Journal of Seed Science
ISSN 2317-1545
www.abrates.org.br/revista

Journal of Seed Science, v.41, n.2, p.168-179, 2019
http://dx.doi.org/10.1590/2317-1545v41n2208263

Gene expression during the germination of coffee seed[1]

Edvaldo Aparecido Amaral da Silva[2*], Marcio Luis Acencio[3], Luiz Augusto Bovolenta[4], Ney Lemke[4], Alessandro de Mello Varani[5], Juliana Pereira Bravo[2], Andrea Akemi Hoshino-Bezerra[2], Eliana Gertrudes Macedo Lemos[5]

ABSTRACT – Germination of the coffee (*Coffea arabica* L.) seed is the result of events that occur simultaneously in the embryo and endosperm. To understand the molecular mechanisms responsible for these events, we undertook a transcriptome analysis of embryo, micropylar and lateral endosperms from 10-day-imbibed seeds. The sequencing yielded contigs coding for 16,813 proteins. From those, 14,005 (~ 83%) were highly similar to at least one protein sequence in the *nr* database. 162 genes were significantly expressed in the embryo, 36 in the micropylar endosperm and 72 in the lateral endosperm. The tissue specificity analysis of the significantly expressed genes showed that the embryo had the highest proportion of specific genes (113/162, ~70%), while 11 were expressed in the micropylar and lateral endosperms. In the embryo, genes were mainly associated with abiotic stress, cell growth, and intercellular communication. In the micropylar and lateral endosperms, they were associated with abiotic stress and cell wall degradation. The accuracy of RNA-seq data was confirmed by RT-qPCR. This work adds new information about the molecular mechanism involved in coffee seed germination.

Index terms: RNA-seq, embryo, micropylar endosperm, lateral endosperm, *Coffea arabica.*

Expressão gênica durante a germinação de sementes de café

RESUMO – A germinação da semente do café (*Coffea arabica* L.) é resultado de eventos que ocorrem simultaneamente no embrião e endosperma. Para entender os mecanismos moleculares responsáveis por estes eventos, nós realizamos uma análise transcricional do embrião e dos endospermas micropilar e lateral de sementes embebidas por 10 dias. O sequenciamento obteve contigs que codificam 16.813 proteínas. Destas, 14.005 (~ 83%) foram altamente similares a pelo menos uma sequência de proteína na base de dados *nr*. 162 genes foram significativamente expressos em embrião, 36 em endosperma micropilar e 72 em endosperma lateral. A análise de tecido-especificidade dos genes significativamente expressos mostrou que a maior proporção de genes específicos está no embrião (113/162, ~70%), enquanto 11 genes foram expressos nos endospermas lateral e micropilar. No embrião, os genes foram associados principalmente com estresse abiótico, crescimento celular e comunicação intercelular. Nos endospermas micropilar e lateral, eles foram associados com estresse abiótico e degradação da parede celular. A acurácia da análise de RNA-seq foi confirmada por RT-qPCR. Este trabalho acrescenta novas informações sobre os mecanismos moleculares envolvidos na germinação de sementes de café.

Termos de indexação: RNA-seq, embrião, endosperma micropilar, endosperma lateral, *Coffea arabica.*

Introduction

Coffee is a member of the Rubiaceae family and the genus *Coffea*. There are many species of coffee in the world but only two are considered to be economically important, *Coffea arabica* and *Coffea canephora* that are responsible, respectively, for 70 and 30% of the coffee traded in the world. Coffee is considered one of the main commodities in the world, and Brazil is the largest producer and exporter, accounting for 25% of the world production. The Brazilian coffee production for the year 2018 is estimated at around 3.48 million tons (CONAB, 2018). The coffee market represents an important

[1]Submitted on 07/18/2018. Accepted for publication on 03/18/2019.
[2]Departamento de Produção e Melhoramento Vegetal, Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, 18610-307 – Botucatu, SP, Brasil.
[3]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, 8905N-7491 – Trondheim, Norway.
[4]Departamento de Física e Biofísica, Instituto de Biociências de Botucatu, Universidade Estadual Paulista, 18618-970 - Botucatu, SP, Brasil.
[5]Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, 14884-900 - Jaboticabal, SP, Brasil.
*Corresponding author: <amaral.silva@unesp.br>

source of yearly income to different segments of the Brazilian economy. Thus, to meet the demands of the coffee production chain and consumers, intensive research is continuously being made to developed new cultivars and new technologies.

Radicle protrusion in *Coffea arabica* seeds at 30 °C in the dark starts around day 5 or 6 of imbibition and 50% of the seed population displays radicle protrusion by day 10. At day 15 of imbibition most of the seeds have completed germination (Silva et al., 2004). Seedling emergence from the soil starts 50 to 60 days after sowing (Maestri and Vieira, 1961). The endosperm, which surrounds the embryo in front of the radicle tip, is referred to as the micropylar endosperm and the rest of the endosperm is called the lateral endosperm (Silva et al., 2004). The endosperm tissue has a high content of polysaccharides (Wolfrom et al., 1961). The cell walls are composed of cellulose and hemicelluloses, mainly insoluble mannans (Wolfrom and Patin, 1964). The lateral endosperm is extremely hard because the mannan is deposited as very thick cell walls; in the micropylar region, however, the walls are thinner (Silva et al., 2004). Coffee mannans contain 2% of galactose, as a side chain to the mannan backbone (Bewley et al., 2013). For germination to take place, the micropylar endosperm needs to be weakened (Silva et al., 2004). The main enzymes involved in the weakening of the endosperm of coffee seeds are the endo-β-mannanase, and β-mannosidase (Silva et al., 2004; 2005) and also the α-galactosidase (Petek and Dong, 1961). The weakening of the micropylar endosperm results in a decrease of the required puncture force to penetrate the endosperm. At this stage, we also observed porosity in the cell walls in the micropylar endosperm (Silva et al., 2004).

In addition, germination in the coffee seeds also depends on embryo growth (Silva et al., 2004). The coffee embryo is 3-4 mm long and is a differentiated tissue, in other words, when the seed is fully mature, the radicle, axis and two cotyledons are visible. During germination embryo growth in coffee seeds is controlled by cell expansion and cell wall loosening that occurs at the same time as the DNA synthesis and accumulation of β-tubulin (Silva et al., 2004; 2005).

Transcriptome studies have proven to be useful to identify the major genes expressed in some parts of the coffee plant. For example, recently transcriptome studies have been carried out on different parts of the coffee fruit during maturation (Cheng et al., 2018; 2017) in leaves, flowers and fruits (Ivamoto et al., 2017) and to identify resistance genes for *Hemileia vastatrix* infection (Flores et al., 2017). However, there is a lack of transcriptome studies in coffee seed during germination considering the different parts of the seed (embryo, lateral and micropylar endosperms). Such study may contribute to our better understanding of the molecular physiology associated with coffee seed germination. In addition, such work may provide tools to accelerate coffee seed germination and seedling formation. Thus, in the present study we established an analysis in gene expression during germination of the coffee seed in the embryo and micropylar and lateral endosperms.

Material and Methods

Coffea arabica fruits from the cultivar Catuai Vermelho were harvested in Botucatu-SP-Brazil. Fruits were manually depulped and dried to 12% moisture content at room temperature. Germination testes were done according to Silva et al. (2004). Four replications of 25 seeds were placed in Petri dishes (9.0 x 1.5 cm) on filter paper. Seeds were considered germinated when the radicle protrusion was at least one millimeter long. Germinated seeds were scored daily (Silva et al., 2004).

Total RNA was extracted separately from the embryos (Figure 1B) and the micropylar and lateral endosperms (Figure 1A) from seeds at 10 days of imbibition (prior to radicle protrusion). One hundred embryos and lateral and micropylar endosperms were frozen in liquid nitrogen and then ground to a powder with a mortar and pestle in the presence of liquid nitrogen and stored at −80 °C. Total RNA extractions were performed using the commercial NucleoSpin RNA Plant® kit (Macherey–Nagel). Extracted RNAs were quantified in a Nanodrop-2000 spectrophotometer (Thermo Scientific) and the integrity was evaluated using a Bioanalyzer (Agilent Technologies). Samples with integrity superior to 7.0 were

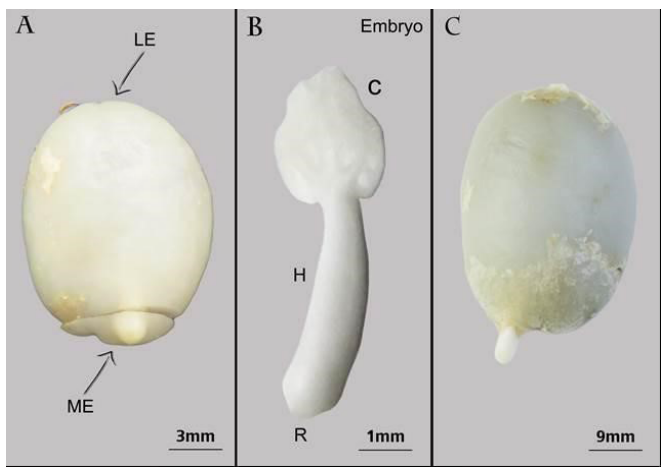

Figure 1. A. Coffee seed with indication of the micropylar (ME) and lateral endosperms (LE). B. Embryo, with indication of the cotyledons (C), axis (H) and radicle (R). C. Germinated seed (radicle protrusion).

used for the library construction.

Samples were sequenced using commercially available kits and the HiScan platform sequencing equipment (Illumina) in a 50 bp single run module. Libraries were prepared using the TruSeq RNA sample prep protocol v2. Data generated were analyzed for the assembly of the transcriptome and quantification of the expression in RPKM (Reads per Kilobase per Million of Mapped Reads) by the CLC Genomics platform Workbench Version 6.0.2 (Bio CLC). All data were deposited in the NCBI Sequence Read Archive (SRA) (BioProject PRJNA305756).

Functional annotation of contigs was performed in two steps: (1) determination of the coding potential, and (2) assignment of Gene Ontology (GO) terms (Consortium TGO, 2013). We determined the coding potential of all contigs using the support vector machine-based classifiers Coding Potential Calculator (Kong et al., 2007) and Portrait (Arrial et al., 2009). We only retained the coding contigs for the assignment of GO terms, which will be referred to as genes hereafter. For the assignment of GO terms to these genes, we used the Blast2GO (Conesa et al., 2005) with default parameters and NR database, a non-redundant set of all CDS translations from GenBank along with sequences from RefSeq, UniProtKB/Swiss-Prot, PDB and PRF proteins (NCBI Resource Coordinators, 2014). Next, GO-based functional enrichment analysis was carried out. First, the most abundant genes (*hiClu*) in each seed tissue were found using a clustering algorithm, namely simple k-means algorithms (Arthur and Vassilvitskii, 2007). The version implemented in the WEKA (Waikato Environment for Knowledge Analysis) (Hall et al., 2009) were used, with default parameters (number of clusters = 2 and Euclidean distance as a function of distance). Second, for each tissue and each GO term, that is, biological process, molecular function and cellular component, frequencies of GO terms mapped to genes within the *hiClu* were compared to those of the entire set of genes. For this, the Biological Networks Gene Ontology tool (BiNGO) (Maere et al., 2005) was used with default values for the statistical test (hypergeometric test), multiple testing corrections (Benjamini and Hochberg correction) and confidence level of 5% ($p<0.05$). For each part of the seed studied and type of ontology, the test set consisted in the set of genes forming the *hiClu* and the reference set consisted in the whole set of genes.

Tissue specific expression was checked using an online tool to draw Venn diagrams for the *hiClu* genes. In order to confirm the accuracy of the expression observed in RNA-seq data, 10 genes were selected and used in RT-qPCR reactions. The primers were obtained using the Perl Primer (v1.1.21.), the sequence resulted of RNAseq was used as model to development of primers and validation. Amplification efficiencies were obtained from the amplification plots using the program LinRegPCR the efficiency of the primers was close to 1.8 and showed an R^2 of approximately 1.0. Total RNA was extracted from 25 seeds at 10 days of imbibition using biological triplicate the cDNA synthesis was performed using the High Capacity Reverse Transcription Kit (Applied Biosystems). Real-time PCR reactions were performed in triplicate of reaction using 48 well plates (Illumina) and LuminoCt SYBR Green qPCR ReadyMix kit (Sigma). The amplification conditions were: 2 min at 50 °C, 5 min at 95 °C, followed by 40 cycles of 15 sec at 95 °C and 60 sec at 60 °C each. The final reaction volume was 10 uL, 2 uL cDNA (30 ng. ml^{-1}), 0.2 uM of each specific primer (Table 1), and 1 X Luminit Ct Mix. Data were normalized using results from 18S (Farias et al., 2015). Relative gene expression was calculated in relation to 1 day of imbibition, using $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Table 1. Primers sequences of genes from embryo and micropylar and lateral endosperms of coffee seed used to perform RT-qPCR to validate the RNA-seq data.

Gene ID	Gene name	Primer Forward (5'-3')	Primer Reverse (5'-3')
2890	Indole-3-acetic acid induced protein 4	AACTAGTCACCGTTGCTTCAT	ATTACACCTTAGGCTTCATGGA
8606	Indole-3-acetic acid induced protein 9	ATTGTCTGAGTGTTCATCGGT	GGCAAAGTTCTGAATCCCTC
3078	α-Tubulin	AGAGGGCATTATACAGTCGG	CCAATCCAGAACCAGTACCA
898	Actin 7	TGCCATCAAGAATCACCCTG	TGACTCACACCATCTCCAGA
396	Glyceraldehyde 3-phosfate dehydrogenase	TTTGGAATCAGGAATCCCGAA	AGCCTTGTCCTTGTCTGTAAA
475	Fructose bisphosfate aldolase	TTGTGGATGTGTTGAAGGAGG	TCTCACCATTAGTGCCAGGA
698	Heat Shock Protein Hsp40	TGCCATCAAGAATCACCCTG	CCACCCATTCCTTCCTTCAAT
1023	Catalase 3	AAGGTACGATCCTGTTCGTC	GTTGCTTGAAGTTGTTCTCCTT
104	Endo-β-mannanase	AGGATTTGTTAGGACTCAGGG	GACACTTTATGCCTCTCGCTA
3111	β-mannosidase	GACGAGTACCACAGATACAAGA	TTCCAATTTACTTTCCCAGTTCC

Results and Discussion

Radicle protrusion in the coffee seeds started at day five of imbibition, and 50% of the seeds had germinated by day ten (Figure 1C). To obtain approximately 100% of germination, fifteen days of imbibition were required as expected (data not shown), indicating the high quality of the seeds.

High quality RNAs extracted from the embryo and the micropylar and lateral endosperms were sequenced. The embryo yielded 7,460,714 reads, the micropylar endosperm 6,545,944 reads, and lateral endosperm 6,434,060 reads. After removal of contaminants, sequences in duplicate, and contigs with RPKM = 0 we obtained 40,381 contigs with average of 691 bp in length. In relation to the protein coding potential, 14,005 contigs were highly similar to at least one protein sequence in the *nr* database, as determined by BLASTX analysis, the first step of Blast2GO.

Thus, for each tissue, we obtained the most abundant genes (*hiClu*). For the embryo, the *hiClu* group contained 162 genes. For the micropylar and lateral endosperms, *hiClu* groups contained, respectively, 36 and 72 genes.

The expression profile attained by RNA-seq analysis was confirmed by RT-qPCR. Thus, the results obtained confirmed the presence of these ten genes during coffee seed germination (Figure 2). In addition, the results showed that the relative expression profile encountered by RT-qPCR gives reliability to the sequencing and to reference transcriptome assembly of coffee seed during germination and, also, to the analysis of the data presented.

Using hypergeometric test of false positive correction of Benjamini and Hochberg, we compared the frequencies of GO terms mapped to *hiClu* genes with the entire set of genes expressed for each part of the seed. Thus, we determined the GO terms significantly enriched (p corrected < 0.05) in the *hiClus*.

For the embryo tissue, the most prevalent biological process was response to abiotic stress, and its related terms, making up 40% of all significantly enriched terms. The three main molecular functions were those related to translation, metal ion binding and glycolysis. Among the 16 significantly enriched terms that are related to cellular components, 31% were terms related to translation and 20% were related to intercellular communication (Figure 3).

For the micropylar endosperm, as observed in the embryo, the predominant biological process was response to abiotic stress, and its related terms, making up 41% of all significantly enriched terms. The activity of the enzymes related to cell wall degradation was the predominant function (50%) among the terms related to the molecular function. Among the 12 significantly enriched terms for cellular components, 41% were related to intercellular communication (Figure 4).

For the lateral endosperm, as observed for the embryo and micropylar region, the most prevalent biological process was response to abiotic stress, and its related terms (55% of all significantly enriched terms). Among the 15 significantly

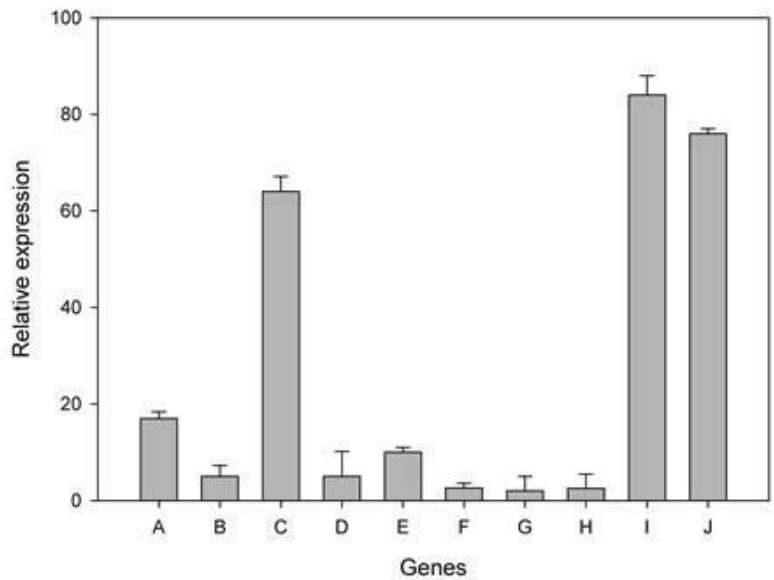

Figure 2. Relative expression of genes identified by RNA-seq in the embryo and the micropylar and lateral endosperms of coffee seeds. A. Indole-3 acetic acid-induced protein 4. B. Indole-3 acetic acid-induced protein 9. C. α-tubulin. D. Actin 7. E. Glyceraldehyde3-phosphate. F. Fructose-bisphosphate-aldolase. (G). Heat Shock Protein (Hsp40). H. Catalase 3. I. Endo-β-mannanase. J. β-mannosidase. Error bars indicate standard deviation.

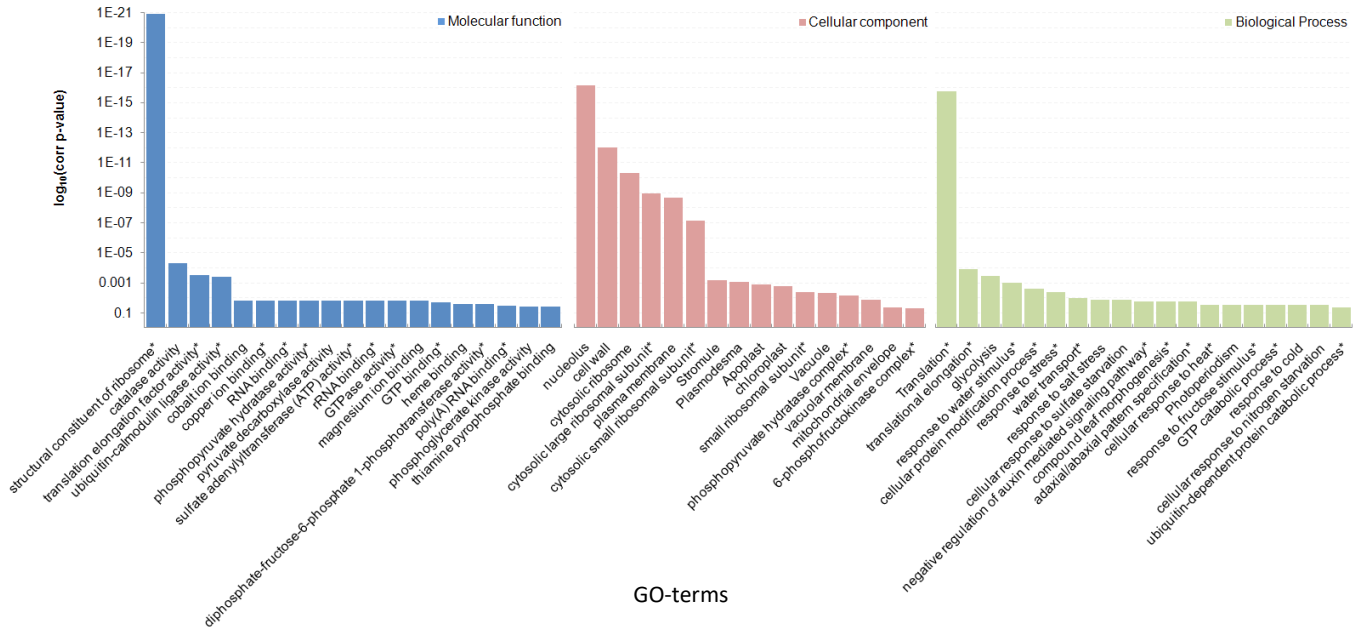

Figure 3. GO terms significantly enriched in the most abundant genes (*hiClu)* from the embryo. We carried out a log-transformation so that the smallest p-values are paired with the highest values. The terms with the larger bars are those with the lowest p-values and, consequently, they are the most significantly enriched terms. Bars marked with asterisks indicate GO terms uniquely enriched in this tissue.

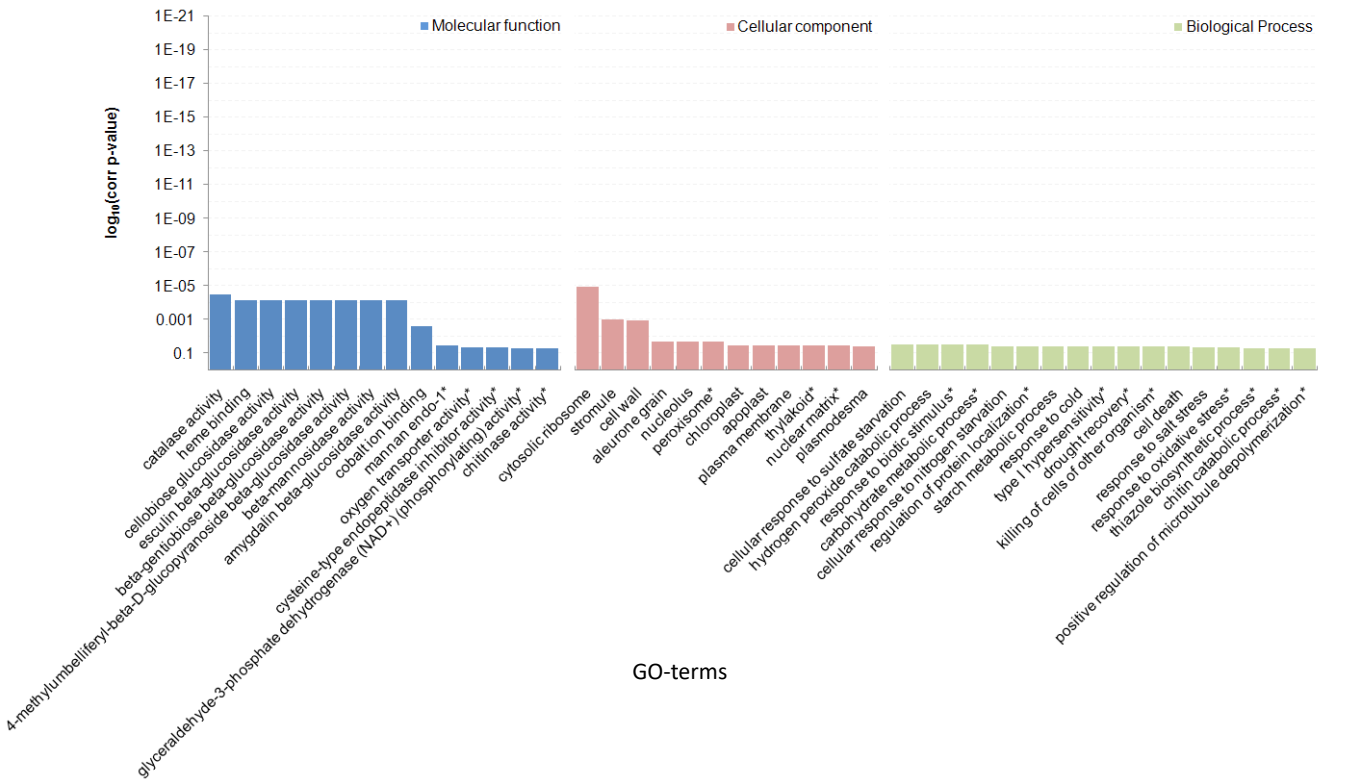

Figure 4. GO terms significantly enriched in the most abundant genes (*hiClu)* from the micropylar endosperm. We carried out a log-transformation so that the smallest p-values are paired with the highest values. The terms with the larger bars are those with the lowest p-values and, consequently, they are the most significantly enriched terms. Bars marked with asterisks indicate GO terms uniquely enriched in this tissue.

enriched terms for cellular components, 33% were related to intercellular communication and 20% were related to chloroplast (Figure 5).

To know which *hiClu* genes are specifically expressed in each tissue, we compared the presence of *hiClu* genes from each tissue to the other two sets of genes and showed the results in Table 2. The embryo showed the highest number of *hiClu* genes and also the highest proportion of specific genes (113/192, ~59%). The micropylar and lateral endosperms had 11 *hiClu* genes specifically expressed in both tissues, and eighteen genes were expressed in all three tissues (Figure 6).

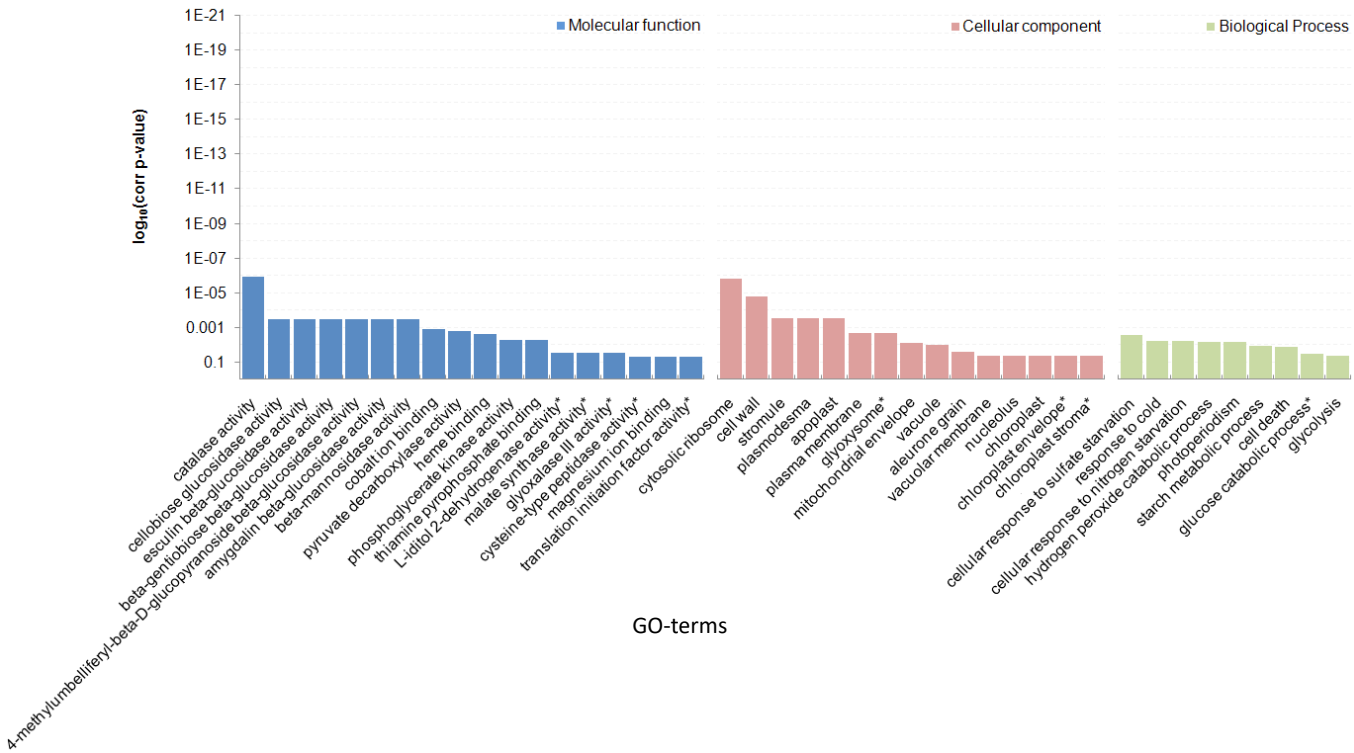

Figure 5. GO terms significantly enriched in the most abundant genes (*hiClu)* from the lateral endosperm. We carried out a log-transformation so that the smallest p-values are paired with the highest values. The terms with the larger bars are those with the lowest p-values and, consequently, they are the most significantly enriched terms. Bars marked with asterisks indicate GO terms uniquely enriched in this tissue.

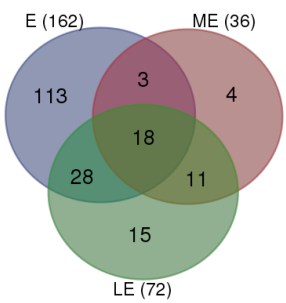

Figure 6. Venn diagram of *hiClu* genes in the samples. The embryo showed the highest number and proportion of specific genes (113/192, ~59%). The micropylar and lateral endosperm had 11 (~6%) *hiClu* genes expressed, and eighteen genes (~9%) were expressed in all three tissues.

Among genes that were expressed in all three tissues it was observed the expression, for example, of some catalases. Catalase is an intracellular enzyme, found in most organisms and their activity is associated with the removal of H_2O_2. Hydrogen peroxide is toxic to cells and may cause damage to lipids, proteins and nucleic acids. Therefore, the presence of catalase is an efficient mechanism of detoxification during imbibition (McDonald, 1999), since imbibition and the beginning of the germination process lead to increased respiratory activity and production of free radicals (Bailly, 2004). According to Hite et al. (1999), there is an increase in the activity of catalase in the scutellum of maize seeds during imbibition, and this increase occurs in parallel to the decrease in H_2O_2 levels. Apparently, the presence of catalase in coffee seeds is important for maintaining cellular integrity and avoiding damage caused by reactive oxygen species.

Table 2. Genes specifically expressed in the embryo and micropylar and lateral endosperms of coffee seeds during germination.

Gene ID	Gene name	Gene ID	Gene name
Specifically expressed in embryos			
2	60S ribosomal protein l31-like	2280	60S ribosomal protein l30
20	Adenosylhomocysteinase	2281	Polyadenylate-binding protein 2-like
80	60S ribosomal protein l28-1-like isoform 1	2300	Methionine aminopeptidase 2b
85	Polyadenylate-binding protein 2-like	2360	60S ribosomal protein l5-like
137	Eukaryotic initiation factor 4a-2-like	2397	Pectin methylesterase
174	S-adenosylmethionine synthase 1-like isoform 1	2432	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha-like
205	Elongation factor 2-like	2575	40S ribosomal protein
229	Elongation factor 1-	2649	Heat shock protein 70
232	Bag family molecular chaperone regulator 7-like	2660	Serine arginine rich splicing isoform 1
345	Small nuclear ribonucleoprotein sm d3-like	2702	60S ribosomal protein l15-like
369	Ubiquitin-like protein	2743	Glucose-1-phosphate adenylyltransferase large subunit 1-like
371	Polyadenylate-binding protein 2-like	2890	Indole-3-acetic acid induced protein 4
453	Desacetoxyvindoline 4-	2947	RNA-binding protein 39-like
502	60S acidic ribosomal protein p0	2996	Pollen-specific protein c13-like
535	Burp domain protein	3038	60S ribosomal protein l23a-like
565	60S ribosomal protein l3-like	3078	A-tubulin
568	40S ribosomal protein s5	3267	Ubiquitin-like partial
604	Probable phosphatidylinositol 4-kinase type 2-beta at1g26270-like	3296	Cysteine protease 1 precursor
610	40S ribosomal protein sa-like	3323	Alba DNA RNA-binding protein
639	Ubiquitin-like protein	3510	Vesicle-associated membrane protein 726-like
673	60S ribosomal protein l21	3542	Elongation factor 1-delta-like
690	U1 small nuclear ribonucleoprotein 70 kda-like	3726	40S ribosomal protein S9-2-like
691	Polyadenylate-binding protein rbp47-like	4054	Cathepsin b-like
784	Atp binding	4096	40S ribosomal protein
820	Hydroxyproline-rich glycoprotein	4131	Tonoplast intrinsic protein
826	Calreticulin	4445	40S ribosomal protein S15-like isoform 1
903	60S ribosomal protein l10a-1-like	4522	Pyruvate cytosolic isozyme-like
914	Nep1-interacting protein 2-like	4734	Actin-depolymerizing factor 2-like
949	Signal recognition particle 68 kda	4778	40S ribosomal protein s5-1-like
976	Calmodulin	4863	60S ribosomal protein l44-like
995	Dehydrin	4941	Gtp-binding protein sar1a-like
1015	Elongation factor 1- partial	5256	Polyadenylate-binding protein rbp45-like
1131	Enolase	5513	Serine mitochondrial-like
1151	Cbs domain protein	5840	Mlp-like protein 423-like
1176	Heat shock	6138	60s ribosomal protein l27a-3-like
1217	40S ribosomal protein s8	6674	Transcription factor bhlh63-like
1252	Pyruvate cytosolic isozyme-like	7048	Oleosin kda-like
1312	40S ribosomal protein S13-like	7450	Burp domain-containing protein
1319	Ac012654_15est gb	7525	Cation exchanger 1
1346	40S ribosomal protein S23-like	7549	Serine acetyltransferase 5-like

Table 2. Continuation.

Gene ID	Gene name	Gene ID	Gene name
Specifically expressed in embryos			
1427	40S ribosomal protein S11	7583	60S ribosomal protein l18a-like
1434	Cyclin-dependent kinase g-2-like	7823	P-loop containing nucleoside triphosphate hydrolases superfamily protein
1528	Pointed first leaf	7912	Dehydrin-like protein
1562	40S ribosomal protein S4-like	8033	60S ribosomal protein l34-like
1578	Carrier protein mitochondrial-like	8579	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha-like
1596	Protein-protein interaction regulator family	8606	Indole-3-acetic acid induced protein 9
1693	Inositol transporter 1	8891	Galactose oxidase kelch repeat superfamily protein isoform partial
1711	Probable aquaporin pip-type 7a-like	10672	Chaperone protein dnaj 20
1726	Protein disulfide-isomerase-like	10711	Rop4 small gtp binding protein
1787	Heat shock protein 90	12711	Ent-kaurene oxidase
1819	Aspartate aminotransferase	14259	Ferritin 2
1834	Ubiquitin-conjugating enzyme e2 7-like	14595	Serine carboxypeptidase-like 18-like
1945	RNA-binding protein 39-like	14935	40S ribosomal protein S14
2037	Ubiquitin-like partial	16395	Osmotin-like protein
2038	Ubiquitin-like partial	19893	Thaumation-like
2074	Hyaluronan mrna binding family isoform 1	42855	Protodermal factor
2266	Tubulin alpha chain-like		
Expressed in micropylar and lateral endosperms			
29	BURP domain-containing protein	292	Thaumation-like protein
32	Nonspecific lipid-transfer protein	376	Glutathione s-transferase t1-like
70	β-glucosidase 44-like	1238	MLP-like protein 28
104	Endo-β-mannanase	1318	Dormancy auxin associated protein
110	Cysteine proteinase inhibitor	3111	β-mannosidase
273	Cysteine proteinase		
Expressed in embryo, micropylar and lateral endosperms			
156	Protein translation factor sui1 homolog isoform 1	636	Thiazole biosynthetic enzyme
158	Elongation factor-1 alpha	698	Heat Shock Protein Hsp40
374	Catalase	765	Ascorbate peroxidase
396	Glyceraldehyde 3-Phosfate dehydrogenase	898	Actin 7
408	Ubiquitin-conjugating enzyme	1022	Catalase
423	Translationally controlled tumor protein	1379	High mobility group b3 protein
475	Fructose bisphosfate aldolase	2508	Mlp-like protein 423-like
516	Polyadenylate-binding protein-interacting p 2	2709	Catalase
608	Ubiquitin-conjugating enzyme e2 10-like	2780	Non-symbiotic hemoglobin class 1

Another protein expressed in all tissues of coffee seed was the Hsp40. Heat shock proteins are molecular chaperones expressed in response to stresses. In a study with *Arabidopsis thaliana*, there was a higher tolerance to stress caused by NaCl when the expression of protein Hsp40 was enhanced (Zhichang et al., 2010). In addition to Hsp40, this work also identified another heat shock protein, Hsp70, expressed in embryos. Although both Hsp40 and Hsp70 were identified during germination of coffee seeds in absence of stress, they may be necessary to protect the breakdown of protein bodies present

as protein reserve, which are required to meet the energy and amino acids needs during germination (Sung et al., 2001).

In the same way of heat shock proteins, calreticulins can act during the germination of coffee seeds by protecting other proteins. Calreticulin was identified in the embryo of coffee seeds during germination (Table 2). There are three calreticulins in Arabidopsis, AtCRT3 is associated with calcium homeostasis in the cell, while AtCRT1 and AtCRT2 have more general functions, as those of chaperone proteins (Christensen et al., 2010).

Actin was also observed expressed in all three tissues of coffee seeds analyzed (Table 2). Actins compose the microfilaments, which are part of the cell cytoskeleton. The cytoskeleton has important functions: mitosis, meiosis, cytokinesis, cell wall deposit, and maintenance of cell shape and cell differentiation (Taiz and Zeiger, 2004). Farias et al. (2015), in a study of gene expression in embryos of coffee (*Coffea arabica*), during germination, showed that there is an increase in the actin expression during imbibition, before radicle protrusion. Actin is responsible for movement of particles and organelles in the cytoplasm, and is also important in cell growth, since some proteins bind to actin and organize correct cell growth (Hussey et al., 2006). The actin identified in this work is highly homolog to actin 7 of *A. thaliana*, that is described as important for cell elongation.

Moreover, the β-tubulin protein, which is part of the microtubule's composition, was also observed specifically expressed in embryos in this transcriptome analyses (Table 2). In coffee embryos, microtubules have an important function in cell division and expansion and is accumulated during germination, coinciding with the expansion of the embryo cells and, also, with embryo growth, before radicle protrusion (Silva et al., 2008). Thus, the expression of microtubules (β-tubulin) and microfilaments (actin) during the germination of coffee seeds points to the contribution of these two transcripts in the embryo growth.

Regarding the control of embryo growth of coffee seeds, auxins are apparently involved in this process during germination. Auxin is known as a hormone that controls cell division, elongation and differentiation. Auxin exerts apical dominance in the formation of lateral/adventitious radicles, tropism, and cell elongation (Bewley et al., 2013). Apparently, auxin is important for the elongation of the embryo of coffee seeds, more precisely in the axis, which has been shown by Silva et al. (2008), to be the part of the embryo that shows superior growth during germination. Two transcription factors, indole-3-acetic acid induced protein 9 (IAA9) and indole-3-acetic acid induced protein 4 (IAA4), were identified in the embryo of coffee seeds (Table 2). These transcription factors act in various aspects related to short-term response to auxins, which can act as a repressor of auxin-response genes. Repression leads to degradation of AUX/AIA proteins, which are induced by AIA. Protein degradation results in interaction with auxin response factors (ARFs). ARFs bind to auxin response elements present in the promoter of the gene and initiate the transcription. Gene transcription can lead to auxin-coordinated growth, development, and elongation of cells (Taiz and Zeiger, 2004).

On the other hand, some genes were observed expressed just in the endosperms, micropylar and lateral (Table 2), as BURP domain-containing protein that is found only in plants and has increased expression in the presence of stress (Xu et al., 2010). Also, thaumation-like protein that is related to the protection of plants against fungal attack. This protein is part of the PR5 family, which is one of the 17 families of PR proteins that are present in higher amounts after a pathogen attack (Wang et al., 2010). Endosperm degradation may release sugars, which are food for fungi. Therefore, the presence of transcripts that encode proteins associated with protection against biotic stress in the endosperm may have the function of protecting the seed against the action of these fungi, which may develop as a result of lateral endosperm degradation during germination.

Other genes expressed in the micropylar and lateral endosperms were endo-β-mannanase and β-mannosidase (Table 2) that have been described as important for weakening the micropylar endosperm during germination of coffee seeds (Silva et al., 2004; Farias et al., 2015). The endosperm of coffee seeds consists of thick cell walls, which are composed of polymers rich in hemicellulose and cellulose (Bewley et al., 2013). The main hemicellulose is formed by mannans with 2% galactose present on the side chain (Wolfrom et al., 1961). Three enzymes are involved in the hydrolysis of cell wall mannans: endo-β-mannanase, β-mannosidase and β-galactosidase.

In the present work, the transcript that encodes the enzyme endo-β-mannanase, present in the micropylar endosperm, had a high level of expression and high homology with *ManA* cloned by Marracini et al. (2001), in coffee seeds after radicle protrusion. Endo-β-mannanase works on the weakening of the micropylar endosperm in coffee (Silva et al., 2004), tomato (Toorop et al., 2000), pepper (*Capsicum annuum*) (Watkins and Cantliffe, 1983), tobacco (*Nicotiana tabacum*) (Leubner-Metzger et al., 1995), and melon (*Cucumis melo*) (Welbaum et al., 1995) seeds. In coffee seeds, the activity of endo-β-mannanase begins in the micropylar endosperm before radicle protrusion, and only during germination the activity is observed in the lateral endosperm (Silva et al., 2004). Farias et al. (2015) also showed an increased expression of endo-β-

mannanase in the micropylar region of coffee seeds before radicle protrusion.

Three different isoforms of endo-β-mannanase were found in the micropylar endosperm of coffee seeds before germination, with isoelectronic points (pI) of 4.5, 6.5 and 7.0 (Silva et al., 2004). However, in the present work, we identified just one endo-β-mannanase in the micropylar and lateral endosperms. This result indicates that isoforms previously identified by other authors may be isoforms that appear after transcription. However, studies still need to be carried out to confirm this assumption.

The enzyme β-manosidase hydrolyzes the oligomers of mannose that result from endo-β-mannanase activity (Bewley et al., 2013), and it was found and characterized in seeds of several species (McCleary and Matheson, 1975; McCleary, 1982). In coffee seeds, Silva et al. (2004) quantified the activity of β-mannosidase during germination and observed increased activity of this enzyme, before radicle protrusion, in the micropylar endosperm. More recently, Farias et al. (2015) showed the presence and increase of the expression of the gene coding the enzyme β-mannosidase in the micropylar endosperm of coffee seeds during germination.

In addition to endo-β-mannanase and β-mannosidase, it was observed the presence of the enzymes β-glucosidase with expression before radicle protrusion in endosperms of coffee seeds (Table 2). β-glucosidase is a type of cellulase that works in cellulose degradation, converting it into glucose (Muñoz et al., 2001). In coffee seeds, cellulase activity occurs before radicle protrusion, and this activity is located in both the micropylar and lateral endosperms. Therefore, cellulase activity is observed throughout the (micropylar and lateral) endosperm (Silva et al., 2004), and there is no separation, in time and space, for its activity, as observed for endo-β-mannanase. According to Bewley et al. (2013), endo-β-mannanase activity, which occurs first in the micropylar endosperm, before radicle protrusion, is associated with the weakening of the endosperm in order to facilitate radicle protrusion; in comparison, the activity that occurs later during germination, in the lateral endosperm, is associated with degradation of the endosperm and contributes to the supply of reserves required for the formation and growth of seedlings. In coffee seeds, abscisic acid inhibits seed germination, but it does not inhibit cellulase activity neither in micropylar endosperm nor in lateral endosperm (Silva et al., 2004). Thus, it is unlikely that β-glucosidase has the function of weakening the endosperm to facilitate radicle protrusion in coffee seeds, as is proposed for endo-β-mannanase and β-mannosidase. Apparently, β-glucosidase works in the degradation of the endosperm to provide reserves to the embryo during germination and seedling establishment.

Conclusions

Sequencing and transcriptome analysis identified various genes expressed in the embryo of coffee seeds during germination. These transcripts play a role in the protection of the embryo against abiotic stresses, the production of energy for cellular metabolism, in embryonic growth (cell elongation and expansion), in the control of cell elongation (auxin) and in the transcription and regulation of proteins. On the other hand, in the micropylar and lateral endosperms, genes associated with the degradation of the cell wall, growth, biotic and abiotic stress were observed.

Acknowledgments

The authors thank the staff of the Genomics section of the Department of Technology, of the School of Agricultural and Veterinary Sciences, *Universidade Estadual Paulista* (UNESP-Jaboticabal), for their contribution to RNA sequencing.

References

ARRIAL, R.T.; TOGAWA, R.C.; BRIGIDO, M. Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus *Paracoccidioides brasiliensis*. *BMC Bioinformatics*, v.10, p.239, 2009. https://doi.org/10.1186/1471-2105-10-239

ARTHUR, D.; VASSILVITSKII, S. k-means++: the advantages of careful seeding. In: Proceedings of the eighteenth annual acm-siam symposium on discrete algorithms. New Orleans, 2007. p.1027-1035. https://dl.acm.org/citation.cfm?id=1283383.1283494

BAILLY, C. Active oxygen species and antioxidants in seed biology. *Seed Science Research*, v.14, p.93-107, 2004. https://doi.org/10.1079/SSR2004159

BEWLEY, J.D.; BRADFORD, K.J.; HILHORST, H.W.M.; NONOGAKI, H. *Seeds*: physiology of development, germination and dormancy. New York: Springer, 2013. 408p.

CONAB. Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira. *Café*, v.5, n.2, 2018. https://www.conab.gov.br/index.php/info-agro/safras/cafe

CHENG, B.; FURTADO, A.; HENRY, R.J. Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. *GigaScience,* v.6, p1-13, 2017. https://doi.org:10.1093/gigascience/gix086

CHENG, B.; FURTADO, A.; HENRY, R.J. The coffee bean transcriptome explains the accumulation of the major bean components through ripening. *Scientific Reports*, v.8, p.11414, 2018. https://doi.org:10.1038/s41598-018-29842-4

CONESA, A.; GÖTZ, S.; GARCIA-GOMEZ, J.M.; TEROL, J.; TALON, M.; ROBLES, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, v.21, p.3674-3676, 2005. https://doi.org/10.1093/bioinformatics/bti610

CONSORTIUM TGO. Gene Ontology Annotations and Resources. *Nucleic Acids Research*, v.41, p.D530-D535, 2013. https://doi.org/10.1093/nar/gks1050

CHRISTENSEN, A.; SVENSSON, K.; THELIN, L.; ZHANG, W.; TINTOR, N.; PRINS, D.; FUNKE, N.; MICHALAK, M.; SCHULZE-LEFERT, P.; SAIJO, Y.; SOMMARIN, M.; WIDELL, S.; PERSSON, S. Higher plant calreticulins have acquired specialized functions in Arabidopsis. *PloS One*, v.5, e11342, 2010. https://doi.org/10.1371/journal.pone.0011342

FARIAS, E.T.; SILVA, E.A.A.; TOOROP, P.E.; BEWLEY, J.D.; HILHORST, H.W.M. Expression studies in the embryo and in the micropylar endosperm of germinating coffee (*Coffea arabica* cv. Rubi) seeds. *Plant Growth Regulation*, v.75, p.575-581, 2015. https://rdcu.be/SNXm

FLORES, J.C.; MOFATTO, L.S.; FREITAS-LOPES, R.L.; FERREIRA, S.S.; ZAMBOLIM, E. M.; CARAZZOLLE, M.F.; ZAMBOLIM, L.; CAIXETA, E.T. High throughput transcriptome analysis of coffee reveals prehaustorial resistance in response to *Hemileia vastatrix* infection. *Plant Molecular Biology*, v.95, n.6, p1-17, 2017. https://doi.org:10.1007/s11103-017-0676-7

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I.H. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, v.11, p.10, 2009. http://www.kdd.org/exploration_files/p2V11n1.pdf

HITE, D.R.C.; AUH, C.; SCANDALIOS, J.G. Catalase activity and hydrogen peroxide levels are inversely correlated in maize scutella during seed germination. *Redox Report*, v.4, p.29-34, 1999. https://doi.org/10.1179/135100099101534710

HUSSEY, P.J.; KETELAAR, T.; DEEKS, M.J. Control of the Actin Cytoskeleton in Plant Cell Growth. *Annual Review of Plant Biology*, v.57, p.109-125, 2006. https://doi.org/10.1146/annurev.arplant.57.032905.105206

IVAMOTO, S.T.; REIS, O.J.; DOMINGUES, D.S.; SANTOS, T.B.; OLIVEIRA, F.F.; POT, D. LEROY, T.; VIEIRA, L.G.E.; CARAZZOLLE, M.F.; PEREIRA, G.A.G.; PEREIRA, L.F.P. Transcriptome Analysis of leaves, flowers and fruits perisperm of *Coffea arabica* L. Reveals the Differential Expression of Genes Involved in Raffinose Biosynthesis. *PLoS One*, v.12, n.1, 2017. https://doi.org/10.1371/journal.pone.0169595

KONG, L.; ZHANG, Y.; YE, Z.Q.; LIU, X.Q.; ZHAO, S.Q.; WEI, L.; GAO, G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, v.35, p.W345-W349, 2007. https://doi.org/10.1093/nar/gkm391

LEUBNER-METZGER, G.; FRUNDT, C.; VOGELI-LANGE, R.; MEINS, F. Class I [beta]-1,3-glucanases in the endosperm of tobacco during germination. *Plant Physiology*, v.109, p.751-759, 1995. https://doi.org/10.1104/pp.109.3.751

LIVAK, K.J.; SCHMITTGEN, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}$T method. *Methods*, v.25, p.402-408, 2001. https://doi.org/10.1006/meth.2001.1262

MAERE, S.; HEYMANS, K.; KUIPER, M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics*, v.21, p.3448-3449, 2005. https://doi.org/10.1093/bioinformatics/bti551

MAESTRI, M.; VIEIRA, C. Nota sobre a redução da porcentagem de germinação de sementes de café por efeito do ácido giberélico. *Revista Ceres*, v.11, p.247-249, 1961.

MARRACCINI, P.; ROGERS, W.J.; ALLARD, C. Molecular and biochemical characterization of endo-b-mannanase from germinating coffee (*Coffea arabica*) grains. *Planta,* 213, 296-308, 2001. https://doi.org/10.1007/s004250100541

MCCLEARY, B.V.; MATHESON, N.K. Galactomannan structure and β-mannanase and β-mannosidase activity in germinating legume seeds. *Phytochemistry*, v.14, p.1187-1194, 1975. https://doi.org/10.1016/S0031-9422(00)98592-3

MCCLEARY, B.V. Purification and properties of a β-d-mannoside mannohydrolase from guar. *Carbohydrate Research*, v.101, p.75-92, 1982. https://doi.org/10.1016/S0008-6215(00)80796-X

MCDONALD, M. Seed deterioration: Physiology, repair and assessment. *Seed Science and Technology*, v.27, p.177-237, 1999. https://www.researchgate.net/publication/279901910_Seed_deterioration_Physiology_repair_and_assessment

MUÑOZ, I.G.; UBHAYASEKERA, W.; HENRIKSSON, H.; SZABÓ, I.; PETTERSSON, G.; JOHANSSON, G.; MOWBRAY, S.L.; STÅHLBERG, J. Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 A resolution and homology models of the isozymes. *Journal of Molecular Biology*, v.314, p.1097-1111, 2001. https://doi.org/10.1006/jmbi.2000.5180

NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, v.42, issue D1, p.D7-D17, 2014. https://doi.org/10.1093/nar/gkt1146

PETEK, F.; DONG, T. Séparation et etude de deux alfa-galactosidases des grains du café. *Enzimologia*, v.23, p.133-142, 1961.

SILVA, E.A.A.; TOOROP, P.E.; VAN AELST, A.C.; HILHORST, H.W.M. Abscisic acid controls embryo growth potential and endosperm cap weakening during coffee (*Coffea arabica* cv. Rubi) seed germination. *Planta*, v.220, p.251-261, 2004. https://rdcu.be/SMW8

SILVA, E.A.A.; TOOROP, P.E.; NIJSSE, J.; BEWLEY, J.D.; HILHORST, H.W.M. Exogenous gibberellins inhibit coffee (*Coffea arabica* cv. Rubi) seed germination and cause cell death in the embryo. *Journal of Experimental Botany,* v.56, p.1029-1038, 2005. https://doi.org/10.1093/jxb/eri096

SILVA, E.A.A.; TOOROP, P.E.; VAN LAMMEREN, A.A.M.; HILHORST, H.W.M. ABA inhibits embryo cell expansion and early cell division events during coffee (*Coffea arabica* 'Rubi') seed germination. *Annals of Botany*, v.102, p.425-433, 2008. https://doi.org/10.1093/aob/mcn112

SUNG, D.Y.; VIERLING, E.; GUY, C.L. Comprehensive expression profile analysis of the Arabidopsis Hsp70 gene family. *Plant Physiology*, v.126, p.789-800, 2001. https://doi.org/10.1104/pp.126.2.789

TAIZ, L.; ZEIGER, E. Fisiologia do estresse. In: TAIZ, L.; ZEIGER, E. (Ed.). *Fisiologia Vegetal*. Porto Alegre: ARTMED, 2004. p.613-642.

TOOROP, P.E.; VAN AELST, A.C.; HILHORST, H.W.M. The second step of the biphasic endosperm cap weakening that mediates tomato (*Lycopersicon esculentum*) seed germination is under control of ABA. *Journal of Experimental Botany*, v.51, p.1371-1379, 2000. http://www.jstor.org/stable/23696653

XU, H.; LI, Y.; YAN, Y.; WANG, K.; GAO, Y.; HU, Y. Genome-scale identification of soybean BURP domain-containing genes and their expression under stress treatments. *BMC Plant Biology*, v.10, p.197, 2010. https://doi.org/10.1186/1471-2229-10-197

WANG, X.; TANG, C.; DENG, L.; CAI, G.; LIU, X.; LIU, B.; HAN, Q.; BUCHENAUER, H.; WEI, G.; HAN, D.; HUANG, L.; KANG, Z. Characterization of a pathogenesis-related thaumatin-like protein gene TaPR5 from wheat induced by stripe rust fungus. *Physiologia Plantarum*, v.139, p.27-38, 2010. https://doi.org/10.1111/j.1399-3054.2009.01338.x

WATKINS, J.T.; CANTLIFFE, D.J. Mechanical resistance of the seed coat and endosperm during germination of *Capsicum annuum* at low temperature. *Plant Physiology*, v.72, p.146-150, 1983. http://www.jstor.org/stable/4267990

WELBAUM, G.E.; MUTHUI, W.J.; WILSON, J.H.; GRAYSON, R.L.; FELL, R.D. Weakening of muskmelon perisperm envelope tissue 4. *Journal of Experimental Botany*, v.46, p.391-400, 1995. https://doi.org/10.1093/jxb/46.4.391

WOLFROM, M.L.; LAVER, M.L.; PATIN, D.L. Carbohydrates of the Coffee Bean. II. Isolation and characterization of a Mannan. *Journal of Organic Chemistry*, v.26, p.4533-4535, 1961. https://pubs.acs.org/doi/pdf/10.1021/jo01069a080

WOLFROM, M.L.; PATIN, D.L. Coffee Constituents, isolation and characterization of cellulose in coffee bean. *Journal of Agricultural and Food Chemistry*, v.12, p.376-377, 1964. https://pubs.acs.org/doi/pdf/10.1021/jf60134a020

ZHICHANG, Z.; WANRONG, Z.; JINPING, Y.; JIANJUN, Z.; XUFENG, L.Z.; YANG, Y. Over-expression of Arabidopsis DnaJ (Hsp40) contributes to NaCl-stress tolerance. *African Journal of Biotechnology*, v.9, p.972-978, 2010. https://www.ajol.info/index.php/ajb/article/download/78163/68545