GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Identification of expressed sequences in the coffee genome potentially associated with somatic embryogenesis

A.T. Silva[1], L.V. Paiva[1,2], A.C. Andrade[3] and D. Barduche[1]

[1]Labóratorio Central de Biologia Molecular, Universidade Federal de Lavras, Lavras, MG, Brasil
[2]Departamento de Química, Universidade Federal de Lavras, Lavras, MG, Brasil
[3]Embrapa Recursos Genéticos e Biotecnologia (CENARGEN), Brasília, DF, Brasil

Corresponding author: L.V. Paiva
E-mail: luciano@dqi.ufla.br

Genet. Mol. Res. 12 (2): 1698-1709 (2013)
Received Noverber 10, 2011
Accepted February 20, 2012
Published May 21, 2013
DOI http://dx.doi.org/10.4238/2013.May.21.1

ABSTRACT. Brazil possesses the most modern and productive coffee growing farms in the world, but technological development is desired to cope with the increasing world demand. One way to increase Brazilian coffee growing productivity is wide scale production of clones with superior genotypes, which can be obtained with *in vitro* propagation technique, or from tissue culture. These procedures can generate thousands of clones. However, the methodologies for *in vitro* cultivation are genotype-dependent, which leads to an almost empirical development of specific protocols for each species. Therefore, molecular markers linked to the biochemical events of somatic embryogenesis would greatly facilitate the development of such protocols. In this context, sequences potentially involved in embryogenesis processes in the coffee plant were identified *in silico* from libraries generated by the Brazilian Coffee Genome Project. Through these *in silico* analyses, we identified 15 EST-*contigs* related to the embryogenesis process. Among these, 5

EST-*contigs* (3605, 9850, 13686, 17240, and 17265) could readily be associated with plant embryogenesis. Sequence analysis of EST-*contig* 3605, 9850, and 17265 revealed similarity to a polygalacturonase, to a cysteine-proteinase, and to an allergenine, respectively. Results also show that EST-*contig* 17265 sequences presented similarity to an expansin. Finally, analysis of EST-*contig* 17240 revealed similarity to a protein of unknown function, but it grouped in the similarity dendrogram with the *WUSCHEL* transcription factor. The data suggest that these EST-*contigs* are related to the embryogenic process and have potential as molecular markers to increase methodological efficiency in obtaining coffee plant embryogenic materials.

Key words: EST-*contigs*; Brazilian coffee genome project; *In silico*; Embryogenic; Electronic Northern

INTRODUCTION

The sequencing of expressed sequence tags (ESTs) is a method for obtaining genomic data of interest. This method has a positive cost-benefit relationship. In model plants such as *Arabidopsis thaliana* (Höfte et al., 1993) and *Oryza sativa* (Yamamoto and Sasaki, 1997), the sequencing of ESTs has contributed to the rapid identification of genes responsible for characteristics of agronomic interest, thus enabling the manipulation of these genes through biotechnological tools.

Coffee researchers have benefited from the Brazilian Coffee Genome Project (PBGC), in which approximately 33,000 unigenes have been identified starting from 214,964 ESTs obtained from 37 cDNA libraries in different physiologic stages of *Coffea arabica*, *C. canephora*, and *C. racemosa*. These ESTs have been grouped, resulting in 17,982 *contigs* and 32,155 *singlets* (Vieira et al., 2006). The large number of PBGC data generated for several species has been analyzed using bioinformatics tools. Among the various algorithms, the Fisher exact test, available at the Embrapa Genetic Resources and Biotechnology Coffee Database (Sales et al., 2007), enables *in silico* comparisons of groups of sequences formed by various libraries.

The larger objective in a genomic work is to identify genes responsible for relevant biological characteristics. Genes related to embryogenic tissues are of interest for researchers working on the *in vitro* cultivation of a species, because they are involved in responses that increase embryogenic competence. Thus, this study was conducted with the objective of identifying, through *in silico* analysis, PBGC-database sequences differentially expressed among the groups formed by libraries of embryogenic and non-embryogenic calli and cellular suspension. The identified *contigs* will be useful for the development of new and efficient molecular markers to assist in the somatic embryogenesis processes of the coffee plant.

MATERIAL AND METHODS

Finding differentially expressed sequences

To search for sequences potentially involved in the embryogenesis of the coffee plant,

we analyzed PBGC ESTs through the bioinformatics platform of the Embrapa Genetic Resources and Biotechnology, starting with comparisons among the groups formed by libraries of embryogenic and non-embryogenic calli and cellular suspension. This strategy was based on the data supplied by the Fisher exact test.

All of the ESTs (16703 in the embryogenic callus group, 8558 in the non-embryogenic callus group, and 15908 in the cellular suspension group) from 10 complementary DNA libraries (Table 1) were compared to analyze their differential expression *in silico*. The comparison combinations were: 1) embryogenic versus non-embryogenic calli, 2) embryogenic calli versus cellular suspension, and 3) cellular suspension versus non-embryogenic calli.

Table 1. Groups formed by the libraries of the Brazilian Coffee Genome Project with the respective numbers of clusters.

Library	Description	Clusters
Embryogenic callus group		
EA1	Embryogenic callus - *Coffea arabica*	3919
EC1	Embryogenic callus - *C. canephora*	4616
IA1	Embryogenic callus - *C. arabica*	72
IA2	Embryogenic lineage (leaf) with induction of 2,4D - *C. arabica*	1889
PA1	Embryogenic lineage (primary callus) - *C. arabica*	1534
Non-embryogenic callus group		
CA1	Callus - *C. arabica*	3416
IC1	Non-embryogenic lineage (leaf) without induction of 2,4D - *C. arabica*	1552
PC1	Non-embryogenic lineage (leaves) with induction of 2,4D - *C. arabica*	1422
Cellular suspension group		
CB1	Cells in suspension with bion and brassinosteroids - *C. arabica*	5409
CS1	Cells in suspension with salts - *C. arabica*	4921

2,4D = 2,4-dichlorophenoxyacetic acid.

In silico gene expression - electronic Northern

For analysis of *in silico* gene expression, the read frequencies of each EST-*contig* expressed in the libraries were normalized because the libraries were of unequal size. The normalization consisted of multiplying the frequency of each read by the ratio between the total number of reads of all the libraries and the number of reads of the library in which the read data were expressed. Using the normalization results, we processed a matrix with Cluster and TreeView (Eisen et al., 1998), in which related libraries and EST-*contig* clusters were grouped by hierarchal clustering. From the *in silico* results, we selected 15 EST-*contigs* to further study its involvement in the somatic embryogenesis of the coffee plant.

Similarity dendrogram

The 15 EST-*contig* sequences differentially expressed in the coffee plant libraries (Vieira et al., 2006) were evaluated for alignment by ClustalW (Thompson et al., 1994) and grouped with the Molecular Evolutionary Genetics Analysis 4 software (Tamura et al., 2004) using the neighbor-joining comparison model (Saitou and Nei, 1987) with the *p* distance method and pairwise suppression. The following gene sequences described as important in embryogenesis were used: *BBM* (Boutilier et al., 2002), *PSK* (Igasaki et al., 2003), *LEC* (Lo-

tan et al., 1998; Stone et al., 2001), *AGL1* (Heck et al., 1995), *PKL* (Dean Rider et al., 2003), *WUS* (Zuo et al., 2002), and *SERK* (Schmidt et al., 1997). The validity of the dendrogram related to the distance of the clusters was assessed using the bootstrap probabilistic test (Sitnikova et al., 1995). From the *in silico* results, 4 EST-*contigs* were selected. The amino acid sequences of these EST-*contigs* were deduced using the ExPASY interface (http://web.expasy.org/translate/), and the integrity of the respective domains was verified using the National Center for Biotechnology Information Conserved Domain Search program.

RESULTS AND DISCUSSION

Sequences potentially involved in coffee plant embryogenesis were identified *in silico* with the Fisher exact test, starting from the search for differentially expressed sequences in the embryogenic material libraries. From the *in silico* searches, 15 EST-*contigs* were selected with high expression in the embryogenic callus or cellular suspension libraries (Table 2).

Table 2. *Contigs* differentially expressed *in silico* in embryogenic materials of coffee plant.

Contig	Homology	No. Reads
1661	No hits found	172
1671	No hits found	204
2021	No hits found	71
3022	No hits found	19
3231	No hits found	62
3605	gb|AAC70951.1| Polygalacturonase [*Lycopersicon esculentum*]	19
3642	No hits found	58
3688	No hits found	102
3997	gb|AAO43000.1| Early tobacco anther 1 [*Nicotiana tabacum*]	48
8969	No hits found	154
9221	No hits found	86
9850	|CAA88629.1| Cysteine proteinase [*L. esculentum*]	66
13686	gb|AAX18296.1| Major allergen [*Malus domestica*]	78
17240	ref|XP_550180.1| Hypothetical protein [*Oryza sativa*]	25
17265	gb|AAT11859.2| Expansin 1 [*Mangifera indica*]	39

EST-*contigs* that displayed expression *in silico* in the embryogenic material libraries (embryogenic callus and cellular suspension) and, simultaneously, in the non-embryogenic callus libraries, were discarded from subsequent analyses (Figure 1). Although electronic Northern is a normalization of the EST-*contigs* expressed in the coffee plant libraries, low expressed sequences cannot be visualized in the graph. The program that performs the Fisher exact test provides graphs that show where the relative expressions of EST-*contigs* are observed in each library. In this context, EST-*contigs* that presented basal expressions in the non-embryogenic callus library, similar to that of EST-*contig* 7474 (Figure 2), were also discarded.

Among the 15 EST-*contigs* selected, EST-*contig* 3997 presented the domain Glo_EDI_BRP_like (Figure 3), which is structurally related metalloproteins, including dioxygenases, glyoxalase, and a group of proteins related to antibiotic resistance (Marchler-Bauer et al., 2011). These proteins are involved in stress processes, such as saline stress (Yadav et al., 2005). Because somatic embryogenesis induction is a stressful process (Nolan et al., 2006), this EST-*contig* may help protect the plant against stress.

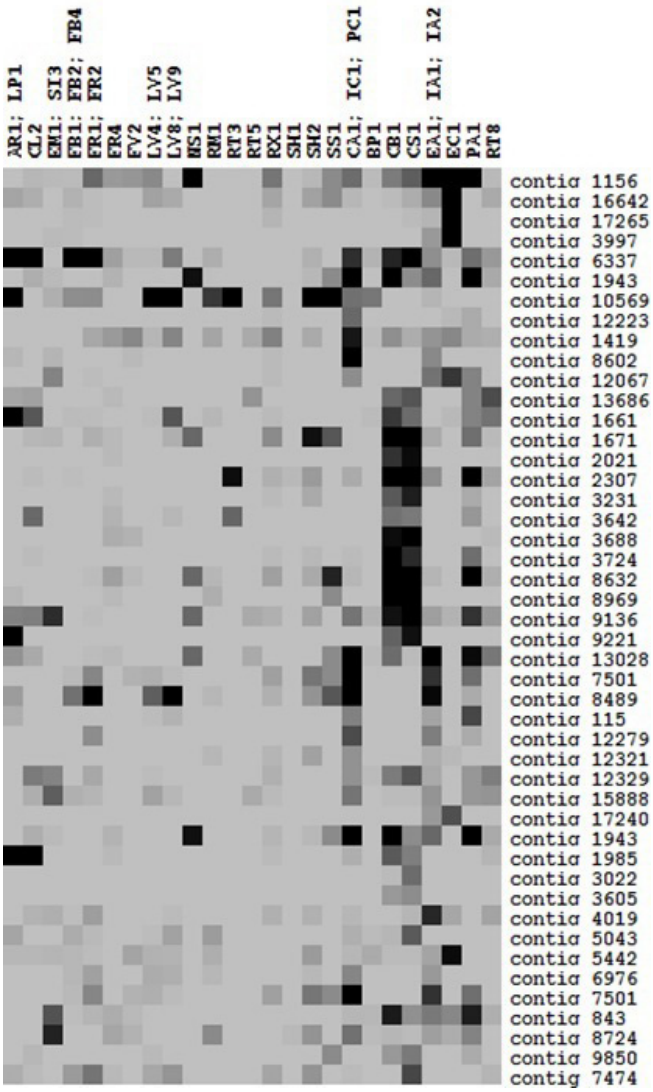

Figure 1. Electronic Northern representing EST-*contig* expression levels in the coffee libraries. The darker the gray tones, the higher the expression. **AR1**, **LP1** = leaves and plantlets with arachidonic acid treatment; **BP1** = cells in suspension treated with acibenzolar-S-methyl; **CB1** = cells in suspension treated with acibenzolar-S-methyl and brassinosteroids; **CL2** = hypocotyls treated with acibenzolar-S-methyl; **CS1** = cells in suspension treated with NaCl; **EA1**, **IA1**, **IA2** = embryogenic calus; **EC1** = embryogenic callus of *Coffea canephora*; **EM1**, **SI3** = germinating seeds (whole seeds and zygotic embryos); **FB1**, **FB2**, **FB4** = floral buds in different development stages; **FR1**, **FR2** = floral buds + fruitlets at the 1st stage + fruits at different stages; **FR4** = fruit (*C. racemosa*); **FV2** = green fruit at the 1st, 2nd, and 3rd stages of *C. racemosa*; **CA1**, **IC1**, **PC1** = non-embryogenic calus with and without 2,4-D; **LV4**, **LV5** = young leaves of orthotropic branches; **LV8**, **LV9** = mature leaves of plagiotropic branches; **PA1** = primary embryogenic callus (*C. arabica* L.); **RM1** = mature leaves infected with rust and leaf miner; **RT3** = root without bion; **RT5** = root with bion; **RT8** = root and cells in suspension in the presence of aluminum; **RX1** = branches infected with *Xylella* ssp; **SH1** = leaves of *C. canephora* under hydric stress; **SH2** = hydric stress in the field; **SS1** = tissue pool under normal conditions.

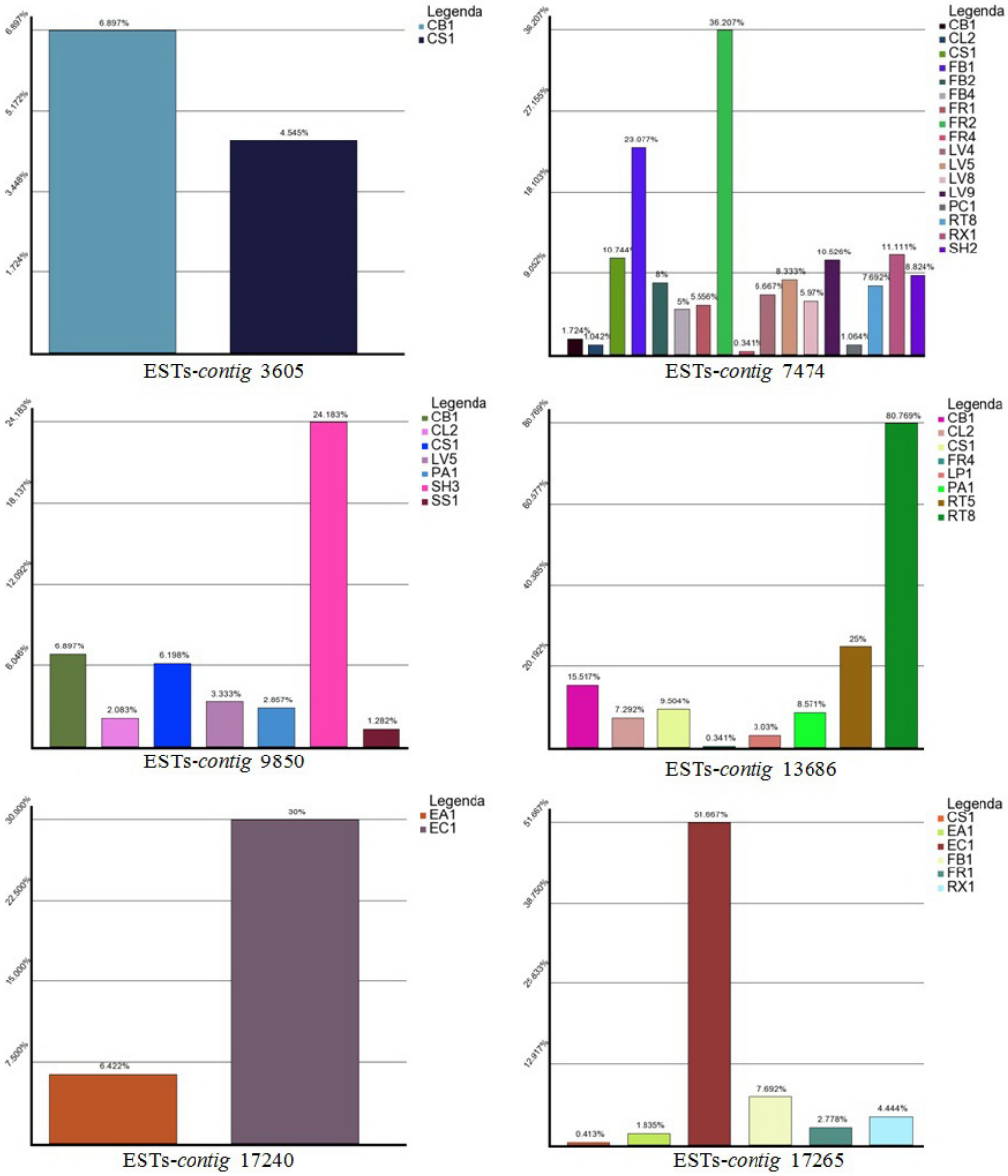

Figure 2. Relative expression in relation to the most expressed EST-*contigs* in each library. For abbreviations, see legend to Figure 1.

Of the 14 remaining EST-*contigs*, 9 were expressed *in silico* exclusively in embryogenic materials (see Figure 1) and presented no similarity with any gene present in the National Center for Biotechnology Information databases (see Table 2). These characteristics indicate that the sequences may be unique to the coffee plant or related to genes that remain

undescribed. Therefore, they are sequences that should be analyzed in more detail.

Five EST-*contigs* (3605, 9850, 13686, 17240, and 17265) were expressed *in silico* in the embryogenic material libraries (Vieira et al., 2006) (see Figure 1), displayed homology with sequences related to the embryogenic process (see Table 2), and presented domains that are potentially associated with somatic embryogenesis (see Figure 3).

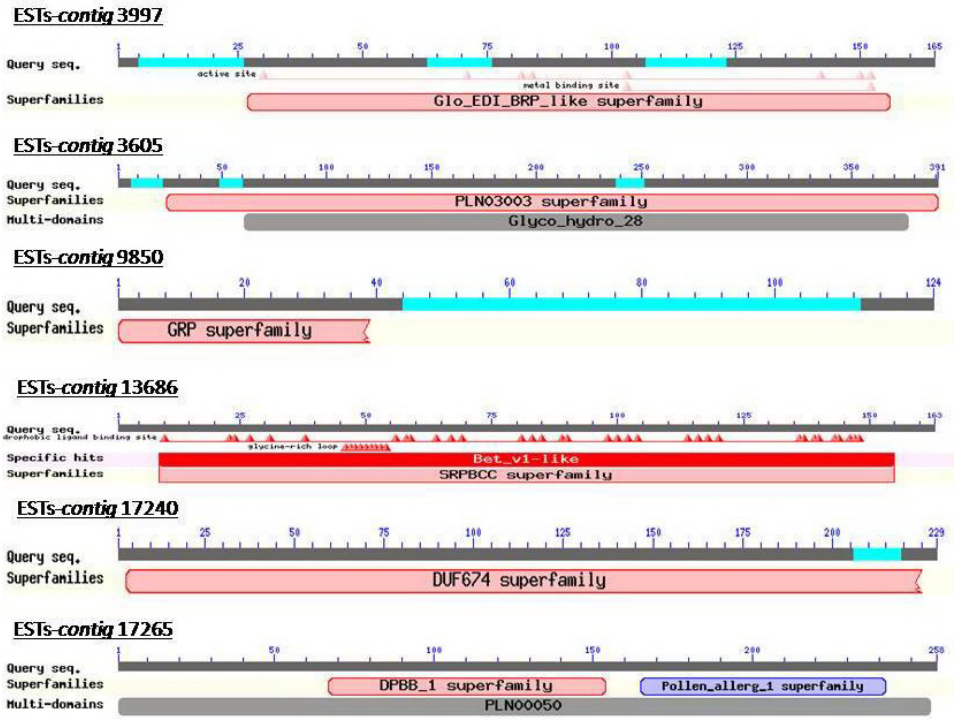

Figure 3. Dominions identified in the NCBI Program Conserved Domain Search program.

EST-*contig* 3605

Protein sequences derived from *contig* 3605 presented the Glyco_hydro_28 domain, which is structurally related to the polygalacturonase of *A. thaliana* (*At3g15720*) (Marchler-Bauer et al., 2011) (see Figure 3). Its sequence is similar to that of the polygalacturonase of *Lycopersicon esculentum* (see Table 2).

Polygalacturonase is responsible for pectin solubilization during the ripening of various fruit species (Huber, 1983), including *C. arabica* L. (Pimenta et al., 2000). In *Hordeum vulgare* L., however, the polygalacturonase gene (*HvPG1*) is expressed after microspore division during gametogenesis (Pulido et al., 2009), and in seeds of *L. esculentum* Mill., polygalacturonase (*LeXPG1*) is involved in embryonic development as well as in the relaxation of the endosperm cell walls during the protrusion of the radicle and subsequent plantlet growth (Sitrit et al., 1999).

Conversely, *contig* 3605 was not expressed *in silico* in the fruit libraries (FR1, FR2, and FR4) or germinating seeds [whole seeds and zygotic embryos (EM1, SI3)] (Vieira et

al., 2006) but only in the libraries of cell suspensions treated with acibenzolar-*S*-methyl and brassinosteroids (CB1) and cell suspensions treated with NaCl (CS1) (see Figure 2). Therefore, considering the *in silico* expression of *contig* 3605 and the expression of polygalacturonase during embryonic development in tomato (Sitrit et al., 1999) and during gametogenesis in *H. vulgare* L. (Pulido et al., 2009), *contig* 3605 may be related to coffee plant somatic embryogenesis because the process is similar to that of the zygote (Zimmerman, 1993).

EST-*contig* 9850

Analysis of protein sequences from *contig* 9850 showed that it contains part of the glycine-rich protein domain (see Figure 3) of proteins induced in response to stress (Marchler-Bauer et al., 2011) and it is similar to that of the cysteine peptidase of *L. esculentum* (see Table 2). Cysteine peptidases participate directly in the processing of seed globulin reserves, mainly in the dicotyledons, in the accumulation of these reserves during the late stages of embryogenesis, and in their mobilization during germination (Fischer et al., 2000). *Contig* 9850 was expressed *in silico* in the tissue pool libraries under normal field conditions (SS1), in calli (CL2 and PA1), and in cellular suspensions (CB1 and CS1; see Figure 2). In *Medicago truncatula*, the transient expression of β-glucuronidase by the cysteine-peptidase promoter *PsCys15a* was detected during organogenesis in embryonic calli (Vincent et al., 2000), and in *Elaeis guineensis* Jacq., cysteine peptidases were detected in somatic embryos (Aberlenc-Bertossi et al., 2008). Therefore, *contig* 9850 may be related to somatic embryogenesis in the coffee plant.

EST-*contig* 13686

The protein sequence of *contig* 13686 presented the major allergen domain Bet_v1_ like (see Figure 3) and showed similarity with the major allergen sequence of *Malus domestica* (see Table 2). This domain includes proteins related to plant pathogenesis (PR10) (Marchler-Bauer et al., 2011). These proteins were initially thought to be induced by different biotic and abiotic stresses, but later their constitutive expression was detected in various plant organs during growth (Van Loon et al., 2006).

Many PR10 proteins are constitutively expressed in various development stages and in seeds, roots (Sikorski et al., 1999), flowers, and leaves (Sikorski et al., 1999). Conversely, in *Arachis hypogaea*, 3 genes of allergenine (*ara h1*, *ara h2*, and *ara h3*) are exclusively expressed during seed development (Kang et al., 2007), and 2 homologs of the major allergen in *Dacus carrota* were highly expressed in the initial stages of embryogenesis and cellular suspension (Sano et al., 2004). Therefore, considering that *contig* 13686 was expressed *in silic*o in a non-constitutive form, in lower intensities, in the libraries of leaves, plantlets, hypocotyls, and roots (AR1, LP1, CL2, FR4, RT5, and RT8) and in higher intensities in the embryogenic callus and cellular suspension libraries (CB1, CS1, PA1, and RT8), this *contig* may be involved in coffee plant embryogenesis.

EST-*contig* 17240

Contig 17240 was composed of reads expressed only in the embryogenic callus libraries of *C. canephora* (EC1) and *C. arabica* (EA1, IA1, IA2) (Vieira et al., 2006) (see

Figure 2) and displayed homology with a protein of *O. sativa* of unknown function with the domain DUF674 (see Table 2 and Figure 3). It is grouped with the *WUSCHEL* (*WUS*) sequence [bootstrap value <50% (Figure 4)] (Laux et al., 1996).

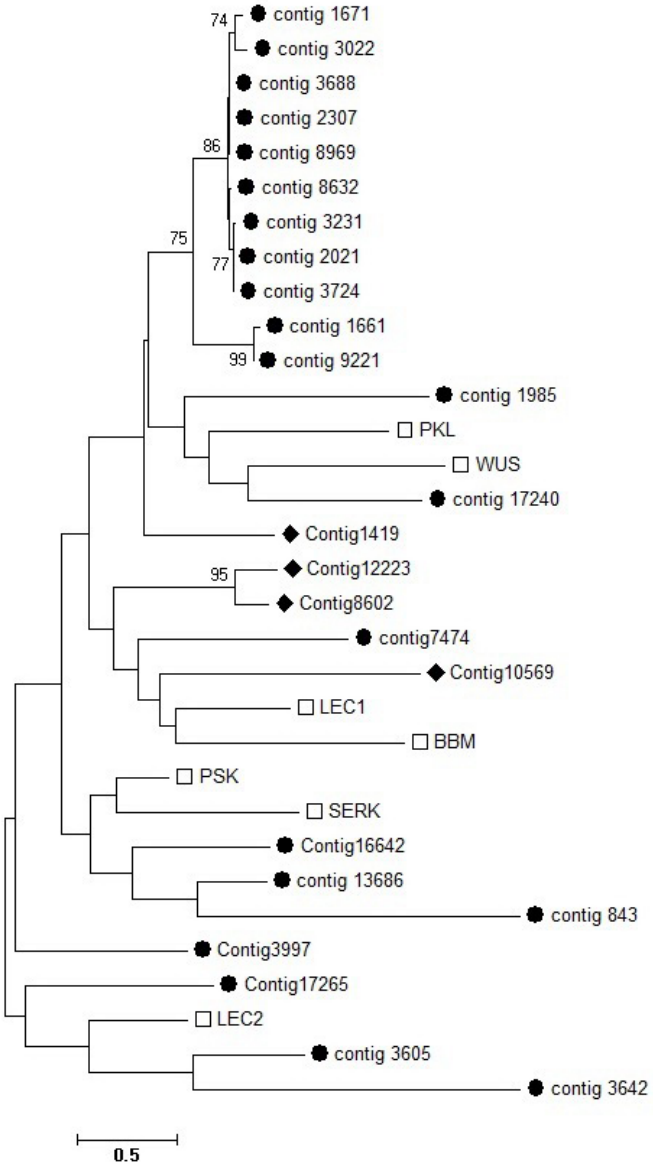

Figure 4. Similarity dendrogram relating EST-*contig* nucleotide sequences. Circles = most expressed EST-*contig* sequences in embryogenic materials; lozenges = most expressed EST-*contig* sequences in non-embryogenic callus (Vieira et al., 2006); squares = sequences involved in the embryogenic process; bootstrap values lower than 50% were omitted.

WUS is a transcription factor of the family *WUS*-related homeobox that is expressed during the embryogenesis (Haecker et al., 2004) specifically, in the 2 central cells of the 16-cell embryo during the vegetative reproduction (Laux et al., 1996). In *C. canephora*, a heterologous of *WUS* was capable of promoting the transition from the vegetative to the embryogenic state in somatic cells and, eventually, the formation of somatic embryos (Arroyo-Herrera et al., 2008).

Contig 17240 presented an *in silico* expression approximately 480 times higher in the embryogenic EC1. Therefore, because it did not present expression *in silico* in the meristematic growth libraries (see Figure 2), *contig* 17240 is most likely related to coffee plant embryogenesis; in *A. thaliana*, *WUS* expression is delimited to the regions that create the new meristems (Gordon et al., 2007). In transgenic *C. canephora* plants, however, the expression of the *WUS* heterologous gene increased the induction of somatic embryos by 400% (Arroyo-Herrera et al., 2008), a result similar to that found in the *in silico* analyses of our study.

EST-*contig* 17265

Encoding protein sequence of *contig* 17265 presented the domain of expansin (PLN00050; see Figure 3) and was similar to the *Mangifera indica* expansin sequence (see Table 2). Expansins promote the relaxation of the cellular wall at the onset of ripening of various fruit species (Harrison et al., 2001). In *C. arabica*, 2 expansin genes (*CaEXPA1* and *CaEXP3*) are directly related to the size of the fruit (Budzinski et al., 2011). In *L. esculentum*, however, expansin genes (*LeEXP4*, *LeEXP8*, and *LeEXP10*) are involved in embryonic development and germination (Chen and Bradford, 2000).

Therefore, considering that *contig* 17265 was expressed approximately 57 times higher *in silico* in the EC1 than in the fruit library, it may be more involved in coffee plant embryogenesis than in the increase of fruit size, as described in *C. canephora* (Budzinski et al., 2011). Although the EST-*contigs* analyzed indicate involvement in the somatic embryogenesis of the coffee plant, functional genomics analyses are necessary to evaluate whether these genes can be used as markers of embryogenic acquisition processes in coffee.

ACKNOWLEDGMENTS

Research supported by Fundação de Amparo à Pesquisa de Minas Gerais, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, and Conselho Nacional de Desenvolvimento Científico e Tecnológico.

REFERENCES

Aberlenc-Bertossi F, Chabrillange N, Duval Y and Tregear J (2008). Contrasting globulin and cysteine proteinase gene expression patterns reveal fundamental developmental differences between zygotic and somatic embryos of oil palm. *Tree Physiol.* 28: 1157-1167.

Arroyo-Herrera A, Ku Gonzalez A, Canche Moo R, Quiroz-Figueroa FR, et al. (2008). Expression of *WUSCHEL* in *Coffea canephora* causes ectopic morphogenesis and increases somatic embryogenesis. *Plant Cell Tissue Organ Culture* 94: 171-180.

Boutilier K, Offringa R, Sharma VK, Kieft H, et al. (2002). Ectopic expression of *BABY BOOM* triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14: 1737-1749.

Budzinski IG, Santos TB, Sera T, Pot D, et al. (2011). Expression patterns of three alpha-expansin isoforms in *Coffea arabica* during fruit development. *Plant Biol.* 13: 462-471.

Chen F and Bradford KJ (2000). Expression of an expansin is associated with endosperm weakening during tomato seed germination. *Plant Physiol.* 124: 1265-1274.

Dean Rider S Jr, Henderson JT, Jerome RE, Edenberg HJ, et al. (2003). Coordinate repression of regulators of embryonic identity by *PICKLE* during germination in *Arabidopsis*. *Plant J.* 35: 33-43.

Eisen MB, Spellman PT, Brown PO and Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95: 14863-14868.

Fischer J, Becker C, Hillmer S, Horstmann C, et al. (2000). The families of papain- and legumain-like cysteine proteinases from embryonic axes and cotyledons of *Vicia* seeds: developmental patterns, intracellular localization and functions in globulin proteolysis. *Plant Mol. Biol.* 43: 83-101.

Gordon SP, Heisler MG, Reddy GV, Ohno C, et al. (2007). Pattern formation during de novo assembly of the *Arabidopsis* shoot meristem. *Development* 134: 3539-3548.

Haecker A, Gross-Hardt R, Geiges B, Sarkar A, et al. (2004). Expression dynamics of *WOX* genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Development* 131: 657-668.

Harrison EP, McQueen-Mason SJ and Manning K (2001). Expression of six expansin genes in relation to extension activity in developing strawberry fruit. *J. Exp. Bot.* 52: 1437-1446.

Heck GR, Perry SE, Nichols KW and Fernandez DE (1995). AGL15, a MADS domain protein expressed in developing embryos. *Plant Cell* 7: 1271-1282.

Höfte H, Desprez T, Amselem J, Chiapello H, et al. (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* 4: 1051-1061.

Huber DJ (1983). Polyuronide degradation and hemicellulose modifications in ripening tomato fruit. *J. Am. Soc. Hort. Sci.* 108: 405-409.

Igasaki T, Sato T, Akashi N, Mohri T, et al. (2003). Somatic embryogenesis and plant regeneration from immature zygotic embryos of *Cryptomeria japonica* D. Don. *Plant Cell Rep.* 22: 239-243.

Kang IH, Srivastava P, Ozias-Akins P and Gallo M (2007). Temporal and spatial expression of the major allergens in developing and germinating peanut seed. *Plant Physiol.* 144: 836-845.

Laux T, Mayer KF, Berger J and Jurgens G (1996). The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. *Development* 122: 87-96.

Lotan T, Ohto M, Yee KM, West MA, et al. (1998). *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93: 1195-1205.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, et al. (2011). CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39: D225-D229.

Nolan KE, Saeed NA and Rose RJ (2006). The stress kinase gene *MtSK1* in *Medicago truncatula* with particular reference to somatic embryogenesis. *Plant Cell Rep.* 25: 711-722.

Pimenta CJ, Chagas SJR and Costa L (2000). Pectinas e enzimas pectinolíticas em café (*Coffea arabica* L.) colhido em quatro estádios de maturação. *Cienc. Agrotec.* 24: 1079-1083.

Pulido A, Bakos F, Devic M, Barnabas B, et al. (2009). HvPG1 and ECA1: two genes activated transcriptionally in the transition of barley microspores from the gametophytic to the embryogenic pathway. *Plant Cell Rep.* 28: 551-559.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.

Sales RMOB, Barbosa AV and da Silva FR (2007). Coffee Genome Project Database at Embrapa Genetic Resources and Biotechnology (CENARGEN). V Simpósio de Pesquisa dos Cafés do Brasil, Águas de Lindóia.

Sano T, Nishimoto M, Saburi W, Kimura A, et al. (2004). Isolation and characterization of cDNA encoding P-19.5 protein accumulated preferentially at early stage of carrot somatic embryogenesis. *Plant Sci.* 167: 1211-1217.

Schmidt ED, Guzzo F, Toonen MA and de Vries SC (1997). A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development* 124: 2049-2062.

Sikorski MM, Biesiadka J, Kasperska AE, Kopciñska J, et al. (1999). Expression of genes encoding PR10 class pathogenesis-related proteins is inhibited in yellow lupine root nodules. *Plant Sci.* 149: 125-137.

Sitnikova T, Rzhetsky A and Nei M (1995). Interior-branch and bootstrap tests of phylogenetic trees. *Mol. Biol. Evol.* 12: 319-333.

Sitrit Y, Hadfield KA, Bennett AB, Bradford KJ, et al. (1999). Expression of a polygalacturonase associated with tomato seed germination. *Plant Physiol.* 121: 419-428.

Stone SL, Kwong LW, Yee KM, Pelletier J, et al. (2001). *LEAFY COTYLEDON2* encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci. U. S. A.* 98: 11806-11811.

Tamura K, Nei M and Kumar S (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101: 11030-11035.

Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple

sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

Van Loon LC, Rep M and Pieterse CM (2006). Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* 44: 135-162.

Vieira LGE, Andrade AC, Colombo CA, Moraes AHA, et al. (2006). Brazilian coffee genome project: An est-based genomic resource. *Braz. J. Plant Physiol.* 18: 95-108.

Vincent JL, Knox MR, Ellis TH, Kalo P, et al. (2000). Nodule-expressed *Cyp15a* cysteine protease genes map to syntenic genome regions in *Pisum* and *Medicago* spp. *Mol. Plant Microbe Interact.* 13: 715-723.

Yadav SK, Singla-Pareek SL, Ray M, Reddy MK, et al. (2005). Methylglyoxal levels in plants under salinity stress are dependent on glyoxalase I and glutathione. *Biochem. Biophys. Res. Commun.* 337: 61-67.

Yamamoto K and Sasaki T (1997). Large-scale EST sequencing in rice. *Plant Mol. Biol.* 35: 135-144.

Zimmerman JL (1993). Somatic Embryogenesis: A Model for Early Development in Higher Plants. *Plant Cell* 5: 1411-1423.

Zuo J, Niu QW, Frugis G and Chua NH (2002). The *WUSCHEL* gene promotes vegetative-to-embryonic transition in *Arabidopsis*. *Plant J.* 30: 349-359.

