values are reported only for nodes with a-LRT > 50%

Expression analysis of *CaSMP* in coffee fruits and seeds

To gain insights into the transcriptional regulation of *CaSMP*, a detailed analysis of its spatiotemporal expression was undertaken using entire coffee fruits harvested at different developmental stages and coffee seeds at different imbibition times. The results obtained using entire coffee fruits (*C. arabica* cv. Obatã) revealed the presence of *CaSMP* transcripts in all developmental stages examined (Fig. 3). Noteworthy, the lowest level of transcript accumulation was detected during fruit expansion (stage 2), while peak accumulation was observed at stage 4 (yellow/green). A gradual decrease in transcript levels was observed thereafter [ripening stages 5 (green/cherry) and 6 (cherry), respectively], but this reduction was not statistically significant compared to the pinhead stage used as calibrator. The yellow/green phenological stage, when the fruit pericarp turns yellow, coincides with the final stage of endosperm development (Salmona et al. 2008; Joët et al. 2009).

CaSMP expression was further evaluated in the pulp and endosperm separated from coffee fruits (*C. arabica* cv. Catuaí Vermelho) harvested at different stages of development (90, 180, and 240 DAF). The results obtained revealed a low expression of *CaSMP* in the pulp of all fruit stages examined (Fig. 4). Relatively low *CaSMP* transcript levels were also found in the endosperm at 90 and 180 DAF, but a significant increase was detected at 240 DAF, reaching a maximum at this latter developmental stage (Fig. 4). It should be emphasized that Obatã and Catuaí Vermelho are late and intermediate maturation cultivars, respectively, a feature that, associated with punctual environmental effects, could explain the observed differences in gene expression timing of *CaSMP* (Gaspari-Pezzopane et al. 2012). Nevertheless, taken together, these results indicate that *CaSMP* is preferentially expressed in the endosperm.

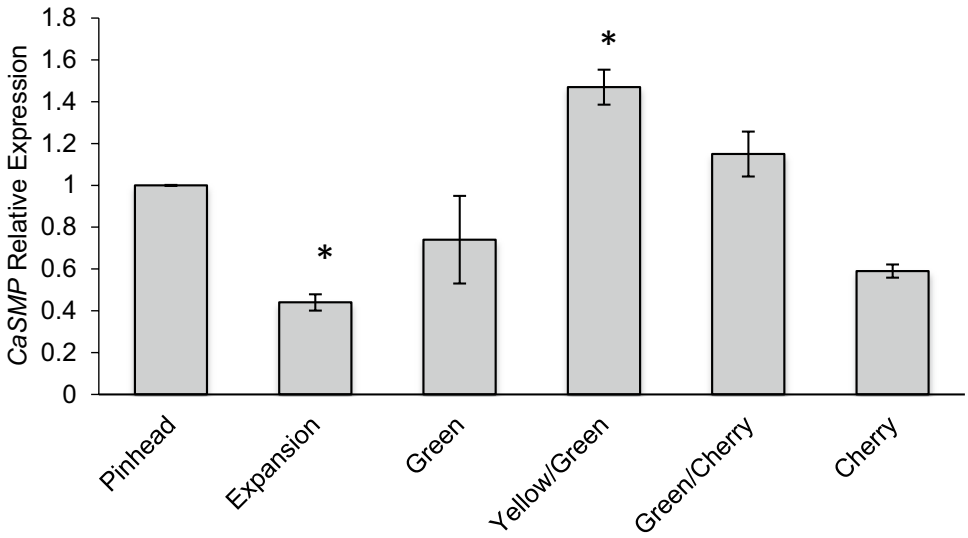

Fig. 3 Relative expression of *CaSMP* in coffee (*C. arabica* cv. Obatã) fruits at different stages of development: (1) pinhead; (2) expansion; (3) green; (4) yellow/green; (5) green/cherry and (6) cherry. The relative expression in stage 1 was arbitrarily set to 1 and used as calibrator. Average data with standard errors from three replicates is presented (*$p < 0.05$). Data analysis was performed using REST software

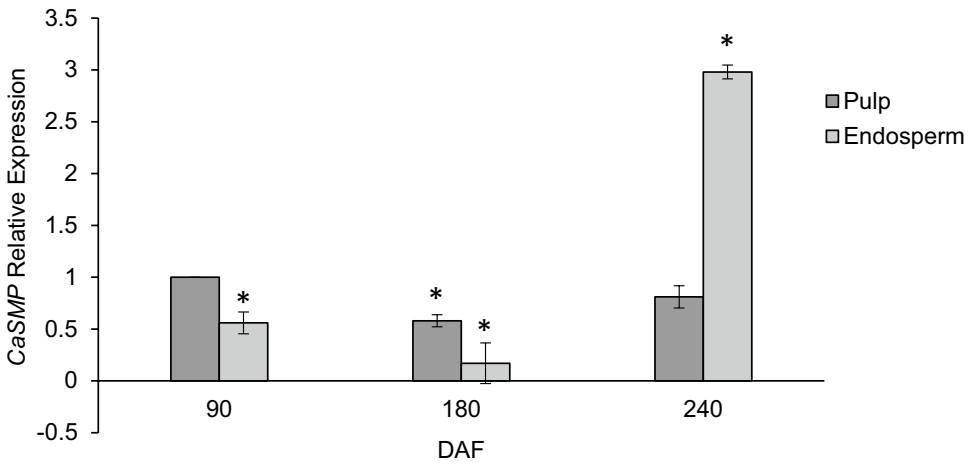

Fig. 4 Relative expression of *CaSMP* in the pulp and endosperm of coffee fruits (*C. arabica* cv. Catuaí Vermelho) harvested at different sampling times: 90, 180 and 240 DAF. The relative expression in pulp samples at 90 DAF was arbitrarily set to 1 and used as calibrator. Average data with standard errors from three replicates is presented (*$p < 0.05$). Data analysis was performed using REST software

Additional expression analysis performed using coffee seeds at different imbibition times revealed no substantial variation in *CaSMP* expression from 0 to 24 h post-imbibition (Fig. 5). On the other hand, a significant decrease in *CaSMP* transcript accumulation was observed after 48 and 72 h of imbibition, with the lowest expression level observed after 72 h. These results suggest that *CaSMP* expression is down-regulated during seed imbibition. Consistent with this, a decreased expression of two *SMP* genes from Brassica was observed during seed germination on water (Soeda et al. 2005).

Amplification and analysis of the *CaSMP* promoter region

As a subsequent step towards functional analysis of the *CaSMP* promoter, a primer-assisted GW approach was employed to amplify its 5′-flanking region. The nucleotide sequence of the GW-amplified fragment (~ 1.4 kb) was determined (Supplementary Figure S1) and then subjected to a search for *cis*-regulatory elements using specific databases. As shown in Table 1, this analysis enabled the identification of regulatory elements known to be involved in seed-specific (numerous E-box motifs) and endosperm-specific (one prolamin-box and 6 Skn-1-motifs) gene expression, which were also retrieved in the promoter region of a bean-specific dehydrin gene (*CcDH2*) from *C. canephora*. The *CaSMP* promoter region also harbored regulatory sequences required for hormone regulation, such as the abscisic acid (ABA)-responsive (ABRE) and the gibberellic acid (GA)-responsive (GARE) elements, and two motifs involved in heat stress (HSE) and low temperature (LTR) responsiveness, respectively (Table 1). Both GA and ABA have been implicated in the control of the germination process in coffee (Silva et al. 2005, 2008) and in the regulation of *LEA* genes in coffee and other species (Hinniger et al. 2006; Hundertmark and Hincha 2008; Pedrosa et al. 2015). On the

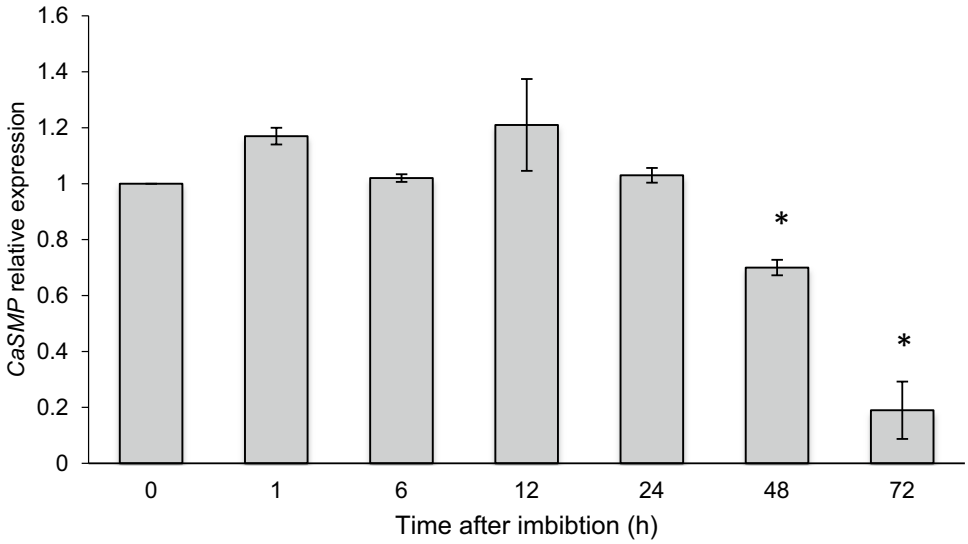

Fig. 5 Relative expression of *CaSMP* in coffee (*C. arabica* cv. Catuaí Vermelho) seeds at different imbibition times (0, 1, 6, 12, 24, 48 and 72 h). The relative expression at time 0 was arbitrarily set to 1 and used as calibrator. Average data with standard errors from three replicates is presented (*$p < 0.05$). Data analysis was performed using REST software

Table 1 Main *cis*-regulatory elements found in the *CaSMP* 5′-upstream region (1.4 kb)

Element	Sequence	Position (strand)*	Function	References
ABRE	GACATGTGGC	− 984 (+)	Abscisic acid-responsive element	Guiltinan et al. (1990)
GARE	AAACAGA	− 1300 (−)	Gibberellin-responsive element	Skriver et al. (1991)
Prolamin-box	TGTAAAG	− 546 (+)	Endosperm-specific expression	Forde et al. (1985)
E-box	CANNTG	Numerous	Seed-specific expression	Stalberg et al. (1996)
LTR-motif	CCGAAA	− 1033 (+)	Response to low temperature	Dunn et al. (1998)
Skn-1-motif	GTCAT	− 34 (+) − 188 (−) − 228 (−) − 529 (+) − 640 (+) − 1203 (+)	Endosperm-specific expression	Washida et al. (1999)
HSE	AAAAAATTTC	− 1115 (+)	Heat stress responsiveness	Pastuglia et al. (1997)

*Positions of the regulatory elements relative to the start codon (A + 1)

other hand, the HSE and LTR elements are generally retrieved in the promoter regions of *LEA* genes from other plant species (Hundertmark and Hincha 2008; Pedrosa et al. 2015). Furthermore, different putative MYB recognition binding sites and a predicted "TATA box" motif were found. These results confirm the presence within the investigated promoter region of regulatory elements potentially involved in the tissue-specific control of *CaSMP* and suggest its regulation by ABA and GA, hormones typically associated with the regulation of genes encoding LEAs (Arenas-Mena et al. 1999).

Functional analysis of the *CaSMP* promoter in tomato cv. Micro-Tom

To investigate the transcriptional and spatiotemporal activities of the *CaSMP* promoter, a *CaSMP:GUS* expression cassette was stably transformed into tomato cv. Micro-Tom using an *Agrobacterium*-mediated method (Pino et al. 2010). Of note, tomato has been considered an excellent model system to investigate promoters specifically active in reproductive organs (Sorkina et al. 2011). Analysis of GUS activity in four independent transgenic Micro-Tom lines revealed consistent GUS staining in the ovules of immature fruits (Fig. 6; immature fruits; left panel). In contrast, reporter activity was undetectable in different vegetative organs/tissues and in flowers. Likewise, GUS staining was prominent in the ovules of immature fruits from control plants carrying the *DR5:GUS* cassette (Fig. 6; right panel). In these plants, however, a low basal reporter activity was also detected in the other organs/tissues analyzed, while in flowers, GUS staining was only evident in the peduncle (Fig. 6; right panel). DR5 is a synthetic auxin-responsive promoter that drives basal GUS expression in regions associated with auxin accumulation (Ulmasov et al. 1997). No staining was observed in non-transgenic Micro-Tom plants used as negative control (data not shown).

To gain complementary information about the *CaSMP* promoter activity in fruits, GUS expression was evaluated in transgenic tomato fruits at different stages of development and ripening (from 10 to 50 DAF). As shown in Fig. 7, GUS activity was clearly evident in the ovule of immature green fruits (between 20 and 30 DAF) and undetectable in the seeds of mature red fruits. Overall, our results indicate that the *CaSMP* promoter is specifically active in the ovule of immature tomato fruits and does not promote expression in any vegetative organs/tissues.

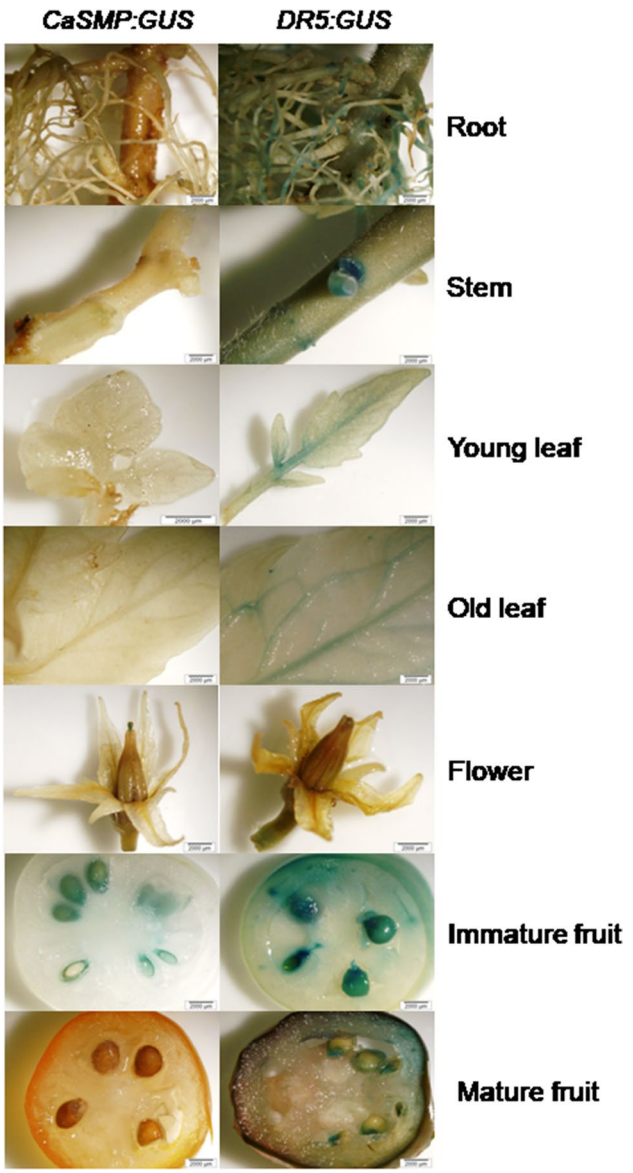

Fig. 6 Histochemical analysis of GUS activity in different organs/tissues of tomato (cv. Micro-Tom) transformed with the expression cassette *CaSMP:GUS*. The observed expression patterns were compared with those obtained in tomato (cv. Micro-Tom) plants carrying the *DR5:GUS* expression cassette. Images of a representative transgenic tomato line were used

Discussion

Seed-specific promoters represent useful tools to fine-tune the expression of transgenes in the seeds aiming at different biotechnological applications. In coffee, such promoters are particularly interesting to improve grain quality and to engineer resistance to biotic factors. Currently available coffee promoters showing seed specificity include the 11S seed storage protein (Marraccini et al. 1999) and the non-specific lipid-transfer protein (Cotta et al. 2014) gene promoters from *C. arabica* and the dehydrin gene promoter from *C. canephora* (Hinniger et al. 2006).

Here we employed an EST-based approach to identify a novel seed-specific promoter from *C. arabica*. The cognate gene (*CaSMP*) codes for a not yet characterized coffee

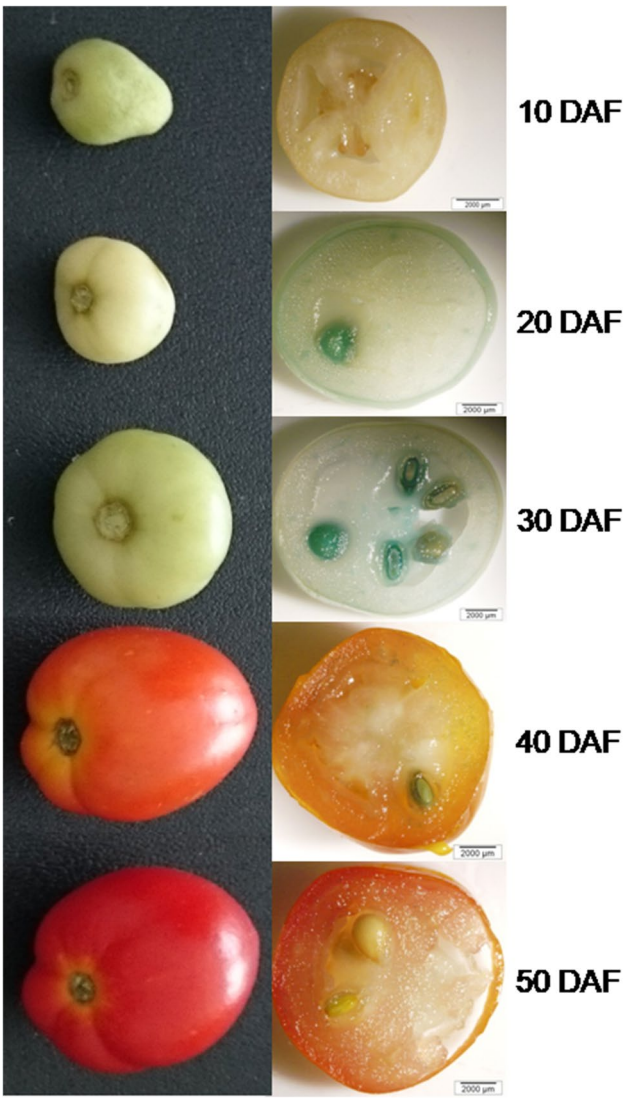

Fig. 7 Histochemical analysis of GUS activity in transgenic tomato (cv. Micro-Tom) fruits at different developmental stages recorded as DAF. Images of a representative *CaSMP:GUS* transgenic tomato line used

SMP-like protein belonging to the LEA family. SMPs, first described as LEA D-34 in cotton seeds (Baker et al. 1988), are atypical LEA proteins having an amphipathic structure and a high content of hydrophobic residues. Although poorly characterized, SMPs usually accumulate in seeds at the final stage of maturation (Chatelain et al. 2012) and have been proposed to play a role in the acquisition of desiccation tolerance and longevity (Soeda et al. 2005; Boudet et al. 2006; Chatelain et al. 2012).

Our data demonstrate that *CaSMP* was predominantly expressed during the intermediate stages of coffee fruit development and, despite a residual expression in the pulp, was significantly accumulated in the endosperm at 240 DAF. This is consistent with the presence of different conserved motifs related to endosperm-specific expression in the *CaSMP* promoter region. It should be emphasized that a delay between transcript and protein accumulation has been observed for different LEAs from *Medicago truncatula*, including a SMP, suggesting that SMP abundance during seed maturation is under post-transcriptional regulation (Verdier et al. 2013; Leprince et al. 2017). Additional expression analysis revealed that *CaSMP* transcript accumulation during seed imbibition was low, compared to the control, and was significantly down-regulated after 48 h. This is fully consistent with previous studies showing the down-regulation of *SMP* expression in imbibed seeds and during seed germination (Arenas-Mena et al. 1999; Soeda et al. 2005).

Based on gene expression clustering, Salmona et al. (2008) reported the existence of three major phenological gene clusters in developing coffee seeds that oppose the late maturation stages to the early and intermediate stages. Interestingly, *CaSMP* showed an expression pattern similar to the genes grouped in the so-called C-III3 cluster, which comprises marker genes of the developing endosperm. In this respect, SMPs are often reported as preferentially expressed in the endosperm and embryonic axes of the seed. In rice, for example, a genomic survey of endosperm-specific genes identified at least three genes encoding SMPs (Nie et al. 2013). In a proteomic study employing *M. truncatula* seeds, a high content of the SMP D-34.II isoform was observed in the seed embryonic axis (Chatelain et al. 2012). Of note, CaSMP is closely related to the Atrab28 protein from Arabidopsis (named AtSMP2 in the phylogenetic tree), which is abundant in the provascular tissues of mature embryos and in the seed coat outer tegument (Arenas-Mena et al. 1999). Taking into account that developing coffee seeds do not undergo desiccation upon maturation, and that water content at the end of endosperm development is relatively high (~ 50%) compared to orthodox seeds (Eira et al. 2006), a more general role for CaSMP in coffee seed stress tolerance could be envisaged. This hypothesis is consistent with previous studies highlighting the protective role of these proteins in seeds during stress (Boucher et al. 2010; Amara et al. 2013), but further experimentation is needed to unravel this point.

In line with its specific expression in coffee seeds, inspection of the amplified *CaSMP* 5′-upstream region revealed the presence of *cis*-regulatory elements typically found in the promoters of seed-specific genes. Among them, the E-box has been shown to function in mediating high expression levels in seeds (Chandrasekharan et al. 2003; Kim et al. 2005). In this regard, validation of the cloned promoter fragment in transgenic tomato Micro-Tom lines corroborated promoter activity in the ovules of immature fruits, while no activity was observed in ripening fruits and in the other organs/tissues analyzed. Furthermore, GUS staining was

apparently restricted to the seed coat and was not detectable in the endosperm, a feature that requires further examination. Also noteworthy is the fact that the cloned promoter was not able to completely reproduce the temporal expression pattern of the endogenous *CaSMP* gene, a feature that could be attributed to the lack of additional *cis*-regulatory elements within the cloned promoter or the absence of a *Coffea*-specific transcription factor in tomato seeds.

In summary, through its seed-specific expression pattern, the *CaSMP* promoter may be a suitable tool to direct transgene expression at a specific stage of coffee seed development, which is especially attractive for targeting pathways involved, for example, in caffeine biosynthesis and cup quality. Moreover, the successful validation of *CaSMP* promoter activity using Micro-Tom indicates that it might be potentially useful for a wide application in plant biotechnology.

Author contribution statement FOQ, MPM and IGM conceived the work and wrote the manuscript. FOQ, LGP, CFBC and MLCA performed the expression and the phylogenetic analyses. FOQ, LEP and LEPP performed the tomato transformation assays and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements This work was supported by a research Grant from FAPESP (Processo 2009/54361-9). We are very grateful to Dr. Anne-Lise Haenni for her careful and critical reading of the manuscript. Fabiola O. Quintero was recipient of a PhD fellowship from FAPESP (11/50746-3). Layra G. Pinto and Mariana L. C. Arcuri were recipients of undergraduate FAPESP fellowships (2007/59963-1 and 2014/03241-1). Ivan G. Maia, Mirian P. Maluf and Lázaro E. P. Peres are recipients of research fellowships from CNPq.

Compliance with ethical standards

Conflict of interest All authors declare they have no conflict of interest.

References

Alves GSC, Torres LF, Déchamp E, Breitler JC, Joët T, Gatineau F, Andrade AC, Bertrand B, Marraccini P, Etienne H (2017) Differential fine-tuning of gene expression regulation in coffee leaves by CcDREB1D promoter haplotypes under water deficit. J Exp Bot 68:3017–3031

Amara I, Capellades M, Ludevid MD, Pagès M, Goday A (2013) Enhanced water stress tolerance of transgenic maize plants overexpressing LEA Rab28 gene. J Plant Physiol 170:864–873

Anthony F, Combes C, Astorga C, Bertrand B, Graziosi G, Lashermes P (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theor Appl Genet 104:894–900

Arenas-Mena C, Raynal M, Borrell A, Varoquaux F, Cutanda MC, Stacy RA, Pagès M, Delseny M, Culiáñez-Macià FA (1999) Expression and cellular localization of Atrab28 during arabidopsis embryogenesis. Plant Mol Biol 40:355–363

Baker J, Van Dennsteele C, Dure L 3rd (1988) Sequence and characterization of 6 Lea proteins and their genes from cotton. Plant Mol Biol 11:277–291

Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG (2009) Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. BMC Mol Biol 10:1

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) The Pfam protein families database. Nucleic Acids Res 30:276–280

Boucher V, Buitink J, Lin X, Boudet J, Hoekstra FA, Hundertmark M, Renard D, Leprince O (2010) MtPM25 is an atypical hydrophobic late embryogenesis-abundant protein that dissociates cold and desiccation-aggregated proteins. Plant Cell Environ 33:418–430

Boudet J, Buitink J, Hoekstra FA, Rogniaux H, Larré C, Satour P, Leprince O (2006) Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. Plant Physiol 140:1418–1436

Brandalise M, Severino FE, Maluf MP, Maia IG (2009) The promoter of a gene encoding an isoflavone reductase-like protein in coffee (*Coffea arabica*) drives a stress-responsive expression in leaves. Plant Cell Rep 28:1699–1708

Cao J, Li X (2015) Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*). Planta 241:757–772

Chandrasekharan MB, Bishop KJ, Hall TC (2003) Module-specific regulation of the beta-phaseolin promoter during embryogenesis. Plant J 33:853–866

Chatelain E, Hundertmark M, Leprince O, Gall SL, Satour P, Deligny-Penninck S, Rogniaux H, Buitink J (2012) Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. Plant Cell Environ 35:1440–1455

Chen PY, Wang CK, Soong SC, To KY (2003) Complete sequence of the binary vector pBI121 and its application in cloning T-DNA insertion from transgenic plants. Mol Breed 11:287–293

Cotta MG, Barros LM, de Almeida JD, de Lamotte F, Barbosa EA, Vieira NG, Alves GS, Vinecky F, Andrade AC, Marraccini P (2014) Lipid transfer proteins in coffee: isolation of *Coffea* orthologs, *Coffea arabica* homeologs, expression during coffee fruit development and promoter analysis in transgenic tobacco plants. Plant Mol Biol 85:11–31

da Silva EA, Toorop PE, Nijsse J, Bewley JD, Hilhorst HW (2005) Exogenous gibberellins inhibit coffee (*Coffea arabica* cv. Rubi) seed germination and cause cell death in the embryo. J Exp Bot 56:1029–1038

da Silva EA, Toorop PE, van Lammeren AA, Hilhorst HW (2008) ABA inhibits embryo cell expansion and early cell division events during coffee (*Coffea arabica* 'Rubi') seed germination. Ann Bot 102:425–433

Dickinson CD, Evans RP, Nielsen NC (1988) RY repeats are conserved in the 5′-flanking regions of legume seed-protein genes. Nucleic Acids Res 16:371

Dunn MA, White AJ, Vural S, Hughes MA (1998) Identification of promoter elements in a low-temperature-responsive gene (blt4.9) from barley (*Hordeum vulgare* L.). Plant Mol Biol 38:551–564

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5:113

Eira MTS, da Silva EA, de Castro RD, Dussert S, Walters C, Bewley JD, Hilhorst HWM (2006) Coffee seed physiology. Braz J Plant Physiol 18:149–163

Ezcurra I, Ellerstrom M, Wycliffe P, Stalberg K, Rask L (1999) Interaction between composite elements in the napA promoter: both the

B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. Plant Mol Biol 40:699–709
Fauteux F, Strömvik MV (2009) Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. BMC Plant Biol 9:126
Forde BG, Heyworth A, Pywell J, Kreis M (1985) Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucleic Acids Res 13:7327–7339
Gaspari-Pezzopane C, Bonturi N, Guerreiro-Filho O, Favarin JL, Maluf MP (2012) Gene expression profile during coffee fruit development and identification of candidate markers for phenological stages. Pesqui Agropecu Bras 47:972–982
Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224
Guiltinan MJ, Marcotte WR Jr, Quatrano RS (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. Science 250:267–271
Hernandez-Garcia CM, Finer JJ (2014) Identification and validation of promoters and *cis*-acting regulatory elements. Plant Sci 217–218:109–119
Hinniger C, Caillet V, Michoux F, Ben Amor M, Tanksley S, Lin C, McCarthy J (2006) Isolation and characterization of cDNA encoding three dehydrins expressed during *Coffea canephora* (Robusta) grain development. Ann Bot 97:755–765
Hundertmark M, Hincha DK (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genom 9:118
Illing N, Denby KJ, Collett H, Shen A, Farrant JM (2005) The signature of seeds in resurrection plants: a molecular and physiological comparison of desiccation tolerance in seeds and vegetative tissues. Integr Comp Biol 45:771–787
Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6:3901–3907
Joët T, Laffargue A, Salmona J, Doulbeau S, Descroix F, Bertrand B, de Kochko A, Dussert S (2009) Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. New Phytol 182:146–162
Kim MJ, Shin JS, Kim JK, Suh MC (2005) Genomic structures and characterization of the 5′-flanking regions of acyl carrier protein and Δ4-palmitoyl-ACP desaturase genes from *Coriandrum sativum*. Biochim Biophys Acta 1730:235–244
Kumar V, Satyanarayana KV, Ramakrishna A, Chandrashekar A, Ravishankar GA (2007) Evidence for localization of *N*-methyltransferase (MMT) of caffeine biosynthetic pathway in vacuolar surface of *Coffea canephora* endosperm elucidated through localization of GUS reporter gene driven by NMT promoter. Curr Sci 93:383–389
Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
Leprince O, Pellizzaro A, Berriri S, Buitink J (2017) Late seed maturation: drying without dying. J Exp Bot 68:827–841
Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van De Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 1:325–327
Livak K, Schmittgen TD (2001) Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 25:402–408
Maluf MP, Silvestrini M, Ruggiero LMD, Guerreiro O, Colombo CA (2005) Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. Sci Agric 62:366–373
Marraccini P, Deshayes A, Pétiard V, Rogers WJ (1999) Molecular cloning of the complete 11S seed storage protein gene of *Coffea arabica* and promoter analysis in transgenic tobacco plants. Plant Physiol Biochem 37:273 – 282
Marraccini P, Courjault C, Caillet V, Lausanne F, Lepage B, Rogers WJ, Tessereau S, Deshayes A (2003) Rubisco small subunit of *Coffea arabica*: cDNA sequence, gene cloning and promoter analysis in transgenic tobacco plants. Plant Physiol Biochem 41:17–25
Mishra MK, Slater A (2012) Recent advances in the genetic transformation of coffee. Biotechnol Res Int 2012:580857
Murashige T, Skoog F (1962) A revised medium for rapid growth and bio-assays with tobacco tissue cultures. Physiol Plant 15:473–497
Nie DM, Ouyang YD, Wang X, Zhou W, Hu CG, Yao J (2013) Genome-wide analysis of endosperm-specific genes in rice. Gene 530:236–247
Nobres P, Patreze CM, Waltenberg FP, Correa MF, Tavano ECR, Mendes BMJ, Alves-Ferreira M (2016) Characterization of the promoter of the homeobox gene CaHB12 in *Coffea arabica*. Trop Plant Biol 9:50–62
Pastuglia M, Roby D, Dumas C, Cock JM (1997) Rapid induction by wounding and bacterial infection of an S gene family receptor-like kinase gene in *Brassica oleracea*. Plant Cell 9:49–60
Pedrosa AM, Martins Cde P, Gonçalves LP, Costa MG (2015) Late embryogenesis abundant (LEA) constitutes a large and diverse family of proteins involved in development and abiotic stress responses in sweet orange (*Citrus sinensis* L. Osb.). PLoS One 10:e0145785
Petitot AS, Barsalobres-Cavallari C, Ramiro D, Albuquerque Freire E, Etienne H, Fernandez D (2013) Promoter analysis of the WRKY transcription factors CaWRKY1a and CaWRKY1b homoeologous genes in coffee (*Coffea arabica*). Plant Cell Rep 32:1263–1276
Pezzopane JRM, Pedro Júnior MJ, Camargo MBP, Thomaziello RA (2003) Escala para avaliação de estádios fenológicos do cafeeiro arábica. Bragantia 62:501–505
Pino LE, Lombardi-Crestana S, Azevedo MS, Scotton DC, Borgo L, Quecini V, Figueira A, Peres LEP (2010) The *Rg*1 allele as a valuable tool for genetic transformation of the tomato Micro-Tom model system. Plant Methods 6:23
Salmona J, Dussert S, Descroix F, de Kochko A, Bertrand B, Joët T (2008) Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. Plant Mol Biol 66:105–124
Satyanarayana KV, Kumar V, Chandrashekar A, Ravishankar GA (2005) Isolation of promoter for *N*-methyltransferase gene associated with cafeine biosynthesis in *Coffea canephora*. J Biotechnol 119:20–25
Severino FE, Brandalise M, Costa CS, Wilcken SR, Maluf MP, Gonçalves W, Maia IG (2012) CaPrx, a *Coffea arabica* gene encoding a putative class III peroxidase induced by root-knot nematode infection. Plant Sci 191–192:35–42
Skriver K, Olsen FL, Rogers JC, Mundy J (1991) *Cis*-acting DNA elements responsive to gibberellin and its antagonist abscisic acid. Proc Natl Acad Sci USA 88:7266–7270
Soeda Y, Konings MC, Vorst O, van Houwelingen AM, Stoopen GM, Maliepaard CA, Kodde J, Bino RJ, Groot SP, van der Geest AH (2005) Gene expression programs during *Brassica oleracea* seed maturation, osmopriming, and germination are indicators of progression of the germination process and the stress tolerance level. Plant Physiol 137:354–368
Sorkina A, Bardosh G, Liu YZ, Fridman I, Schlizerman L, Zur N, Or E, Goldschmidt EE, Blumwald E, Sadka A (2011) Isolation of a citrus promoter specific for reproductive organs and its functional analysis in isolated juice sacs and tomato. Plant Cell Rep 30:1627–1640
Stalberg K, Ellerstom M, Ezcurra I, Ablov S, Rask L (1996) Disruption of an overlapping E-box/ABRE motif abolished high transcription

of the napA storage-protein promoter in transgenic Brassica napus seeds. Planta 199:515–519

Tran HT, Lee LS, Furtado A, Smyth H, Henry RJ (2016) Advances in genomics for the improvement of quality in coffee. J Sci Food Agric 96:3300–3312

Ulmasov T, Murfett J, Hagen G, Guilfoyle TJ (1997) Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. Plant Cell 9:1963–1971

Verdier J, Lalanne D, Pelletier S, Torres-Jerez I, Righetti K, Bandyopadhyay K, Leprince O, Chatelain E, Vu BL, Gouzy J, Gamas P, Udvardi MK, Buitink J (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. Plant Physiol 163:757–774

Vieira LGE et al (2006) Brazilian coffee genome project: an EST-based genomic resource. Braz J Plant Physiol 18:95–108

Washida H, Wu CY, Suzuki A, Yamanouchi U, Akihama T, Harada K, Takaiwa F (1999) Identification of *cis*-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. Plant Mol Biol 40:1–12