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

In silico characterization of putative members of the coffee (*Coffea arabica*) ethylene signaling pathway

A.A. Lima[1], S.A. Ságio[1], A. Chalfun-Júnior[1] and L.V. Paiva[2]

[1]Departamento de Biologia, Universidade Federal de Lavras, Lavras, MG, Brasil
[2]Departamento de Química, Universidade Federal de Lavras, Lavras, MG, Brasil

Correspondence authors: A. Chalfun-Júnior
E-mail: chalfunjunior@dbi.ufla.br

Genet. Mol. Res. 10 (2): 1277-1289 (2011)
Received February 7, 2011
Accepted March 16, 2011
Published June 28, 2011
DOI 10.4238/vol10-2gmr1314

ABSTRACT. The plant hormone ethylene is involved in several developmental and physiological processes in plants, including senescence, fruit ripening and organ abscission, as well as in biotic and abiotic stress responses. Initiation of these processes involves complex regulation of both ethylene biosynthesis and the ability of cells to perceive the hormone and respond in an appropriate manner, a process which is regulated both spatially and temporally. Ethylene is a gaseous hormone whose sensitivity is a key factor to limiting its response in target cells. We made a search of the Coffee Expressed Sequence Tag (CAFEST) database for expressed sequence tags related to known elements of the ethylene signaling pathway. Sequences showing a reliable similarity were clusterized, annotated and analyzed for conserved domains. Multiple alignments comprising the sequences that we found and sequences of ethylene signaling elements from other species were made, and their phylogeny was assessed by phylogenetic trees constructed with the MEGA4 software. The expression profile was assessed by *in silico* Northern blot analysis performed using the Cluster and TreeView programs. The CAFEST database was found to have a

large number of sequences related to previously described ethylene signaling pathway elements, allowing identification of putative members from almost every step of this pathway. The phylogenetic trees demonstrated high similarity between the sequences found in the CAFEST and those from other species, and the electronic Northern blot analysis detected their expression in various tissues, development stages and stress conditions.

Key words: Sensitivity; Ethylene receptors; *CTR1*; *EIN2*; *EIN3*; *ERF*

INTRODUCTION

The plant hormone ethylene plays an important role in various diverse physiological and developmental processes including organ senescence, seed germination, stem elongation, fruit ripening, as well as biotic and abiotic stress responses. Its role in these agronomically important processes has made ethylene a target of intense studies related to its action and regulation.

A good understanding of a hormone action is achieved based on the comprehension of the main factors that control its actions, which include concentration, regulated by biosynthesis, degradation and conjugation, localization, influenced by transport, and sensitivity (or responsiveness), which involves the presence of receptors and a signaling transduction pathway (Davies, 2004). Once produced, being a gaseous hormone, ethylene easily defuses through the intercellular spaces and adjacent tissues. Without the possibility of having a transport regulation mechanism, the controlling of its sensitivity is a key factor in limiting its responses at target cells (Alonso and Ecker, 2001).

The morphological changes displayed by dark-grown seedlings under ethylene effect, the so-called triple response phenotype, along with genetic and molecular analyses, allowed the identification of many key components of the ethylene signaling pathway, uncovering a linear framework leading from ethylene perception to transcription regulation. Ethylene is perceived by a family of five membrane bound receptors, ETHYLENE RECEPTOR1 (ETR1), ETR2, ETHYLENE SENSOR1 (ERS1), ERS2 and ETHYLENE INSENSITIVE4 (EIN4), that along with CONSTITUTIVE TRIPLE RESPONSE (CTR1), a Raf protein kinase, act as negative regulators. Downstream CTR1, the integral membrane protein ETHYLENE INSENSITIVE2 (EIN2), the transcriptional factors ETHYLENE INSENSITIVE3 (EIN3) and ETHYLENE RESPONSE FACTORS (ERFs) have been identified as positive regulators since their loss of function or overexpression can lead to ethylene insensitivity and a constitutive ethylene response in air, respectively.

Coffee quality, among other factors, has been associated with fruit ripening stage at harvest time, which is often highly asynchronous due to the sequential flowering found in this species, and usually leads to higher production costs and also a lower cup quality (Farnezi et al., 2010). Some recent studies have suggested that coffee may constitute a climateric fruit (Pereira et al., 2005; Salmona et al., 2008), indicating that ethylene may play an important role in coffee fruit ripening.

In this study, we have investigated the Coffee Expressed Sequence Tag (CAFEST)

genome database, employing bioinformatics tools and *in silico* expression analyses, to characterize putative components of coffee (*Coffea arabica*) ethylene signaling pathway.

MATERIAL AND METHODS

Database searches and alignments

In order to identify homologs of functionally characterized genes involved in the basic ethylene signaling pathway (*ETR1*, *ETR2*, *EIN4*, *ERS1*, *ERS2*→*CTR1*→*EIN2*→*EIN3*→*ERF*), data mining in the CAFEST database (http://bioinfo04.ibi.unicamp.br), composed by 214,964 expressed sequence tags (EST) obtained from 37 libraries (Vieira et al., 2006), were carried out using plant gene (BLASTn) and protein (tBLASTn) sequences as bait, as well as key word searches. The sequences with significant similarity (e-value $>10^{-4}$) were selected and sent to the sequence manager and manipulation system, the *GeneProject*, and submitted to clustering by using the CAP3 program (Huang and Madan, 1999), which is integrated into the system, forming the EST-contigs and EST-singlets.

The *C. arabica* EST-contigs (CaC) and EST-singlets (CaS), where Ca stands for *C. arabica*, obtained were manually annotated and data validation was performed by local tBLASTx and tBLASTn searches of the retrieved sequences against the GenBank® database. Then, the selected sequences were used as bait in another search against the CAFEST database, aimed at finding new reads, as well as to remount incomplete clusters. This process was repeated until no more new significant reads were found. The ORFs (open reading frames) of the validated sequences were obtained through the ORFinder tool, from the NCBI homepage (http://www.ncbi.nlm.nih.gov), and their protein sequences were generated through the translate tool found in the ExPASY (http://www.expasy.ch) protein database. The protein sequence alignments were performed by the ClustalW program (Thompson et al., 1994), using default parameters.

Phylogenetic analysis

The putative function of the deduced amino acid sequences of coffee transcripts, compared to homologs from other species, was assessed by phylogenetic trees performed by the MEGA software, version 4.0 (Tamura et al., 2007), with neighbor-joining comparison model (Saitou and Nei, 1987), p-distance method and pairwise suppression. Bootstrap values from 1000 replicates were used to assess the robustness of the trees.

In silico gene expression analysis

In silico qualitative gene expression profiling was performed using virtual Northern blot analyses of the coffee EST database. For each EST-contig and EST-singlet, frequencies of reads that form each EST-contig and EST-singlet 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 libraries were considered in this study.

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 results were plotted in a matrix and gene expression patterns among ESTs and libraries were obtained by hierarchical clustering, performed by the Cluster v.2.11 program (Eisen et al., 1998). Graphic outputs were generated by the TreeView v.1.6 software (Eisen et al., 1998) and presented in a color scale from black to red, where closer to red color the higher the expression level. No expression was represented by gray color.

RESULTS

Analyses of the CAFEST database revealed 24 reads related to ethylene receptors, clusterized into 3 contigs and 4 singlets. The motif analyses showed that only one contig (CaC20), showed a conserved ethylene binding domain, and was selected for further analyses.

As shown in the phylogenetic tree (Figure 1) and in Table 1, CaC20 showed a greater similarity with receptors from ETR1-like subfamily with an amino acid identity ranging from 70% (AtERS1) to 96% (CcETR1), while compared with ETR2-like subfamily it was below 40%. The *in silico* Northern showed that CaC20 is expressed in six different libraries, in a no tissue specific manner, and in conditions which ethylene displays important functions, such as abiotic stresses caused by an aluminum and water deficit (Figure 2).

Table 1. Comparison of the coffee sequences related to ethylene signaling pathway found in the CAFEST database and their best hit of the tBlastx analysis at the NCBI database.

Category	Contig/aa	tBlastx	E-value	Identity	Positives
ETR-like	CaC20/615	ABL63474.1 ethylene receptor isoform 1 [*Coffea canephora*]. 740aa	0.0	594/614 (96%)	601/614 (97%)
EIN2	CaC3/483	ABD65477.1 ethylene signaling protein [*Solanum lycopersicum*]. 1316aa	$6e^{-158}$	288/474 (60%)	361/474 (76%)
	CaC6/243	AAR08678.1 EIN2 [*Petunia x hybrida*]. 1310aa	$2e^{-106}$	185/242 (76%)	211/242 (87%)
EIN3	CaC4/644	AAP03998.1 EIL2 [*Nicotiana tabacum*]. 616aa	0.0	493/636 (77%)	557/636 (87%)
ERF	CaC2/264	Q40479.1 ERF2 [*Nicotiana tabacum*]. 233aa	$3e^{-88}$	175/264 (66%)	199/264 (75%)
	CaC5/304	XP_002313851.1 AP2/ERF domain-containing transcription factor [*Populus trichocarpa*]. 266aa	$3e^{-49}$	128/290 (44%)	172/290 (59%)
	CaC7/249	ERF-like transcription factor [*Coffea canephora*]. 329aa	$2e^{-102}$	175/178 (98%)	177/178 (99%)
	CaC8/218	AP2/EREBP transcription factor ERF-2 [*Gossypium hirsutum*]. 255aa	$8e^{-43}$	109/225 (48%)	135/225 (60%)
	CaC10/193	O04681.1 pathogenesis-related genes transcriptional activator PTI5 [*Solanum lycopersicum*]. 161aa	$2e^{-49}$	108/162 (66%)	119/162 (73%)
	CaC11/404	AAP40022.1 callus-expressing factor [*Nicotiana tabacum*]. 387aa	$2e^{-132}$	257/407 (63%)	304/407 (74%)
	CaC13/387	AAP40022.1 callus-expressing factor [*Nicotiana tabacum*]. 387aa	$2e^{-112}$	225/392 (57%)	276/392 (70%)
	CaC18/222	Q9LW49.1 ethylene-responsive transcription factor 4 [*Nicotiana sylvestris*]. 227aa	$7e^{-55}$	139/237 (58%)	160/237 (67%)
	CaC22/158	Q9SXS8.1 ethylene-responsive transcription factor 3 [*Nicotiana tabacum*]. 225aa	$7e^{-35}$	113/157 (71%)	120/157 (76%)
	CaC23/165	XP_002300797.1 AP2/ERF domain-containing transcription factor [*Populus trichocarpa*]. 179aa	$4e^{-33}$	81/138 (58%)	94/138 (68%)
	CaC32/282	CAN66025.1 hypothetical protein [*Vitis vinifera*]. 421aa	$6e^{-46}$	125/190 (65%)	144/190 (75%)
	CaC33/252	BAF75651.1 transcription factor DcERF1 [*Daucus carota]*. 296aa	$2e^{-50}$	120/246 (48%)	158/246 (64%)
	CaC37/226	BAF75651.1 transcription factor DcERF1 [*Daucus carota*]. 296aa	$1e^{-45}$	107/206 (51%)	137/206 (66%)
	CaS1/275	XP_002325634.1 AP2/ERF domain-containing transcription factor [*Populus trichocarpa*]. 313aa	$3e^{-44}$	131/281 (46%)	160/281 (56%)
	CaS2/158	CBI15759.1| unnamed protein product [*Vitis vinifera*]. 134aa	$5e^{-42}$	90/128 (70%)	103/128 (80%)
	CaS3/245	XP_002313851.1| AP2/ERF domain-containing transcription factor [*Populus trichocarpa*]. 266aa	$1e^{-36}$	99/226 (43%)	133/226 (58%)

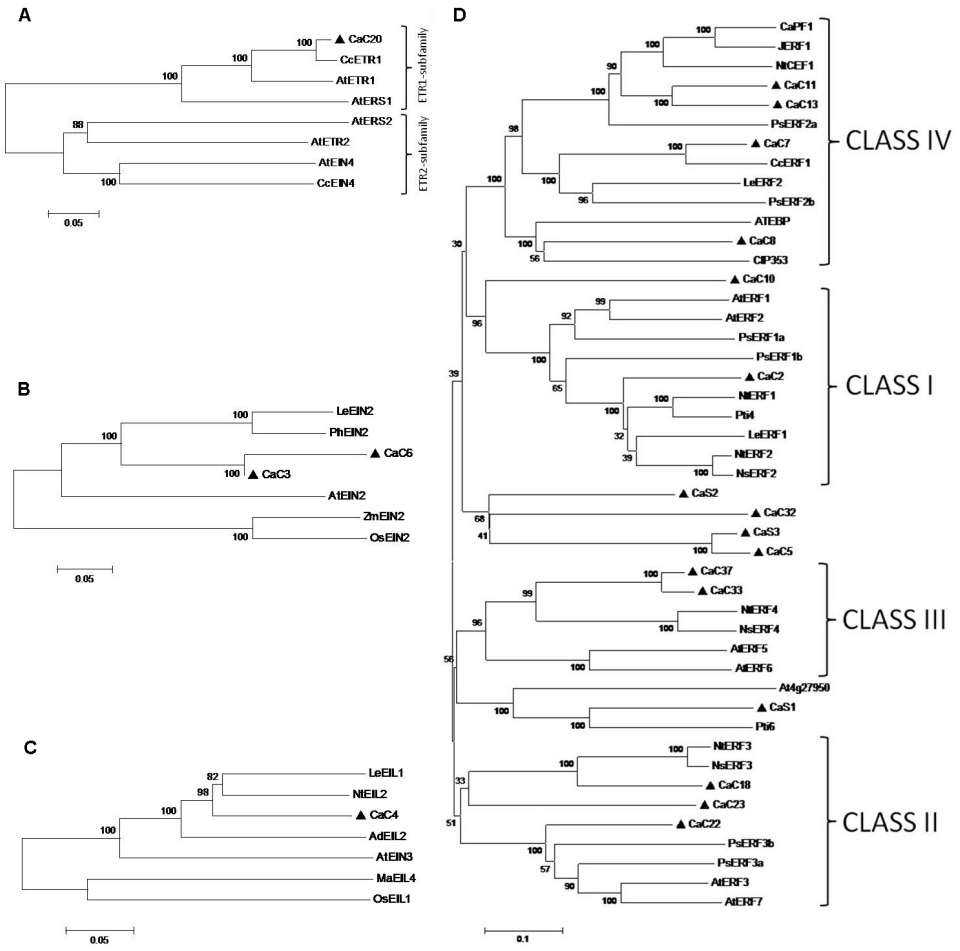

Figure 1. Phylogenetic analysis involving putative coffee ethylene signaling pathway elements (triangles) and homolog sequences obtained from the NCBI database related to **A.** ethylene receptors; **B.** EIN2 proteins; **C.** EIN3 transcriptional factors; **D.** ERF transcriptional factors. Neighbor-joining trees were built for coffee deduced amino acid and protein sequences from other species aligned with ClustalW. Bootstrap values from 1000 replications were used to assess the robustness of the trees. Genetic distances are shown at the given scales. The protein sequences from other species and their respective accession numbers are as follows: **A.** *Arabidopsis thaliana* [AtETR1(P49333), AtERS1(Q38846), AtEIN4(Q9ZTP3), AtERS2(P93825), AtETR2(O82429)]; *Coffea canephora* [CcETR1(ABL63471), CcEIN4(ABZ89180.1]. **B.** *A. thaliana* [AtEIN2(AAD41076)], *Lycopersicum esculentum* [LeEIN2(AAZ95507)], *Petunia x hybrida* [PhEIN2(AAR08678)], *Zea mays* [ZmEIN2(AAR25570)], *Oryza sativa* [OsEIN2(AAQ95276)]. **C.** *L. esculentum* [LeEIL1(AAK58857)], *Nicotiana tabacum* [NtEIL2(AAP03998)], *A. deliciosa* [AdEIL2(ACJ70675)], *A. thaliana* [AtEIN3(O24606)], *Musa acuminata* [MaEIL4(BAF44110)], *Oryza sativa* [OsEIL1(AAZ78349)]. **D.** *A. thaliana* [AtERF1(BAA32418.1), AtERF2(BAA32419.1), AtERF3(BAA32420), AtERF5(BAA32422), AtERF6(NP_567529.1), AtERF7(|NP_188666.1), AtEBP(CAA05084), At4g27950(AAT44939.1)], *L. esculentum* [LeERF1(AAO34703), LeERF2(AAO34704), JERF1(AAK95687), Pti4(AAC50047), Pti6(AAC49741.1)], *N. tabacum* [NtERF1(Q40476), NtERF2(Q40479), NtER3F(Q40477), NtERF4(Q40478), NtCEF1(AAP40022)], *N. sylvestris* [NsERF2(Q9LW50), NsERF3(Q9LW49), NsERF4(QPLW48)], *Prunus salicina* [PsERF1a(ACM49849.1), PsERF1b(ACM49848.1), PsERF3a(ACM49845.1), PsERF3b(ACM49844.1), PsERF2a(ACM49847), PsERF2b(ACM49846)], *C. canephora* [CcERF1(AAS01337)], *Solanum tuberosum* [(BAC56862.1)], *Capsicum annuum* [CaPF1(AAP72289)].

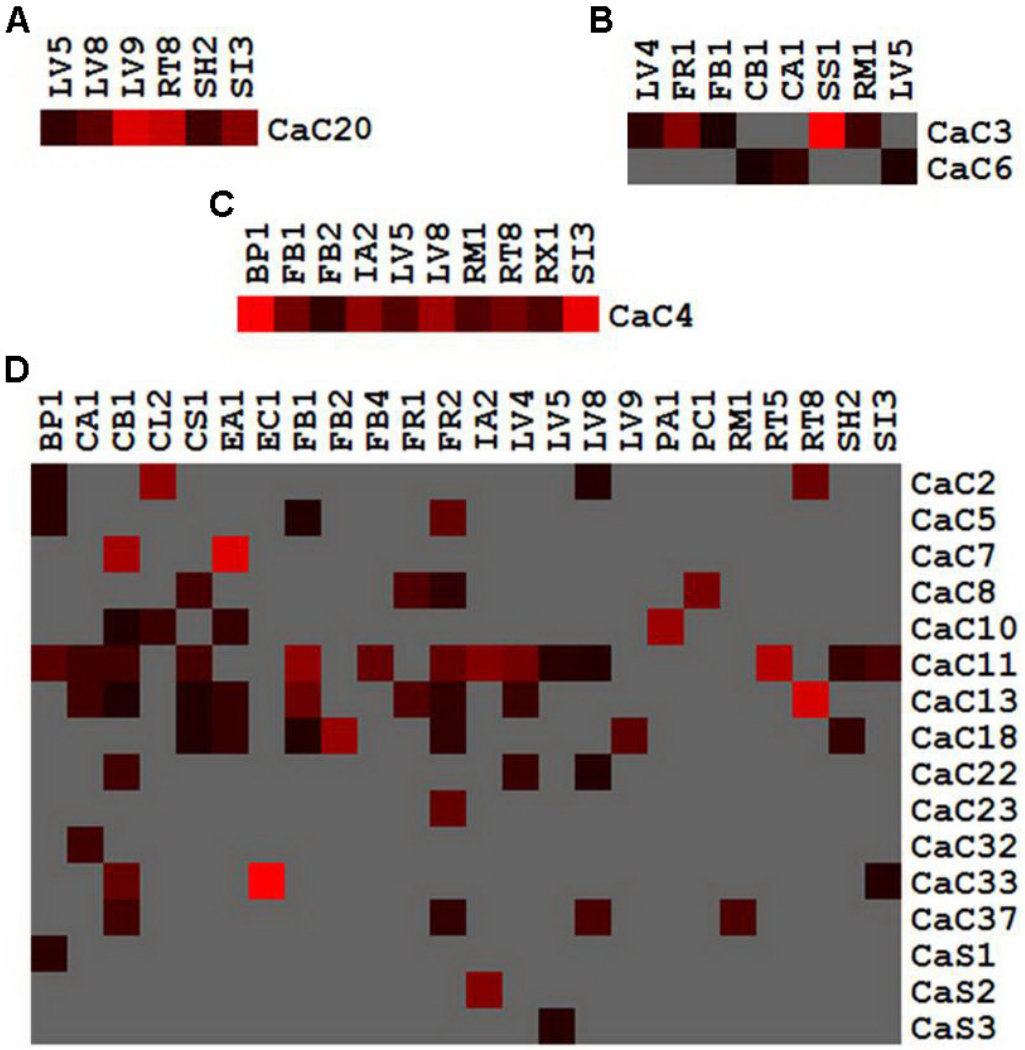

Figure 2. *In silico* expression profile of putative elements of the coffee ethylene signaling pathway: **A.** ethylene receptors; **B.** EIN2 proteins; **C.** EIN3 transcriptional factors; **D.** ERF transcriptional factors. The normalized numbers of reads for the transcripts in each library are represented in a scale from black to red. The contigs (CaC) and singlets (CaS) are represented as lines and the coffee libraries as columns. Coffee libraries are as follows (Vieira et al., 2006): suspension cells treated with acibenzolar-S-methyl (BP1); non-embryogenic calli with and without 2,4 D (CA1, PC1); suspension cells treated with acibenzolar-S-methyl and brassinoesteroids (CB1); hypocotyls treated with acibenzolar-S-methyl (CL2); suspension cells treated with NaCl (CS1); embryogenic calli (EA1, IA2); embryogenic calli from *Coffea canephora* (EC1); flower buds in different developmental stages (FB1, FB2, FB4); flower buds + pinhead fruits + fruits at different stages (FR1, FR2); young leaves from the orthotropic branch (LV4, LV5); mature leaves from plagiotropic branches (LV8, LV9); primary embryogenic calli (PA1); leaves infected with leaf miner and coffee leaf rust (RM1); roots with acibenzolar-S-methyl (RT5); suspension cells stressed with aluminum (RT8); stems infected with *Xylella* spp (RX1); water deficit stresses field plants (pool of tissues) (SH2); germinating seeds (whole seeds and zygotic embryos) (SI3); well-watered field plants (pool of tissues) (SS1).

It was not possible to identify any sequence related to CTR1 in this study, which suggests an under-representation of the ethylene signaling pathway in this database, considering the high conservation of this element in other species.

There were 10 reads related to *EIN2* within the CAFEST database, clusterized into 2 contigs. Only two contigs, CaC3 and CaC6, were found to encode proteins with similarity to the C-terminal portion of EIN2, which does not show similarity to any domain described to date, but is similar between EIN2 proteins already characterized. CaC3 and CaC6 encode for incomplete ORFs and showed an amino acid identity of 96% between each other, being more closely related to PhEIN2 and LeEIN2 (Figure 1) (Table 1). The *in silico* Northern blot analysis showed that CaC3 and CaC6 are expressed in five and three different libraries, respectively, in a no tissue-specific expression pattern, being expressed in biotic and abiotic stress conditions, as well as reproductive tissues like floral buds and fruits (Figure 2).

According to the results obtained in the CAFEST database, 53 reads related to EIN3 were found and clusterized into 4 contigs. Only one contig, CaC4, showed all the characteristic conserved domains found in the EIN3 family of transcriptional factors. CaC4 were most homologous to dicotyledonous EIN3/EIL proteins showing an amino acid identity ranging from 50 to 77% when compared to the sequences used to construct the phylogenetic tree, with NtEIL2 displaying the higher amino acid identity with CaC4 (Figure 1) (Table 1). CaC4 is formed by 29 reads and *in silico* Northern blot analysis showed that this putative coffee EIN3 homolog was expressed in 10 different libraries in a no tissue-specific manner (Figure 2). The libraries involve abiotic and biotic stress conditions, vegetative and reproductive tissues, and also germinating seeds and embriogenic calli.

Among the ethylene signaling pathway genes studied in this study, the transcriptional factors of the ERF family were the most abundant within the CAFEST database. A total of 166 reads related to ERF were found and clusterized into 21 contigs and 9 singlets, with all sequences encoding for a complete or partially complete ERF domain, which characterize these transcriptional factors. According to the amino acid identity within the ERF domain, these sequences could be separated into the two subfamilies that form the ERF family (Sakuma et al., 2002): eight contigs and six singlets belonged to CBF/DREB (C-repeat/DRE-binding factor/dehydration-responsive element binding proteins) subfamily, not involved in the ethylene signaling pathway; 13 contigs and three singlets belonged to ERF subfamily, which participate in the ethylene signaling, and were analyzed in this study.

Multiple alignments between the putative *C. arabica* ERF (Ca-ERF) proteins and previously described ERF sequences from different plant species, highlighted a number of conserved motifs and structural similarities that are commonly associated with the AP2/ERF family of transcriptional factors, and allowed the classification of putative Ca-ERFs into the four ERF classes previously described (Figure 1) (Fujimoto et al., 2000; Tournier et al., 2003). Comparison of ERF domains between Ca-ERFs and other plant ERFs showed a high sequence identity (88-98%), but full-length Ca-ERFs exhibited a considerable divergence (43-98%) with them (Table 1). The *in silico* Northern blot analysis showed that the putative Ca-ERFs are involved in a wide range of processes, since an expression in 24 different libraries, involving different tissues, development stages and stress conditions, could be observed (Figure 2).

DISCUSSION

Phylogenetic analyses

Although part of the sequences found in the CAFEST database did not correspond to complete ORFs, the search for putative homologs of the ethylene signal transduction pathway was very representative and allowed the identification of members from almost every stage of this signaling pathway.

Previous studies in *Arabidopsis* have demonstrated that the ETR family is composed of at least five receptors, divided into two subfamilies: ETR1-like subfamily (ETR1 and ERS1) and ETR2-like subfamily (ETR2, ERS2 and EIN4) (Hua et al., 1998). Bustamante-Porras et al. (2007) have characterized the ethylene receptors ETR1, ETR2 and EIN4 in *C. canephora*. Like receptors from ETR1-like subfamily, CaC20 has three N-terminal transmembrane domains, which have been predicted to adopt a helix coil, thereby creating a hydrophobic pocket that binds to ethylene. It also contains all the primordial residues for ethylene binding in its transmembrane domains (Wang et al., 2006), a GAF domain, involved in the interaction between different receptors types (Gao et al., 2008), which supposedly binds cAMP or GMP, and a C-terminal histidine kinase domain required for the interaction with the downstream component CTR1. ETR2-like subfamily II, which includes ETR2, EIN4, and ERS2, shows four transmembrane regions and a serine-threonine kinase domain in the C-terminus. Although CaC20 lacks the receptor domain in the C-terminal portion of its predicted protein, like ERS1, the results obtained suggest that CaC20 encodes an incomplete ORF being a putative *ETR1* homolog.

EIN2 is a positive regulator of the ethylene signaling pathway and it encodes an integral protein with 21 transmembrane domains on its N-terminal half, which is related to the cation transporter family N-RAMP (resistance-associated macrophage protein) and is responsible for recognizing the ethylene signal. Its unique C-terminal half is responsible for the transduction of ethylene signal to the downstream components of this pathway (Alonso et al., 1999). Although CaC3 and CaC6 encode for incomplete ORFs, both of these fragments were found to be related to the C-terminus of EIN2 proteins previously described, which does not share any homology with other described domains, suggesting that these sequences might represent putative *EIN*2 homologs. The EIN2 phylogenetic analysis confirmed the great similarity between CaC3 and CaC6 and also to other dicotyledonous EIN2 sequences.

The transcription factors EIN3 form a small gene family in *Arabidopsis*, composed by *EIN3* and five *EIN3*-like (EIL) genes, that act as positive regulators, controlling the expression of a variety of ethylene-responsive genes, including other transcriptional factors such as ERF1, a member of the ERF family of transcriptional factors (Solano et al., 1998). The contig CaC4 encodes for a protein with all structure features found in previously described EIN3 proteins such as five basic domains along its sequence, a coil structure, a rich proline region and an N-terminal acid domain. CaC4 is most closely related to dicotyledonous EIN3 sequences. A similar result was found by Mbéguié et al. (2008), where all monocotyledonous EIN3-like proteins were separately clustered from dicotyledonous sequences, except for AtEIL2 and AtEIL3 from *Arabidopsis*.

At the last step of the ethylene signaling pathway, ERFs are uniquely present in the plant kingdom and belong to the AP2/ERF superfamily of transcriptional factors (Nakano et al., 2006). All members of this superfamily are characterized by the AP2/ERF domain, and

according to the number and similarity within it, three families can be distinguished: AP2 (APETALA2), RAV (related to ABI3/VP1) and ERF (ethylene-response factors) (Riechmann et al., 2000). These transcriptional factors regulate the expression of genes involved in many biological processes related to plant growth and development, as well as environmental stimulus responses. The ERF family is composed of two subfamilies: the ERF subfamily, mainly involved in biotic stress responses, and the CBF/DREB subfamily, whose members play an important role in abiotic stress responses (Sakuma et al., 2002). These two subfamilies are separated based on the conserved amino acid residues located within the AP2/ERF domain, A13 and D18 (ERF subfamily), V13 and E18 (CBF/DREB subfamily), which are thought to contribute to a functional interaction to the cis-elements GCC-box (5'-AGCCGCC-3') and DRE/CRT (5'-TACCGACAT-3'/5'-TGGCCGAC-3'), respectively (Sakuma et al., 2002). Only a subgroup of the ERF subfamily is involved in ethylene responses (Rzewuski and Sauter, 2008).

Amino acid sequence comparison and phylogenetic analysis of other species, which characterized ERFs and the putative Ca-ERF found in the CAFEST database, allowed the allocation of many of these sequences into the four ERF classes previously described (Fujimoto et al., 2000; Tournier et al., 2003). These classes are separated according to the amino acid identity within the AP2/ERF domain (Fujimoto et al., 2000), and they are also characterized by other features such as AP2/ERF domain position, number and localization of acid domains, and also by the presence of putative nuclear localization signal (NLS) motifs.

The contig CaC2 has been classified as a member of Class I ERFs (Figure 1) (Fujimoto et al., 2000; Tournier et al., 2003), and like all class I gene members its predict proteins possess a putative NLS motif near the C-terminal region, contain an AP2/ERF domain localized near the middle of the sequence and comprise an acidic domain in the N-terminal region.

The contigs CaC18, CaC22 and CaC23 have been classified as members of Class II ERFs (Figure 1) (Fujimoto et al., 2000; Tournier et al., 2003) and are consistent with the Class II gene members previously described. All of their predict proteins showed an AP2/ERF domain in the N-terminal region and an NLS motif within the AP2/ERF domain. However, only CaC18 showed a complete sequence allowing the identification of an acidic domain and the conserved EAR (ERF-associated amphiphilic repression) repressor motif (Ohta et al., 2001) in the C-terminal end, both characteristic of class II members.

The contigs CaC33 and CaC37 encode for proteins that belong to Class III ERFs (Figure 1) (Fujimoto et al., 2000; Tournier et al., 2003), whose members are characterized by an AP2/ERF domain in the central portion of sequences, an acidic domain on both N- and C-terminal domains, and eight conserved residues (PXXSPXSP, in which X represents any amino acid) in the C-terminal end that may constitute a target of MAP (mitogen-activated protein) kinases (Pearson and Kemp, 1991). CaC33 showed an AP2/ERF domain and an N-terminal acid domain localized in a similar manner of Class III proteins previously described; however, due to its short C-terminal moiety, it was not possible to identify the C-terminal acid domain and the MAP kinase target.

The contigs CaC7, CaC8, CaC11, and CaC13 were classified as Class IV ERFs (Figure 1), which are characterized by an AP2/ERF domain close to the N-terminal portion, an NLS motif just before the AP2/ERF domain, and possess a characteristic N-terminal motif [MCGGAII/L] of unknown function (Tournier et al., 2003). All Class IV predicted proteins found in this study showed these characteristics just mentioned.

The contigs CaC5, CaC10 and CaC32 and the singlets CaS1, CaS2 and CaS3 were

shown to be related to ERFs (Table 1 and Figure 1); however, probably due to their short sequences and/or distinct features present in their sequences, they were not clustered in any of the previously described ERF classes, like other ERF genes (Figure 1).

Expression profile

The expression profile exhibited by the putative components of the coffee ethylene signaling pathway found in this study matches some of the conditions and development stages on which ethylene plays important roles (Figure 2). For instance, the expression of the putative ethylene receptor (CaC20) in the library involving stress caused by aluminum (Al) (RT8) is possibly associated with the production of ethylene under such conditions. Al is the most abundant mineral in the soil and it becomes phytotoxic to plants when it is solubilized to phytotoxic Al^{3+} species under acidic conditions. Inhibition of root elongation is one of the most distinct and earliest symptoms of Al toxicity and is caused by an increase in ethylene synthesis triggered by Al (Sun et al., 2007). Expression of CaC20 in SH2 may be explained by the fact that drought can increase the generation of ethylene in shoots, by up-regulating the synthesis and xylem transport from roots to shoots of the ethylene precursor 1-aminocyclopropane-1-carboxylic acid (ACC) (Sobeih et al., 2004), thus requiring the presence of ethylene receptors to allow ethylene to exert its effects under this condition.

The expression profile exhibited by the putative coffee *EIN2* and *EIN*3 homologs and the transcriptional factors of the ERF subfamily corroborate with the *in silico* expression of the putative coffee ethylene receptor CaC20, and also point to the involvement of ethylene in other processes and organ of plants, such as biotic stresses and reproductive tissues, respectively (Figure 2). The expression of these putative ethylene signaling members in libraries, such as RM1, RX1, BP1, and RT5, is possibly associated with ethylene's role on defense responses against biotic stresses. Ethylene biosynthesis is activated in many plants challenged by pathogens, which induce defense-related effector genes through a cascade of events in which the penultimate stage is the activation of ERF-type transcription factors. The most studied ethylene-induced defense related effector molecules are the so-called pathogenesis-related (PR) proteins, which contain the GCC-box present in their gene promoter sequences, a *cis-acting* ethylene response element that is necessary and sufficient for ERF interaction (Broekaert et al., 2006). Many ERFs have been shown to be induced upon pathogen attack, and transgenic plants over-expressing ERF genes display enhanced stress tolerance against different pathogens (Meng et al., 2010). For instance, the increased disease resistance against *Hemileia vastatrix*, showed by coffee plants treated with acibenzolar-S-methyl (Guzzo et al., 2001), may involve the participation of ERF transcription factors, since putative coffee ERFs were found to be expressed in BP1 and CB1 libraries. ERFs also have been shown to be expressed in callus (Lee et al., 2005) and seeds (Pirrello et al., 2006), corroborating with the results found in this study where the contigs CaC13, CaC32, CaC7, CaC10, CaC18, and the contigs CaC11 and CaC33, were expressed in callus tissue and germinating seeds, respectively.

On reproductive tissues, the *in silico* expression analysis of different putative elements of the coffee ethylene signaling pathway indicates that this hormone possibly plays a central role on coffee flowering and ripening. Coffee trees show an asynchronous flowering and the understanding of the mechanisms of action of the genes expressed in flower tissues, like those related to EIN2, EIN3 and ERFs found in this study, may help to explain how eth-

ylene is involved in coffee flowering. When it comes to fruit ripening, the expression of the putative coffee ethylene signaling members found in this study may be associated with the ethylene's role in ripening of climateric fruits. It has been shown that ethylene regulates different enzymes related to different ripening process such as ethylene biosynthesis (ACC synthase and ACC oxidase), breaking cell wall enzymes (polygalacturonase, pectin methylesterase and expansins), and also enzymes related to the production of volatile compounds (lipoxygenases), which affect fruit aroma (Alexander and Grierson, 2003). This regulation occurs at various levels including biosynthesis, perception, signal transduction, and target gene expression by transcription factors, such as ERFs, which play an important role in modulating ethylene-induced ripening in fruits (Bapat et al., 2010). In tomato, antisense *LeERF1* fruits displayed a longer shelf-life (Li et al., 2007), ethylene production was suppressed in antisense *LeERF2* (Zhang et al., 2009), and Sl-*ERF2* overexpression stimulated seed germination and increased apical hook formation in dark grown seedlings, indicating an increase in ethylene sensitivity (Pirrello et al., 2006). ERFs may also act like potential repressors of ripening-related genes, as has been shown by *AdERF9* in kiwifruit (Yin et al., 2010). These findings show the crucial role of ERFs in fruit ripening and corroborate with the *in silico* expression profile of putative coffee ERFs (CaC5, CaC8, CaC11, CaC13, CaC18, CaC23, CaC37) in fruit libraries (Figure 2).

This preliminary survey of coffee components of the ethylene signaling pathway has provided useful information for further studies of developmental control in this species, allowing the identification of conserved members of this signaling pathway. The results obtained indicate a high conservation of this signaling pathway between coffee and model species and the expression profile exhibited by these putative elements of the coffee ethylene signaling pathway matches many of the conditions and developmental stages in which ethylene plays important roles.

ACKNOWLEDGMENTS

We would like to thank Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café for allowing the access to the data generated by the CAFEST Genome Project, the "Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG" for the financial support and the "Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq" for research fellowships.

REFERENCES

Alexander L and Grierson D (2003). Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. *J. Exp. Bot.* 53: 2039-2055.

Alonso JM and Ecker JR (2001). The ethylene pathway: a paradigm for plant hormone signaling and interaction. *Sci. STKE.* 2001: re1.

Alonso JM, Hirayama T, Roman G, Nourizadeh S, et al. (1999). *EIN2*, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. *Science* 284: 2148-2152.

Bapat VA, Trivedi PK, Ghosh A, Sane VA, et al. (2010). Ripening of fleshy fruit: molecular insight and the role of ethylene. *Biotechnol. Adv.* 28: 94-107.

Broekaert WF, Delaure SL, De Bolle MF and Cammue BP (2006). The role of ethylene in host-pathogen interactions. *Annu. Rev. Phytopathol.* 44: 393-416.

Bustamante-Porras J, Campa C, Poncet V, Noirot M, et al. (2007). Molecular characterization of an ethylene receptor gene (*CcETR1*) in coffee trees, its relationship with fruit development and caffeine content. *Mol. Genet. Genomics* 277: 701-712.

Davies PJ (2004). Plant Hormones: Biosynthesis, Signal Transduction, Action! 3rd edn. Kluwer Academic Publishers, New York.

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.

Farnezi MMM, Silva EB, Guimarães PTG and Pinto NAVD (2010). Levantamento da qualidade da bebida do café e avaliação do estado nutricional dos cafeeiros do Alto Jequitinhonha, Minas Gerais, através do DRIS. *Cienc. Agrotec.* 34: 1191-1198.

Fujimoto SY, Ohta M, Usui A, Shinshi H, et al. (2000). *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant. Cell* 12: 393-404.

Gao Z, Wen CK, Binder BM, Chen YF, et al. (2008). Heteromeric interactions among ethylene receptors mediate signaling in *Arabidopsis*. *J. Biol. Chem.* 283: 23801-23810.

Guzzo SD, Castro RM, Kida K and Martins EMF (2001). Protection of coffee plants against coffee leaf rusty by Acibenzolar-S-Methyl. *Arq. Inst. Biol.* 68: 89-94.

Hua J, Sakai H, Nourizadeh S, Chen QG, et al. (1998). *EIN4* and *ERS2* are members of the putative ethylene receptor gene family in *Arabidopsis*. *Plant Cell* 10: 1321-1332.

Huang X and Madan A (1999). CAP3: a DNA sequence assembly program. *Genome Res.* 9: 868-877.

Lee JH, Kim DM, Lee JH, Kim J, et al. (2005). Functional characterization of *NtCEF1*, an AP2/EREBP-type transcriptional activator highly expressed in tobacco callus. *Planta* 222: 211-224.

Li Y, Zhu B, Xu W, Zhu H, et al. (2007). *LeERF1* positively modulated ethylene triple response on etiolated seedling, plant development and fruit ripening and softening in tomato. *Plant Cell Rep.* 26: 1999-2008.

Mbéguié AM, Hubert O, Fils-Lycaon B, Chillet M, et al. (2008). EIN3-like gene expression during fruit ripening of Cavendish banana (*Musa acuminata* cv. *Grande naine*). *Physiol. Plant* 133: 435-448.

Meng X, Li F, Liu C, Zhang C, et al. (2010). Isolation and characterization of an *ERF* transcription factor gene from cotton (*Gossypium barbadense* L.). *Plant Mol. Biol. Rep.* 28: 176-183.

Nakano T, Suzuki K, Fujimura T and Shinshi H (2006). Genome-wide analysis of the *ERF* gene family in *Arabidopsis* and rice. *Plant Physiol.* 140: 411-432.

Ohta M, Matsui K, Hiratsu K, Shinshi H, et al. (2001). Repression domains of class II *ERF* transcriptional repressors share an essential motif for active repression. *Plant Cell* 13: 1959-1968.

Pearson RB and Kemp BE (1991). Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. *Methods Enzymol.* 200: 62-81.

Pereira LFP, Galvão RM, Kobayashi AK, Cação SMB, et al. (2005). Ethylene production and ACC oxidase gene expression during fruit ripening of *Coffea arabica* L. *Braz. J. Plant Physiol.* 17: 283-289.

Pirrello J, Jaimes-Miranda F, Sanchez-Ballesta MT, Tournier B, et al. (2006). *Sl-ERF2*, a tomato ethylene response factor involved in ethylene response and seed germination. *Plant Cell Physiol.* 47: 1195-1205.

Riechmann JL, Heard J, Martin G, Reuber L, et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290: 2105-2110.

Rzewuski G and Sauter M (2008). Ethylene biosynthesis and signaling in rice. *Plant Sci.* 175: 32-42.

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

Sakuma Y, Liu Q, Dubouzet JG, Abe H, et al. (2002). DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* 290: 998-1009.

Salmona J, Dussert S, Descroix F, Kochko A, et al. (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.

Sobeih WY, Dodd IC, Bacon MA, Grierson D, et al. (2004). Long-distance signals regulating stomatal conductance and leaf growth in tomato (*Lycopersicon esculentum*) plants subjected to partial root-zone drying. *J. Exp. Bot.* 55: 2353-2363.

Solano R, Stepanova A, Chao Q and Ecker JR (1998). Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* 12: 3703-3714.

Sun P, Tian QY, Zhao MG, Dai XY, et al. (2007). Aluminum-induced ethylene production is associated with inhibition of root elongation in *Lotus japonicus* L. *Plant Cell Physiol.* 48: 1229-1235.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.

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.

Tournier B, Sanchez-Ballesta MT, Jones B, Pesquet E, et al. (2003). New members of the tomato *ERF* family show specific expression pattern and diverse DNA-binding capacity to the GCC box element. *FEBS Lett.* 550: 149-154.

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

Wang W, Esch JJ, Shiu SH, Agula H, et al. (2006). Identification of important regions for ethylene binding and signaling in the transmembrane domain of the *ETR1* ethylene receptor of *Arabidopsis*. *Plant Cell* 18: 3429-3442.

Yin XR, Allan AC, Chen KS and Ferguson IB (2010). Kiwifruit *EIL* and *ERF* genes involved in regulating fruit ripening. *Plant Physiol.* 153: 1280-1292.

Zhang Z, Zhang H, Quan R, Wang XC, et al. (2009). Transcriptional regulation of the ethylene response factor *LeERF2* in the expression of ethylene biosynthesis genes controls ethylene production in tomato and tobacco. *Plant Physiol.* 150: 365-377.

