Brazilian Journal of Botany 35(4):357-376, 2012

Influence of air temperature on proteinase activity and beverage quality in *Coffea arabica*

HELLEN MARÍLIA COUTO DE ABREU[1], PAULA MACEDO NOBILE[1,2], MILTON MASSAO SHIMIZU[1], PAULA YURI YAMAMOTO[3], EMERSON ALVES SILVA[4], CARLOS AUGUSTO COLOMBO[3] and PAULO MAZZAFERA[1,5]

(received: November 07, 2012; accepted: November 26, 2012)

ABSTRACT – (Influence of air temperature on proteinase activity and beverage quality in *Coffea arabica*). Fruits were collected from trees of *Coffea arabica* cv. Obatã grown at Mococa and Adamantina in São Paulo State, Brazil, which are regions with marked differences in air temperature that produce coffee with distinct qualities. Mococa is a cooler location that produces high-quality coffee, whereas coffee from Adamantina is of lower quality. The amino acid and protein contents, amino acid profile, and proteinase activity and type in endosperm protein extracts were analysed. Proteinase genes were identified, and their expression was assayed. All results indicate that temperature plays a role in controlling proteinase activity in coffee endosperm. Proteinase activity was higher in the endosperm of immature fruits from Adamantina, which was correlated with higher amino acid content, changes in the amino acid profile, and increased gene expression. Cysteine proteinases were the main class of proteinases in the protein extracts. These data suggest that temperature plays an important role in coffee quality by altering nitrogen compound composition.

Key words - amino acids, coffee quality, cysteine proteinase, endosperm, protein

INTRODUCTION

Coffee is cultivated in more than 80 countries and it is one of the five most important crops exported by developing countries worldwide (Marcone 2004). In recent years the coffee market is demanding increased beverage quality and, as a consequence, studies focusing on beverage quality in areas from crop management to the development of cultivars with different chemical characteristics are being performed (Decazy et al. 2003).

Beverage quality is the final result of interactions among a large number of compounds present in the endosperm (Clifford 1985). After roasting, the coffee flavour and aroma are formed by a complex set of transformations involving sugars, amino acids, peptides, trigonelline, chlorogenic acids, organic acids, lipids and carotenoids as precursors (Montavon et al. 2003).

In addition to these compounds, interactions between reducing sugars and the amino groups of some amino acids, peptides and proteins during the roasting process (the Maillard reaction) are essential for flavour and aroma development (Reineccius 1995). The Maillard reaction is known to be responsible for aroma production as well as for the production of the dark colours in different types of food through the formation of several types of compounds, including pyrazines, pyridines and furans. Most of the data available on the chemical process related to this reaction were obtained by investigating the interaction between reducing sugars and free amino acids (Ho et al. 1993), and highlight the importance of the study of free amino acids and protein profiles in immature coffee beans, which contain all of the precursors needed to develop the final beverage flavour and aroma (Montavon et al. 2003).

In this context, proteinases play a key role in beverage quality development because they alter the seed protein profile, which is likely related to flavour and aroma (Ludwig et al. 2000). Montavon et al. (2003) suggested that the levels of peptide and protein degradation among coffee beans vary with quality differences and that these variations most likely result from the activities of different endogenous proteinases in different beans. Despite the evidence of a relationship between coffee beverage quality and protein profile, there is no concrete evidence that the coffee seed storage proteins that act as aroma and flavour precursors are degraded by specific proteinases (Montavon et al. 2003). Bidimensional electrophoresis analyses of coffee seed protein extracts have identified low molecular weight peptides that are most likely produced by the action of

1. Universidade Estadual de Campinas, Instituto de Biologia, Departamento de Biologia Vegetal, Laboratório de Fisiologia Vegetal, Caixa Postal 6109, 13083-970 Campinas, SP, Brazil.
2. Instituto Agronômico de Campinas, Centro de Cana, Caixa Postal 206, 14001-970 Ribeirão Preto, SP, Brazil.
3. Instituto Agronômico de Campinas, Centro de Recursos Genéticos Vegetais, Caixa Postal 28, 13012-970 Campinas, SP, Brazil.
4. Instituto de Botânica, Núcleo de Pesquisa em Fisiologia, Caixa Postal 68041, 04045-972 São Paulo, SP, Brazil.
5. Corresponding author: pmazza@unicamp.br

proteinases that degrade the α subunit of the 11S storage protein (Rogers et al. 1999, Ludwig et al. 2000,). The 11S component is the major storage protein found in coffee seeds, and it participates in amino acid and nitrogen mobilisation for germination and the initial growth of seedlings (Shimizu *&* Mazzafera 2000). Recently, Lepeley et al. (2012) isolated two cysteine proteinases and four inhibitory gene sequences from *C. canephora* beans, and showed that proteinase expression increased in the beans of fruits as they matured; the expression of both genes was highest at the ripe stage (red stage), but *CcCP1* exhibited 600% higher expression than *CcCP4*. These authors also showed that recombinant *CcCP4* had protease activity against BSA, and assays using a specific inhibitor showed that *CcCP4* is a cysteine proteinase.

Some reports have shown that climate may have a major influence on coffee beverage quality by altering the chemical constitution of the seeds. Decazy et al. (2003) studied coffee quality in different environments in Honduras and showed that high altitudes and rainfall less 1,500 mm have favourable effects on the sensorial quality of the beverage by promoting the production of larger beans with higher lipid contents. Air temperature was the main climate factor that affected the sensory profiles of 16 green coffee samples from different locations in Réunion Island. Samples from warmer regions showed major defects in the sensory analysis, whereas positive attributes were observed in samples grown in in colder regions (Bertrand et al. 2012).

In Brazil, the optimal growth temperature for proper coffee fruit development and high-quality beverage production is between 18 °C and 22 °C (Alègre 1959, Camargo et al. 1992). Moreover, high temperatures also lead to faster coffee bean ripening, which leads to smaller and denser fruit seeds (Silva 2004, Silva et al. 2005).

The planted coffee cultivation area in Brazil is substantially large, and it comprises several climate conditions. Ortolani et al. (2000) studied the main coffee production regions of the State of São Paulo and found that multiple thermic and water conditions and their interactions (arising from the continentality gradient and altitude variations between 400 and 1,100 m) interfere with coffee tree phenology and determine beverage quality.

Sensorial analyses have shown that coffee beans from Adamantina (São Paulo State), where the annual mean temperature is approximately 24 °C, provide lower beverage quality than beans from Mococa (São Paulo State), which has an annual mean temperature of 22.5 °C that is closer to the temperature that is considered adequate for coffee production (Ortolani et al. 2000, 2001, Silva et al. 2005). Thus, the present study evaluated quantitative and qualitative aspects related to coffee proteinases of beans from these regions to establish a relationship between endogenous coffee proteinases, beverage quality and temperature in the planting region. The protein, amino acid and phenolic compound contents were analysed, proteinase genes were isolated and the expression of these proteinases was verified in coffee seeds from Adamantina and Mococa.

MATERIAL AND METHODS

Climate data and plant material

The sampled coffee trees were cultivated at the Experimental Stations of the Agronomic Institute of Campinas in Adamantina (21°41' S, 51°05' W and altitude 443 m) and Mococa (21°28' S, 47°01' W and altitude 663 m), which are both in the state of São Paulo. The trees were of the species *Coffea arabica* cv Obatã IAC 1669-20, and they were 4 years old and grown as a 2.5 × 1.0 m spaced hedgerow. The trees were irrigated with a surface drip irrigation system that provided 4 mm of water per day. The irrigation value was determined based on the mean daily evapotranspiration at Adamantina (3.8 mm) and Mococa (3.2 mm). Therefore, these plants were not drought-stressed.

Climate data (temperature) were obtained from the Integrated Centre of Agrometeorological Information (http://www.iac.sp.gov.br/ciagro). According to the Köppen International Classification (Russo Junior 1984), the macroclimates of both sites are Cwa, and the usual annual mean temperatures and total rainfall for Adamantina and Mococa are 23.1 °C and 21.8 °C and 1165 and 1442 mm, respectively. The soils of the localities are classified (Staff 1999) as Rhodic eutrustox (Adamantina) and Typic hapludult (Mococa) (Prado 2003).

Fruit harvest

Immature and mature fruits were harvested in 2005. Immature fruits were harvested in the third week of January in Adamantina and in the first week of March in Mococa, whereas mature fruits were harvested in the third week of April in Adamantina and in the first week of July in Mococa. The difference in harvesting period resulted from the 1-2 month difference in maturing rhythm in both locations; maturation is faster in Adamantina (Ortolani et al. 2000, 2001). When harvested, the immature cherries had already reached their maximum size, and the endosperm occupied almost the entire interior of the fruit. For a better indicator of the fruit development in both locations, we determined the dry mass percentages of the whole fruits, perisperm, pericarp and endosperm, and green fruits were collected when the endosperm had approximately 40-45% dry mass (Geromel et al. 2006). Mature fruits presented an intense red colour, and the endosperm had approximately 30-35% dry mass at

harvesting. Immediately after handpicking, the fruits were frozen in liquid N_2 and then stored in a -80 °C freezer. For the analyses, cherries were divided into halves with a razor blade, and the endosperm was separated using a scalpel and placed in liquid N_2. A portion of the endosperms were lyophilised for biochemical analysis, and another portion was kept in a -80 °C freezer for protein and RNA extraction.

Biochemical analysis

Lyophilised endosperms were macerated in liquid N_2 and extracted in 70% ethanol (50 mg/500 µL) for 1 h at 4 °C with occasional agitation. The solution was centrifuged at 10,000 × *g* for 20 min, and the free amino acid content of the resulting supernatant was determined (Cocking & Yemm 1954). For qualitative amino acid analysis, the milled endosperms were extracted (100 mg/2 mL) in MCW solution [methanol:chloroform:water, 12:5:3, v/v/v (Bielesk & Turner 1966)] at room temperature for 24 h with constant agitation. The solution was centrifuged at 10,000 × *g* for 15 min, and the supernatant was mixed with 1 volume of chloroform and 1.5 volumes of distilled water, followed by vigorous agitation. After another centrifugation for phase separation, the methanolic-aqueous phase was collected and dried under a speed-vac (Savant), redissolved in water and analysed using HPLC coupled with fluorimetric detection after derivation with *o*-phthaldialdehyde (Marur et al. 1994).

Proteinase activity

Endosperms from immature and mature fruits were macerated in liquid N_2, and protein extraction was performed in 0.1 M Na-phosphate buffer, pH 7.0, with 1% ascorbic acid and PVPP ($1/10^{-1}$, m/v), using 1 g of endosperm for each 7 mL of solution. Extracts were centrifuged at 12,000 × *g* for 20 min at 4 °C, and the supernatant was desalted in PD-10 Sephadex G25 columns (GE HealthCare) pre-equilibrated with 0.1 M Na-phosphate buffer, pH 7.0. The protein concentration in the desalted extract was determined with a ready-to-use Bradford reagent [GE HealthCare, (Bradford 1976)], using bovine serum albumin as a standard. Proteinase activity was determined by adding to this reaction 100 µg of protein, 0.1% azocasein (m/v) and 5 mM DTT in a final volume of 1 mL in Na-phosphate buffer, pH 7.0. DTT was added to all of the reactions because it increases coffee proteinase activity (Paulo Mazzafera, unpublished data). The reaction was performed for 1 h at 37 °C and then quenched by the addition of 500 µL of 5% trichloroacetic acid. The reaction was centrifuged at 12,000 × *g* for 10 min, and the supernatant was collected and its absorbance determined at 280 nm using a non-incubated control reaction as a reference sample. The activity was expressed as absorbance at 280 nm h^{-1} mg^{-1} protein.

For proteinase class determination, 100 µg of extracted protein was separated on an SDS-PAGE gel (10% acrylamide; Mini-Protean II, Bio-Rad), co-polymerised with 0.15% BSA, under a continuous 12 mA current for 2 h. The gel was quickly washed with distilled water, followed by renaturing buffer (100 mM Tris-HCl, pH 7.5, and 1% Triton X-100) for 20 min to remove the SDS, and finally for 20 min with 100 mM Tris-HCl, pH 7.5 buffer. The gel was then cut vertically into strips, which were placed into separate test tubes with 100 mM Tris-HCl, pH 7.5 buffer containing specific proteinase inhibitors (final concentrations: 10 mM EDTA for metalloproteinases, 15 µM iodoacetamide for cysteine proteinases, 1 µM pepstatin for aspartic proteinases, and 1 mM PMSF for serine proteinases) for 1 h at 37 °C. After this period, DTT was added to each tube at a final concentration of 5 mM, and the reactions were incubated overnight at 37 °C with slow agitation. The proteins were stained for 1 h at 37 °C with 0.1% Coomassie Blue R250 (in an aqueous solution of 45% methanol and 9% acetic acid) and then destained in 7% methanol and 5% acetic acid in water. Activity was verified through the visualisation of unstained regions in the gel because the co-polymerised BSA was digested by proteinase activity.

Proteinase gene isolation

The Coffee Genome EST database – CafEST (Vieira et al. 2006) was searched for the keywords "protease" and "proteinase". The sequences of two cysteine and two aspartic proteinases isolated from beans of *Coffea canephora* (McCarthy et al. 2007, Lepelley et al. 2012) were also used as baits. More than 600 reads were returned from these searches and used to form contigs in the BioEdit software (Hall 1999), using the CAP3 tool (Xiaoqiu 1992). Then, the contigs were analysed to determine which of them contained the highest number of CafEST reads representing cDNAs from coffee fruits and seeds. The selected contigs were compared with sequences deposited at the NCBI (National Center for Biotechnology Information), and probable open reading frames (ORF) were identified through comparisons with homologous genes from other species. Amino acid sequences were obtained by translating the nucleotide sequences using Swiss-Prot (http://ca.expasy.org/sprot/).

Total RNA was extracted from endosperms using the TRIzol reagent (Invitrogen). The RNA was used as a template for the production of first-strand cDNA using the AMV Reverse Transcriptase enzyme (Promega) and the primer B26 (Frohmann et al. 1988). The same amount of RNA was used for all samples.

Phylogenetic trees for the amino acid sequences of coffee cysteine and aspartic proteinases and their homologues were generated with a Neighbour Joining distance matrix using the default parameters in ClustalX (Thompson et al. 1997) and visualised by Mega 3.1 (Kumar et al. 2004).

Proteinase gene expression analyses

Primers for quantitative PCR (qPCR) were designed based on the sequences of the isolated genes (table 1) using the Primer Express tool (Applied Biosystems). The glyceraldehyde-3-phosphate-dehydrogenase gene (GAPDH)

Table 1. Primers designed for the real-time qpcr of proteinases and constitutive genes. GAPDH: glyceraldehyde 3-phosphate dehydrogenase. (Tm1 = forward primer fusion temperature; Tm2 = reverse primer fusion temperature; bp = amplicon sizes in base pairs).

	Tm1	%GC	Primer forward	Tm2	%GC	Primer reverse	bp
CaAP2	58	48	TGCCAAGTTTGACGGGATACT	59	43	CCATGTTATCAACAGCGATTTCC	59
CaCP4	58	38	GCTTTCCTTTTTGCTGTTGTATTG	59	55	GCTCATAGCCGCCACTAAGATC	51
CaCP1	59	44	GCAGAGTGATACATACAGCCACAAA	59	52	CATCAGACCTCCGCTTGTCAT	53
CaCP23	59	52	TGTCCATCCTGAGCTACGGTAAC	59	52	TCCCACTGCCACTCGTCTTTA	51
GAPDH	58	60	CAAGCAAGGACTGGAGAGGG	59	36	TGGGAATAATGTTAAATGAAGCAGC	58

was used as an internal control in these analyses (Barsalobres-Cavallari et al. 2009, Cruz et al. 2009). Each reaction contained 2 µL of cDNA (3 ng μL^{-1}), 0.15 mM of each primer (table 1) and 6.25 µL of SYBR Green PCR Master Mix (Applied Biosystems) in a final volume of 12.5 µL. Each sample was processed in triplicate, and a control reaction with no cDNA was performed for each primer combination. The qPCR assays were conducted in an ABI PRISM 7500 Sequence Detection System instrument (Applied Biosystems) under the following conditions: 50 °C for 2 min and 95 °C for 10 min, followed by 40 cycles at 94 °C for 15 s and 60 °C for 1 min. Two pairs of primers with similar behaviour were designed for each sequence. The relative quantification was determined by the $2^{-\Delta\Delta Ct}$ method (ΔCt = Cttag – Ctref), where Ct is the threshold cycle, tag is the tagged gene, and ref is the reference gene (Livak & Schmittgen 2001).

Statistical analysis

Samples collected from four plants were used in the biochemical analysis. The four samples were combined into two groups (two replicates of two samples) for the gene expression analyses, which were performed in triplicate. The biochemical analysis data were analysed by ANOVA, and means were compared at $P < 0.05$ using Duncan's test. For expression analysis, only means were calculated, although the standard deviation in each sample (i.e., among the six technical replicates) was less than 4%.

RESULTS

Mococa and Adamantina climate data

The maximum and minimum monthly mean temperatures registered for the years of 2004 (when the coffee plants flowered) and 2005 showed that Adamantina experienced higher temperatures than Mococa, even during the winter (figure 1). In Adamantina, the average maximum temperatures were higher than 30 °C for several months, but this occurred only occasionally in Mococa. Moreover, from August 24, 1992 to August 10, 2007, the mean maximum temperatures were 30.4 °C and 28.7 °C at Adamantina and Mococa, respectively,

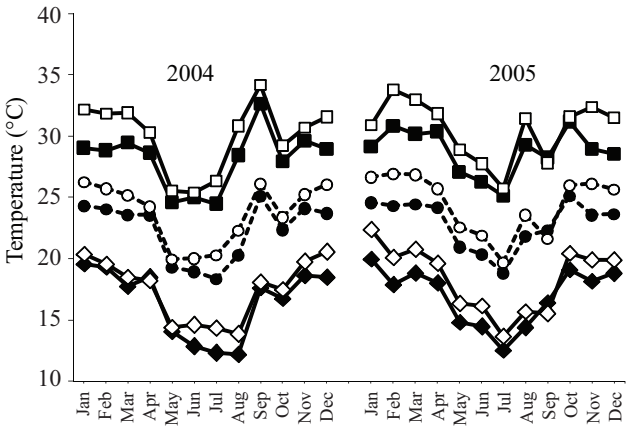


Figure 1. Mean temperature and mean maximum and minimum temperatures in Adamantina and Mococa in São Paulo State, Brazil, during 2004 and 2005. (2004. –■– = MOC – Mean Tmax (°C); -●- = MOC – Mean Temp (°C); –◇– = ADA – Mean Tmin (°C). 2005. –◆– = MOC – Mean Tmin (°C); –□– = ADA – Mean Tmax (°C); -○- = ADA – Mean Temp (°C)).

whereas the mean minimum temperatures were 17.8 °C and 16.9 °C. Thus, Adamantina consistently has a higher temperature than Mococa.

Biochemical analysis

The highest concentrations of free amino acids were found in the endosperm of immature coffee beans at both sites, and these concentrations decreased with maturation (figure 2A). Immature endosperm from Adamantina showed a higher amino acid content than that from Mococa, but no significant difference at maturation was found.

The most abundant amino acids in the immature and mature endosperm were aspartic acid (Asp), glutamic acid (Glu), asparagine (Asn), serine (Ser), glutamine (Gln), glycine (Gly) and lysine (Lys) (figures 2B and 2C), which together accounted for more than 60% of the amino acids identified; Asn was the most abundant amino acid. Only three amino acids (Asp, Asn and Lys)

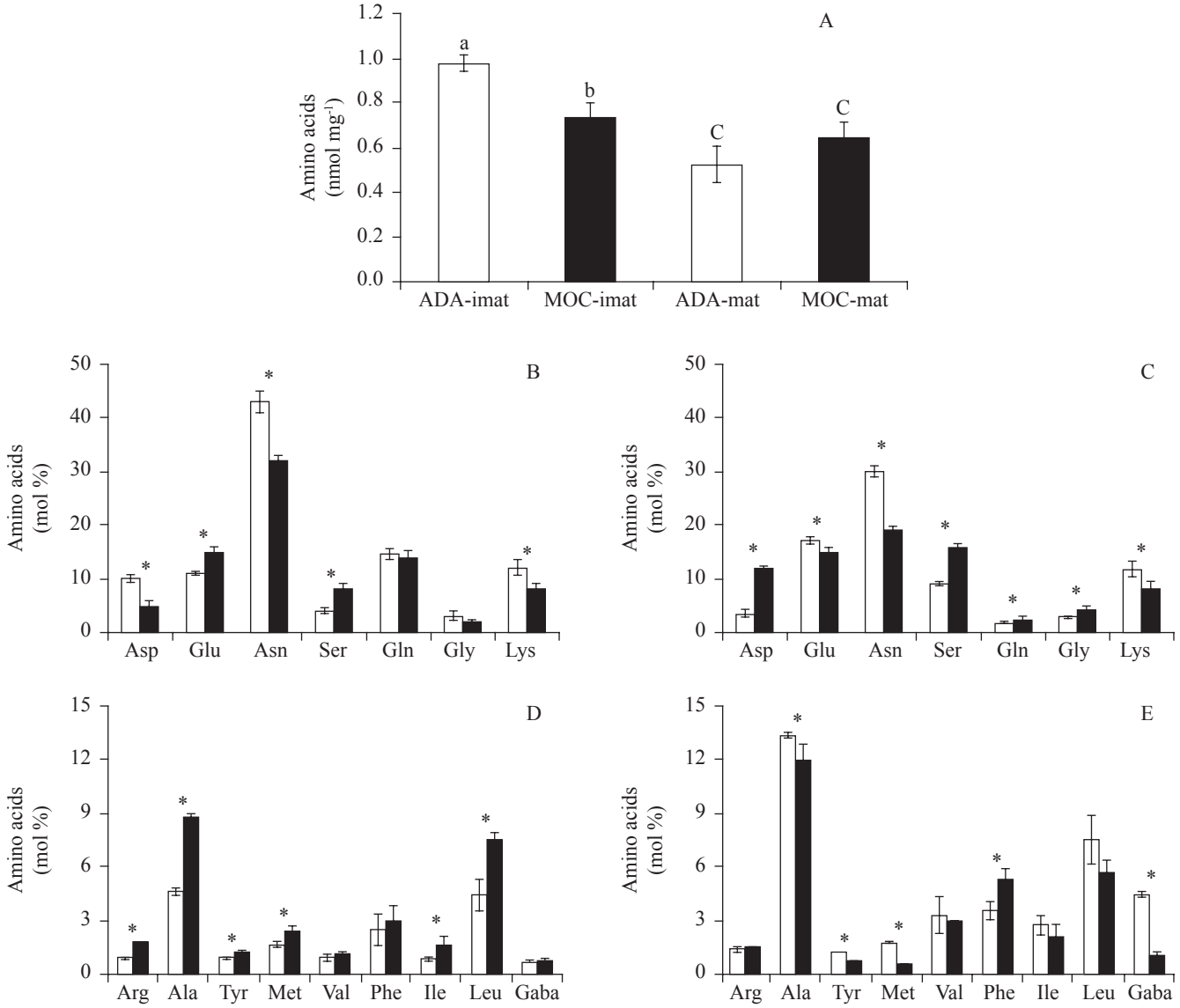


Figure 2. Free amino acid contents (A) and amino acid profiles in endosperm of immature (B, D) and mature (C, E) coffee fruits from Adamantina and Mococa. ADA, Adamantina; MOC, Mococa; Ima, immature endosperm; mat, mature endosperm. The data shown are the means of five replicates. Different letters indicate significant differences at $P < 0.05$ (Duncan's test) among all treatments, and asterisks indicate significant differences at $P < 0.05$ (Duncan's test) between sites for each amino acid. (□ = Adamantina. ■ = Mococa).

were present at significantly higher concentrations in the Adamantina endosperm when compared with the Mococa endosperm at the immature stage (figures 2B and 2D). At maturation, more amino acids were present at higher levels in the Adamantina endosperm: Glu, Asn, Lys, alanine (Ala), tyrosine (Tyr), methionine (Met), leucine (Leu) and gamma-aminobutyric acid (GABA) (figures 2C and 2E). Interestingly, in plants from both sites, there was a reduction in the Asn and Gln content from the immature to mature stages, whereas an increase in Ser, valine (Val), phenylalanine (Phe), isoleucine (Ile) and GABA was observed.

Data on soluble proteins were obtained from the extracts prepared for proteinase activity analysis (figure 3A). No difference was observed between immature and mature samples from Adamantina, but immature and mature samples from Mococa exhibited differences. The proteinase activity was clearly higher in immature endosperm than in mature endosperm (figure 3B), and the highest values were observed for immature endosperm from Adamantina. However, despite an almost four-fold difference between immature and mature endosperms, the proteinase activity was similar in mature endosperm from both sites. These same protein extracts were used

in SDS-PAGE co-polymerised with BSA to identify the proteinase classes present in coffee endosperm. However, because of the low activity in mature fruit samples, only immature fruit samples were used (figure 3C). Because the results were similar for Adamantina and Mococa extracts, only the results obtained with Adamantina extracts are shown. The strong inhibition of extract protease activity by iodoacetamide demonstrates the predominance of cysteine proteinases in the extracts, but the protease activity was also decreased by PMSF, which indicates some serine proteinase activity. Neither pepstatin nor EDTA affected the protease activity, which indicates the absence of aspartic and metalloproteinase activity, respectively.

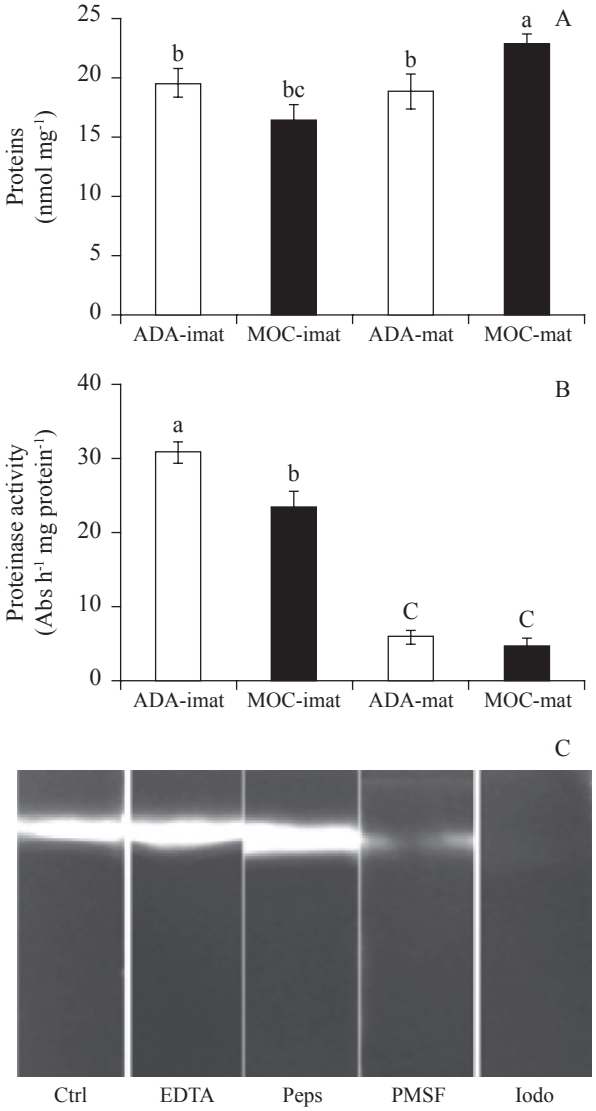


Figure 3. Soluble protein content (A) and proteinase activity (B) in extracts of endosperm from immature and mature coffe fruits from Adamantina and Mococa and SDS-PAGE for proteinase class determination using immature endosperm extract from Adamantina (C). ADA, Adamantina; MOC, Mococa; Ima, immature endosperm; mat, mature endosperm; Ctrl, control; EDTA, ethylenediamine tetra-acetic acid; Peps, pepstatin; PMSF, phenylmethanesulfonyl fluoride; Iodo, iodoacetamide. The data shown are the means of five replicates. Different letters indicate significant differences at $P < 0.05$ (Duncan's test) among all treatments.

Proteinases sequence analysis and gene expression

The CafEST database was searched for proteinases using sequence baits and keywords (protease and proteinase). Many sequences (639 reads) were returned, forming 41 contigs (615 reads) and 24 singlets. Fourteen contigs included sequences (reads) from cDNA libraries generated from fruit tissues, and these reads comprised approximately 4% to 72% of the sequences in each contig. Only three contigs identified as cysteine proteinases (*CaCP1* = 71.4%, *CaCP4* = 62.5%, *CaCP23* = 33.3%) and one aspartic proteinase (*CaAP2* = 12%) were chosen. Among all contigs and singlets, only one contig, assembled by only two reads from a leaf cDNA library, was identified as encoding a serine proteinase. Cysteine proteinases were predominant. Among the 41 contigs, 34 were identified as encoding cysteine proteinases, 6 as encoding aspartic proteinases and one as encoding a serine proteinase.

When the contigs were analysed to determine which of them contained the highest number of CafEST reads representing cDNAs from coffee fruits and seeds, we selected three cysteine proteinase contigs, i.e., *CaCP1* (JU319518), *CaCP4* (JU319519), *CaCP23* (JU319517), and one aspartic proteinase contig: *CaAP2* (JU319520).

The phylogenetic tree built from the amino acid sequences of the three cysteine proteinases from *C. arabica* and those from *C. canephora* (McCarthy et al. 2007, Lepelley et al. 2012) and sequences of homologous proteins from other species showed the formation of two large groups (figure 4A). *Ca*CP1 belonged to the first large group, whereas *Ca*CP4 and *Ca*CP23 belonged to the second large group. Although these proteinases are from the same class and have proteinase domains from the cysteine family (the C1A subfamily of papain proteinases), the grouping tree clearly shows that *Ca*CP1, *Ca*CP4 and *Ca*CP23 belong to different subgroups and are more closely related to sequences from other species than they are to each other. However, *Ca*CP1 and *Ca*CP4 were highly similar to *Cc*CP1 and *Cc*CP4 from *C. canephora*, respectively. Table 2A shows the similarity between coffee cysteine proteinases and protein sequences from other species.

Table 2. Clustalw alignment scores for amino acid coffee proteinase sequences and other sequences obtained from the NCBI OR TIGR databases. The number after the letters is the genbank protein accession number. TC22517 was obtained from the gene index database for coffee (http://compbio.dfci.harvard.edu/tgi/plant.html). (▫ = indicates the ten most similar sequences; ▪ = indicates the sequence with the highest similarity; ■ = indicates the least similar sequence).

A. Scores for cysteine proteinases – *Coffea* versus *Coffea*

	*Ca*CIP4	*Ca*CP1	*Ca*CP23	*Cc*CP1	*Cc*CP4	TC22517
*Ca*CIP4	100	30	43	30	99	39
*Ca*CP1		100	27	97	28	50
*Ca*CP23			100	29	48	31
*Cc*CP1				100	24	42
*Cc*CP4					100	35
TC22517						100

B. Scores for cysteine proteinases – *Coffea* versus other genus

Name and accession number	*Ca*CIP4	*Cc*CP4	*Ca*CP1	*Cc*CP1	*Ca*CP23
Aster42407296	34	29	52	40	34
At17979125	30	28	65	50	24
At18414611	35	31	49	39	31
At308097832	60	64	29	27	48
AtRD21a	47	48	27	29	62
Gm1096153	30	29	62	50	28
Gm31559530	66	67	29	26	48
Gm479060	30	29	62	50	28
Hv109390302	43	49	26	27	69
Ib7211745	32	30	49	37	30
Le47105731	30	30	66	57	27
Le5726641	43	46	31	34	67
Nt5051468	39	35	53	42	34
Nt58531896	65	67	29	27	51
Os115472081	35	30	51	42	28
Pt224082940	32	29	51	41	33
Pt224102377	66	69	29	26	46
Pv1345573	65	66	30	27	49
Pv2511691	33	31	48	38	32
Pv2511693	46	48	27	29	70
Sl157093728	64	66	28	27	51
Sl223049408	36	33	52	41	30
Vm445927	67	68	29	26	48
Vm7242888	33	31	49	38	31
Vs2414683	32	30	61	51	26
Vv157343944	32	30	66	49	28
Vv225427714	38	33	49	39	30
Vv225458701	45	50	28	29	68

Abbreviations of generic names: Ca = *Coffea arabica*; Cc = *Coffea canephora*; At = *Arabidopsis thaliana*; Gm = *Glycine max*; Hv = *Hevea brasiliensis*; Ib = *Ipomoea batatas*; Le = *Lycopersicon esculentum*; Nt = *Nicotiana tabacum*; Os = *Oryza sativa*; Pt = *Populus trichocarpa*; Pv = *Phaseolus vulgaris*; Sl = *Solanum lycopersicum*; Vm = *Vigna mungo*; Vs = *Vicia sativa*; Vv = *Vitis vinifera*.

C. Scores for aspartic proteinases – *Coffea* versus other species

Name and accession number	*Ca*AP2	*Cc*AP2	*Cc*AP1
*Ca*AP2	100	99	12
St82623417	73	71	12
Capsicum171854659	72	70	12
Nt294440430	72	70	12
Sl350535356	71	70	15
St50540937	71	69	15
Gm351725345	65	65	13
Vv225450440	64	63	12
At15221141	60	60	14
Pt224144963	12	12	45
Al297840891	11	14	49

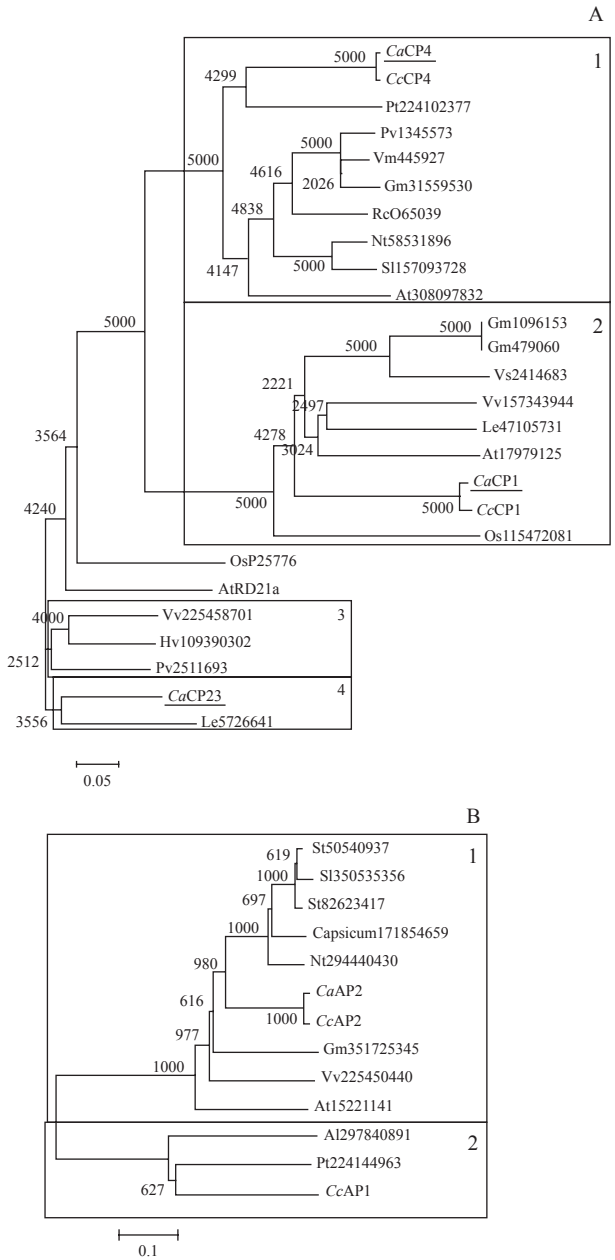


Figure 4. Neighbour-joining tree built with Mega 5.10 software (Tamura et al. 2011) from sequences aligned with ClustalX (Thompson et al. 1997) to show the relationship among the coffee cysteine (A) and aspartic (B) proteinases and several amino acid sequences encoding papain proteinases obtained from NCBI and UNIPROT (O65039 and P25776). Bootstrap values are based on 5,000 repetitions. The numbers after the plant generic names indicate NCBI accession numbers. (Ca = *Coffea arabica*; Cc = *Coffea canephora*; At = *Arabidopsis thaliana*; Gm = *Glycine max*; Hv = *Hevea brasiliensis*; Ib = *Ipomoea batatas*; Le = *Lycopersicon esculentum*; Nt = *Nicotiana tabacum*; Os = *Oryza sativa*; Pt = *Populus trichocarpa*; Pv = *Phaseolus vulgaris*; Sl = *Solanum lycopersicum*; Vm = *Vigna mungo*; Vs = *Vicia sativa*; Vv = *Vitis vinifera*; Rc = *Ricinus communis*). AtRD21a NCBI accession number: gi|18401614.

Alignments of complete sequences with homology to coffee proteinase unigenes provide a better view of the conserved regions (figure 5), which reflect the grouping shown in figure 4A. One interesting observation is that *Ca*CP23 is grouped together with genes that contain the granulin domain, which is not present the other groups (figure 5).

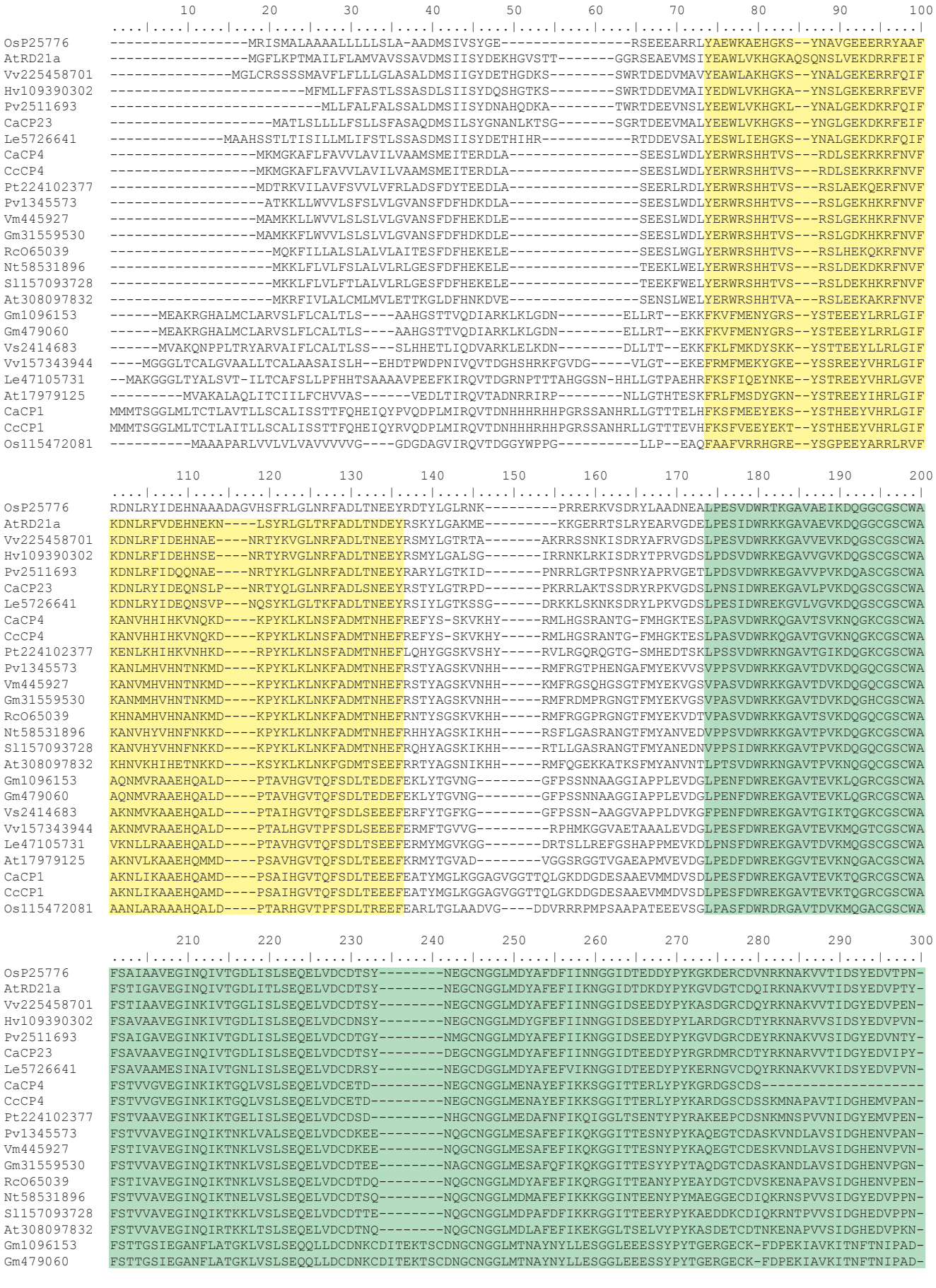
The phylogenetic tree in figure 4B shows that the amino acid sequences of the *Cc*AP2 (McCarthy et al. 2007) and *Ca*AP2 aspartic proteinases are similar, and both belong to a completely different group than *Cc*AP1. The identity between *Ca*AP2 and *Cc*AP1 is only 12% (table 2B), despite their homology within the eukaryotic aspartyl proteinase domain (Asp) region (pfam – PF00026) (figure 6).

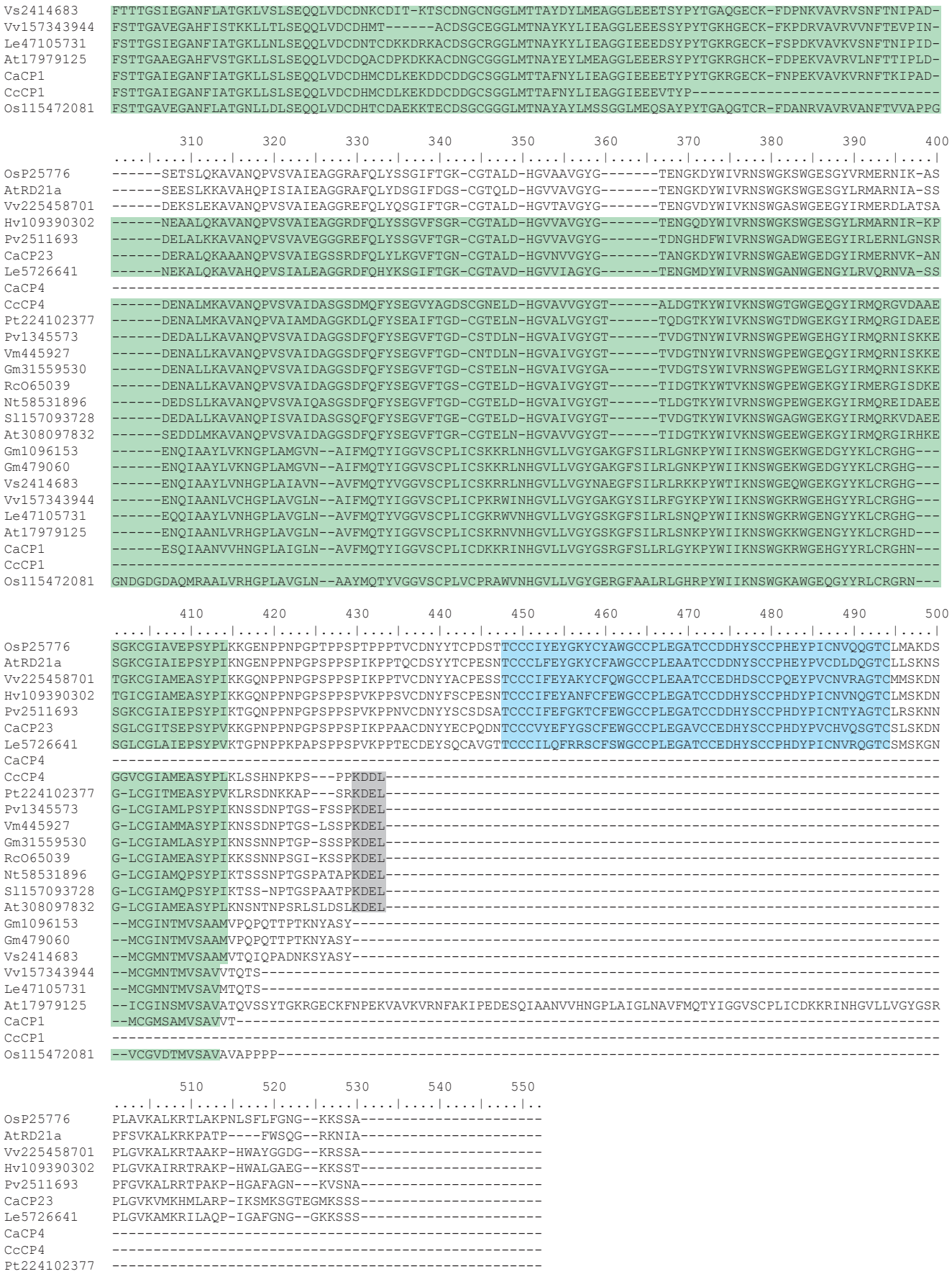
The expression analysis of the identified contigs did not show any clear pattern. Figure 7A shows the ratios of proteinase expression in the endosperm of immature and mature fruits compared between sites. In general, proteinase expression was higher in the endosperm from immature fruits. *CaCP4* was the most highly expressed gene in the Mococa fruits, whereas the *CaCP1* gene was more highly expressed in mature fruits than in green fruits in both Mococa and Adamantina. Figure 7B shows the proteinase expression ratios in Adamantina and Mococa. Proteinase expression was generally higher in Adamantina. The *CaCP4* gene showed the greatestlargest differences in expression; it was most highly expressed in mature fruits from Adamantina and in green fruits from Mococa. *CaCP23* was more highly expressed more in green fruits from Adamantina than in those from Mococa, whereas the opposite pattern was observed for *CaCP1*.

DISCUSSION

Among the compounds suggested to be involved in coffee beverage quality, amino acids and proteins play an important role (Clifford 1985, Rogers et al. 1999, Montavon et al. 2003). Nevertheless, the proportion and concentration of these compounds and how they interact during roasting to influence quality remains to be elucidated (Montavon et al. 2003).

The genetic background (Carvalho 1988, Leroy et al. 2006) and factors related to cultivation (e.g., fertilisation or plague and disease control) certainly influence the types and amounts of compounds stored in coffee seeds (DaMatta & Ramalho 2006, Geromel et al. 2006), but the post-harvest process may also influence beverage quality (Clarke 1985, Vincent 1985). After harvesting, coffee fruits may be dry- or wet-processed (Clarke 1985,


OsP25776
AtRD21a
Vv225458701
Hv109390302
Pv2511693
CaCP23
Le5726641
CaCP4
CcCP4
Pt224102377
Pv1345573
Vm445927
Gm31559530
RcO65039
Nt58531896
Sl157093728
At308097832
Gm1096153
Gm479060
Vs2414683
Vv157343944
Le47105731
At17979125
CaCP1
CcCP1
Os115472081


Vs2414683 FTTTGSIEGANFLATGKLVSLSEQQLVDCDNKCDIT-KTSCDNGCNGGLMTTAYDYLMEAGGLEEETSYPYTGAQGECK-FDPNKVAVRVSNFTNIPAD-
Vv157343944 FSTTGAVEGAHFISTKKLLTLSEQQLVDCDHMT------ACDSGCEGGLMTNAYKYLIEAGGLEEESSYPYTGKHGECK-FKPDRVAVRVVNFTEVPIN-
Le47105731 FSTTGSIEGANFIATGKLLNLSEQQLVDCDNTCDKKDRKACDSGCRGGLMTNAYKYLIEAGGIEEEDSYPYTGKRGECK-FSPDKVAVKVSNFTNIPID-
At17979125 FSTTGAAEGAHFVSTGKLLSLSEQQLVDCDQACDPKDKKACDNGCGGGLMTNAYEYLMEAGGLEEERSYPYTGKRGHCK-FDPEKVAVRVLNFTTIPLD-
CaCP1 FSTTGAIEGANFIATGKLLSLSEQQLVDCDHMCDLKEKDDCDDGCSGGLMTTAFNYLIEAGGIEEEETYPYTGKRGECK-FNPEKVAVKVRNFTKIPAD-
CcCP1 FSTTGAIEGANFIATGKLLSLSEQQLVDCDHMCDLKEKDDCDDGCSGGLMTTAFNYLIEAGGIEEEVTYP----------
Os115472081 FSTTGAVEGANFLATGNLLDLSEQQLVDCDHTCDAEKKTECDSGCGGGLMTNAYAYLMSSGGLMEQSAYPYTGAQGTCR-FDANRVAVRVANFTVVAPPG
OsP25776 ------SETSLQKAVANQPVSVAIEAGGRAFQLYSSGIFTGK-CGTALD-HGVAAVGYG-------TENGKDYWIVRNSWGKSWGESGYVRMERNIK-AS
AtRD21a ------SEESLKKAVAHQPISIAIEAGGRAFQLYDSGIFDGS-CGTQLD-HGVVAVGYG-------TENGKDYWIVRNSWGKSWGESGYLRMARNIA-SS
Vv225458701 ------DEKSLEKAVANQPVSVAIEAGGREFQLYQSGIFTGR-CGTALD-HGVTAVGYG-------TENGVDYWIVKNSWGASWGEEGYIRMERDLATSA
Hv109390302 ------NEAALQKAVANQPVSVAIEAGGRDFQLYSSGVFSGR-CGTALD-HGVVAVGYG-------TENGQDYWIVRNSWGKSWGESGYLRMARNIR-KP
Pv2511693 ------DELALKKAVANQPVSVAVEGGGREFQLYSSGVFTGR-CGTALD-HGVVAVGYG-------TDNGHDFWIVRNSWGADWGEEGYIRLERNLGNSR
CaCP23 ------DERALQKAAANQPVSVAIEGSSRDFQLYLKGVFTGN-CGTALD-HGVNVVGYG-------TANGKDYWIVRNSWGAEWGEDGYIRMERNVK-AN
Le5726641 ------NEKALQKAVAHQPVSIALEAGGRDFQHYKSGIFTGK-CGTAVD-HGVVIAGYG-------TENGMDYWIVRNSWGANWGENGYLRVQRNVA-SS
CaCP4 ----------
CcCP4 ------DENALMKAVANQPVSVAIDASGSDMQFYSEGVYAGDSCGNELD-HGVAVVGYGT------ALDGTKYWIVKNSWGTGWGEQGYIRMQRGVDAAE
Pt224102377 ------DENALMKAVANQPVAIAMDAGGKDLQFYSEAIFTGD-CGTELN-HGVALVGYGT------TQDGTKYWIVKNSWGTDWGEKGYIRMQRGIDAEE
Pv1345573 ------DEDALLKAVANQPVSVAIDAGGSDFQFYSEGVFTGD-CSTDLN-HGVAIVGYGT------TVDGTNYWIVRNSWGPEWGEHGYIRMQRNISKKE
Vm445927 ------DENALLKAVANQPVSVAIDAGGSDFQFYSEGVFTGD-CNTDLN-HGVAIVGYGT------TVDGTNYWIVRNSWGPEWGEQGYIRMQRNISKKE
Gm31559530 ------DENALLKAVANQPVSVAIDAGGSDFQFYSEGVFTGD-CSTELN-HGVAIVGYGA------TVDGTSYWIVRNSWGPEWGELGYIRMQRNISKKE
RcO65039 ------DENALLKAVANQPVSVAIDAGGSDFQFYSEGVFTGS-CGTELD-HGVAIVGYGT------TIDGTKYWTVKNSWGPEWGEKGYIRMERGISDKE
Nt58531896 ------DEDSLLKAVANQPVSVAIQASGSDFQFYSEGVFTGD-CGTELD-HGVAIVGYGT------TLDGTKYWIVRNSWGPEWGEKGYIRMQREIDAEE
Sl1157093728 ------DEDALLKAVANQPISVAIDASGSQFQFYSEGVFTGE-CGTELD-HGVAIVGYGT------TVDGTKYWIVKNSWGAGWGEKGYIRMQRKVDAEE
At308097832 ------SEDDLMKAVANQPVSVAIDAGGSDFQFYSEGVFTGR-CGTELN-HGVAVVGYGT------TIDGTKYWIVKNSWGEEWGEKGYIRMQRGIRHKE
Gm1096153 ------ENQIAAYLVKNGPLAMGVN--AIFMQTYIGGVSCPLICSKKRLNHGVLLVGYGAKGFSILRLGNKPYWIIKNSWGEKWGEDGYYKLCRGHG---
Gm479060 ------ENQIAAYLVKNGPLAMGVN--AIFMQTYIGGVSCPLICSKKRLNHGVLLVGYGAKGFSILRLGNKPYWIIKNSWGEKWGEDGYYKLCRGHG---
Vs2414683 ------ENQIAAYLVNHGPLAIAVN--AVFMQTYVGGVSCPLICSKRRLNHGVLLVGYNAEGFSILRLRKKPYWTIKNSWGEQWGEKGYYKLCRGHG---
Vv157343944 ------ENQIAANLVCHGPLAVGLN--AIFMQTYIGGVSCPLICPKRWINHGVLLVGYGAKGYSILRFGYKPYWIIKNSWGKRWGEHGYYRLCRGHG---
Le47105731 ------EQQIAAYLVNHGPLAVGLN--AVFMQTYVGGVSCPLICGKRWVNHGVLLVGYGSKGFSILRLSNQPYWIIKNSWGKRWGENGYYKLCRGHG---
At17979125 ------ENQIAANLVRHGPLAVGLN--AVFMQTYIGGVSCPLICSKRNVNHGVLLVGYGSKGFSILRLSNKPYWIIKNSWGKKWGENGYYKLCRGHD---
CaCP1 ------ESQIAANVVHNGPLAIGLN--AVFMQTYIGGVSCPLICDKKRINHGVLLVGYGSRGFSLLRLGYKPYWIIKNSWGKRWGEHGYYRLCRGHN---
CcCP1 ----------
Os115472081 GNDGDGDAQMRAALVRHGPLAVGLN--AAYMQTYVGGVSCPLVCPRAWVNHGVLLVGYGERGFAALRLGHRPYWIIKNSWGKAWGEQGYYRLCRGRN---
OsP25776 SGKCGIAVEPSYPLKKGENPPNPGPTPPSPTPPPTVCDNYYTCPDSTTCCCIYEYGKYCYAWGCCPLEGATCCDDHYSCCPHEYPICNVQQGTCLMAKDS
AtRD21a SGKCGIAIEPSYPIKNGENPPNPGPSPPSPIKPPTQCDSYYTCPESNTCCCLFEYGKYCFAWGCCPLEAATCCDDNYSCCPHEYPVCDLDQGTCLLSKNS
Vv225458701 TGKCGIAMEASYPIKKGQNPPNPGPSPPSPIKPPTVCDNYYACPESSTCCCIFEYAKYCFQWGCCPLEAATCCEDHDSCCPQEYPVCNVRAGTCMMSKDN
Hv109390302 TGICGIAMEASYPIKKGQNPPNPGPSPPSPVKPPSVCDNYFSCPESNTCCCIFEYANFCFEWGCCPLEGATCCDDHYSCCPHDYPICNVNQGTCLMSKDN
Pv2511693 SGKCGIAIEPSYPIKTGQNPPNPGPSPPSPVKPPNVCDNYYSCSDSATCCCIFEFGKTCFEWGCCPLEGATCCDDHYSCCPHDYPICNTYAGTCLRSKNN
CaCP23 SGLCGITSEPSYPVKKGPNPPNPGPSPPSPIKPPAACDNYYECPQDNTCCCVYEFYGSCFEWGCCPLEGAVCCEDHYSCCPHDYPVCHVQSGTCSLSKDN
Le5726641 SGLCGLAIEPSYPVKTGPNPPKPAPSPPSPVKPPTECDEYSQCAVGTTCCCILQFRRSCFSWGCCPLEGATCCEDHYSCCPHDYPICNVRQGTCSMSKGN
CaCP4 ----------
CcCP4 GGVCGIAMEASYPLKLSSHNPKPS---PPKDDL----------
Pt224102377 G-LCGITMEASYPVKLRSDNKKAP---SRKDEL----------
Pv1345573 G-LCGIAMLPSYPIKNSSDNPTGS-FSSPKDEL----------
Vm445927 G-LCGIAMMASYPIKNSSDNPTGS-LSSPKDEL----------
Gm31559530 G-LCGIAMLASYPIKNSSNNPTGP-SSSPKDEL----------
RcO65039 G-LCGIAMEASYPIKKSSNNPSGI-KSSPKDEL----------
Nt58531896 G-LCGIAMQPSYPIKTSSSNPTGSPATAPKDEL----------
Sl1157093728 G-LCGIAMQPSYPIKTSS-NPTGSPAATPKDEL----------
At308097832 G-LCGIAMEASYPLKNSNTNPSRLSLDSLKDEL----------
Gm1096153 --MCGINTMVSAAMVPQPQTTPTKNYASY----------
Gm479060 --MCGINTMVSAAMVPQPQTTPTKNYASY----------
Vs2414683 --MCGMNTMVSAAMVTQIQPADNKSYASY----------
Vv157343944 --MCGMNTMVSAVVTQTS----------
Le47105731 --MCGMNTMVSAVMTQTS----------
At17979125 --ICGINSMVSAVATQVSSYTGKRGECKFNPEKVAVKVRNFAKIPEDESQIAANVVHNGPLAIGLNAVFMQTYIGGVSCPLICDKKRINHGVLLVGYGSR
CaCP1 --MCGMSAMVSAVVT----------
CcCP1 ----------
Os115472081 --VCGVDTMVSAVAVAPPPP----------
OsP25776 PLAVKALKRTLAKPNLSFLFGNG--KKSSA----------
AtRD21a PFSVKALKRKPATP----FWSQG--RKNIA----------
Vv225458701 PLGVKALKRTAAKP-HWAYGGDG--KRSSA----------
Hv109390302 PLGVKAIRRTRAKP-HWALGAEG--KKSST----------
Pv2511693 PFGVKALRRTPAKP-HGAFAGN---KVSNA----------
CaCP23 PLGVKVMKHMLARP-IKSMKSGTEGMKSSS----------
Le5726641 PLGVKAMKRILAQP-IGAFGNG--GKKSSS----------
CaCP4 ----------
CcCP4 ----------
Pt224102377 ----------
Pv1345573 ----------

```
Vm445927     --------------------------------------------------
Gm31559530   --------------------------------------------------
RcO65039     --------------------------------------------------
Nt58531896   --------------------------------------------------
Sl157093728  --------------------------------------------------
At308097832  --------------------------------------------------
Gm1096153    --------------------------------------------------
Gm479060     --------------------------------------------------
Vs2414683    --------------------------------------------------
Vv157343944  --------------------------------------------------
Le47105731   --------------------------------------------------
At17979125   GFSILRLGYKPYWIIKNSWGKRWGEHGCYRLCRGHNMCGMSTMVSAVVTQTS
CaCP1        --------------------------------------------------
CcCP1        --------------------------------------------------
Os115472081  --------------------------------------------------
```

Figure 5. Multiple alignment of the deduced amino acid sequences of *Coffea arabica* cysteine proteinases. Protein domains (pfam – http://pfam.sanger.ac.uk/) are highlighted as follows: □ = cathepsin propeptide inhibitor domain (I29); □ = papaya proteinase I; □ = granulin; □ = indicates C-terminal KDEL sequence.

```
                          10        20        30        40        50        60
                  ....|....|....|....|....|....|....|....|....|....|....|....|
CaAP2             ---MERRYLWAAFVLGAIVCSLFPLPSEGLKRISLKKKPLDIQSIRAAKLAHLESTHGAG 57
CcAP2             ---MERRYLWAA--LGAIVCSLFPLPSEGLKRISLKKKPLDIQSIRAAKLAHLESTHGAG 55
St50540937        ---MDKKHLCAALLLWAITCSALPASSGDLLRIGLKKHRLDVNSIKAARVAKLQDRYGKH 57
Sl350535356       ---MDKKHLCAALLLWAIACSALPASSGDLFRIGLKKHRLDVDSIKAARVAKLQDRYGKH 57
St82623417        ---MEKKHLCAALLLWAITCSALPASSGDLLRIGLKKHRLDVNSIKAARVAKLQDRYGKH 57
Capsicum171854659 ---MENKHLCAALLLLAIACSVLPASSDNLLRIGLKKHHVDVNSINAARVARLQDRYGKH 57
Nt294440430       ---MERKHLCAALLLWAIVYFVLPVSSDNLLRVGLKKQSLDVNSINAARVARLQDRYGKN 57
Gm351725345       ---MGQKHLVTVFCLWALTCSLLPSFSFGILRIGLKKRPLDLDSINAARKAREGLRSVRP 57
Vv225450440       ---MRQGVVWAAFCLWALICPLLPVYSHGSVRIGLKKRPLDFNNMRTARIAQMQGKIGGG 57
At15221141        MKIYSRTVAVSLIVSFLLCFSAFAERNDGTFRVGLKKLKLDSKNRLAARVESKQEKP--- 57
Pt224144963       ---------------MVLHDIIIVSFILAAYLVYFVH------------------------ 22
Al297840891       ------MELRRKLCIVVAVFVIVNEFASGNFVFKVQHKFAGKEKKLEHFKSHDTRRHSRM 54
CcAP1             ---------------MLAALDMPLGGNGSPTDA---------------------------- 18
                                 .

                          70        80        90        100       110       120
                  ....|....|....|....|....|....|....|....|....|....|....|....|
CaAP2             RKEMDNNLG--SSNEDILPLKNYLDAQYYGEIGIGTPPQKFTVIFDTGSSNLWVPSAKCY 115
CcAP2             RKEMDNNLG--SSNEDILPLKNYLDAQYYGEIGIGTPPQKFTVIFDTGSSNLWVPSAKCY 113
St50540937        VNGIEKKSS--DSDIDIVPLKNYLDAQYYGEIGIGSPPQKFKVIFDTGSSNLWVPSSKCY 115
Sl350535356       VNGIEKKSS--DSDIYKVPLKNYLDAQYYGEIGIGSPPQKFKVIFDTGSSNLWVPSSKCY 115
St82623417        VNGIEKKSS--DSDIDIVPLKNYLDAQYYGEIGIGSPPQKFKVIFDTGSSNLWVPSSKCY 115
Capsicum171854659 LNGLEKKSD--GSDVDIVPLKNYLDAQYYGEIGIGSPPQKFKVIFDTGSSNLWVPSSRCY 115
Nt294440430       VNGIEKKLG--DSDLDIVSLKNYLDAQYYGEIGVGSPPQKFKVIFDTGSSNLWVPSSRCY 115
Gm351725345       MMGAHDQFIGKSKGEDIVPLKNYLDAQYFGEIGIGIPPQPFTVVFDTGSSNLWVPSSKCY 117
Vv225450440       VMSKYHGFD--DPDGEFVSLKNYLDAQYFGEIGIGTPPQNFTVVFDTGSSNLWVPSSKCY 115
At15221141        -LRAYRLGD--SGDADVVVLKNYLDAQYYGEIAIGTPPQKFTVVFDTGSSNLWVPSSKCY 114
Pt224144963       ----------------------WLSLYFAKIGLGNPSKDYYVQVDTGSDILWVN---CI 56
Al297840891       LASIDLPLG--------GDSRVDSVGLYFTKIKLGSPPKEYHVQVDTGSDILWVN---CK 103
CcAP1             -------------------------ALYFTKLSIG-PPQDYYYQVDIGSDILWVVN--CA 50
                                         . *: :: :* *.: :   .* **. ***    *

                          130       140       140       160       170       180
                  ....|....|....|....|....|....|....|....|....|....|....|....|
CaAP2             FSIACWLHSKYKAKKSSTYTAIGKSCSIRYGSGSISGFSSQDNVEVGDLVVKDQVFIEAS 175
CcAP2             FSIACWLHSKYKAKKSSTYTAIGKSCSIRYGSGSISGFSSQDNVEVGDLVVKDQVFIEAS 173
St50540937        FSIACWIH------------RDGESCSIRYETGSISGHFSMDNVQVGDLVVKDQVFIEAT 163
Sl350535356       FSIACWIHSKYQASKSSTYTRDGESCSIRYGTGSISGHFSMDNVQVGDLVVKDQVFIEAT 175
St82623417        FSIACWIHSKYKASKSSTYTRDGESCSIRYGTGSISGHFSMDNVQVGDLVVKDQVFIEAT 175
Capsicum171854659 FSIACWFHHKYKAGKSSTYTRNGKSCSIRYGTGSISGHFSQDNVQVGDLVVKDQVFIEAT 175
Nt294440430       FSIACWFHSKYKASKSTTYTRNGESCSIRYGTGSISGHFSQDNVQVGDLVVKDQVFIEAT 175
Gm351725345       FTLACYTHNWYTAKKSKTHVKNGTSCKINYGTGSISGFFSQDNVKVGSAVVKHQDFIEAT 177
Vv225450440       FSIACFFHNKYKARLSSTYTKIGRPGEIHYGSGSISGFFSQDNVEVGSLVVKDQVFIEAT 175
At15221141        FSLACLLHPKYKSSRSSTYEKNGKAAAIHYGTGAIAGFFSNDAVTVGDLVVKDQEFIEAT 174
Pt224144963       GCDKCPTKSDLGIKLTLYDPASSVSATRVSCDDDFCTSTYNGLLPD------------CK 104
Al297840891       PCPECPSKTNLNFHLSLFDVNASSTSKKVGCDDDFCS--FISQSDS------------CQ 149
CcAP1             GCVRCPKKSSLGIDLTLYDMKASSTGRLVTCDQDFCLSAFNAPASD------------CK 98
                      *  :              . .         :.                      .

                          190       200       210       220       230       240
                  ....|....|....|....|....|....|....|....|....|....|....|....|
CaAP2             REGSLTFVIAKFDGILGLGFQEIAVDNMVPVWYNMVDQGLVDEQVFSFWLNRDPNAEDGG 235
CcAP2             REGSLTFVIAKFDGILGLGFQEIAVDNMVPVWYNMVDQGLVDEQVFSFWLNRDPNAEDGG 233
St50540937        REPSITFIVAKFDGILGLGFQEISVGNTTPVWYNMVGQGLVKEPVFSFWFNRDANAKEGG 223
Sl350535356       REPSITFIVAKFDGILGLGFQEISVGNTTPVWYNMVGQGLVKEPVFSFWFNRDANAKEGG 235
```

```
St82623417            REPSITFIVAKFDGILGLGFQEISVGNTTPVWYNMVGQGLVKESVFSFWFNRDANAKEGG 235
Capsicum171854659     REPSITFIIGKFDGILGLGFQEISVGNATPVWYNMVDQGLVKEPVFSFWFNRDASTKEGG 235
Nt294440430           REPSITFIIAKFDGILGLGFQEISVGNATPVWYNMVGQGLVKEQVFSFWINRDATAKEGG 235
Gm351725345           HEGSLTFLSAKFDGILGLGFQEISVENAVPVWFKMVEQKLISEKVFSFWLNGDPNAKKGG 237
Vv225450440           REGSLTFALAKFDGIMGLGFQGISVGNATPVWYNMLQQGLLHEELFSFWLNRNPNANEGG 235
At15221141            KEPGITFVVAKFDGILGLGFQEISVGKAAPVWYNMLKQGLIKEPVFSFWLNRNADEEEGG 234
Pt224144963           KELPCQYNVVYGDGSSTAGYFVSDAVQFERVTGNLQTGLSNGTVTFGCGAQQSGGLGTSG 164
Al297840891           PAVGCSYHIVYADESTSEGNFIRDKLTLEQVTGDLQTGPLGQEVVFGCGSDQSGQLGKSD 209
CcAP1                 VGNPCAYSVTYGDGSSTGGYFVRDYAKLNQLTGNLQTIPMNGSIVFGGSSQQSGELGSST 158
                            :     *     *           :  .:          *.   : .     .

                             250       260       270       280       290       300
                             ....|....|....|....|....|....|....|....|....|....|....|
CaAP2                 ELVFGGVDTNHFKGKHTYVPVTQKGYWQFKMGDFLIGNVSTGFCEGGCAAIVDSGTSLLA 295
CcAP2                 ELVFGGVDTNHFKGKHTYVPVTQKGYWQFKMGDFLIGNVSTGFCEGGCAAIVDSGTSLLA 293
St50540937            ELVFGGVDPKHFKGNHTYVPLTQKGYWQFNMGDFLIGNTSTGYCAGGCAAIVDSGTSLLA 283
Sl350535356           ELVFGGVDPKHFKGNHTCVPLTQKGYWQFNMGDFLIGNTSTGYCAGGCAAIVDSGTSLLA 295
St82623417            ELVFGGVDPKHFKGNHTYVPLTQKGYWQFNMGDFLIGNTSTGYCAGGCAAIVDSGTSLLA 295
Capsicum171854659     ELVFGGVDPKHFKGNHTYVPLTQKGYWQFNMGDFLIGNTSTGYCAGGCAAIVDSGTSLLA 295
Nt294440430           ELVFGGVDSNHFKGNHTYVPLTQKGYWQFNMGDFLIGNASTGVCAGGCAAIVDSGTSLLA 295
Gm351725345           ELVFGGVDPKHFKGNHTYVPITEKGYWQIEMGDFFVGGVSTGVCEGGCAAIVDSGTSLLA 297
Vv225450440           EIVFGGVDKRHFRGKHTFVPVTQAGYWQFRMGDFLISNQTTGVCEGGCSAIVDSGTSLIA 295
At15221141            ELVFGGVDPNHFKGKHTYVPVTQKGYWQFDMGDVLIGGAPTGFCESGCSAIADSGTSLLA 294
Pt224144963           EALDGILG----------------------AFAHCLDNVNGGGIFAIGELVSPKVNTTPMV 203
Al297840891           SAVDGVMGFGQSN-TSVLSQLAATGDAKRVFSHCLDNVKGGGIFAVGVVDSPKVKTTPMV 268
CcAP1                 EAVDGIIG----------------------FGQANS------------------------ 172
                      . : * :.                      :..

                             310       320       330       340       350       360
                             ....|....|....|....|....|....|....|....|....|....|....|
CaAP2                 GPTTVVTQINHAIGAEGVVSTECKEIVSQYGELIWDLLVSGVLPDRVCKQAGLCPLRGAQ 355
CcAP2                 GPTTVVTQINHAIGAEGVVSTECKEIVSQYGELIWDLLVSGVLPDRVCKQAGLCPLRGAQ 353
St50540937            GPTTIVAQINHAIGAEGIVSMECKTIVSQYGEMIWDLLVSGVRPDQVCSQAGLCFVDGAQ 343
Sl350535356           GPTTIVTQINHAIGAEGIVSMECKTIVSQYGEMIWDLLVSGIRPDQVCSQAGLCFLDGSQ 355
St82623417            GPTTIVTQINHAIGAEGIVSMECKTIVSQYGEMIWDLLVSGVRPDQVCSQAGLCFVDGAQ 355
Capsicum171854659     GPTTIVTQLNHAIGAEGVVSAECKTIVSQYGEVLWDLLVSGVRPDQVCSQAGLCFFNGAE 355
Nt294440430           GPTTVVTQINHAIGAEGIVSMECKTIVSQYGEMIWNLLVSGVKPDQVCSQAGLCYFNGAQ 355
Gm351725345           GPTPVVAEINHAIGAEGVLSVECKEVVSQYGELIWDLLVSGVKPDDICSQVGLCSSKRHQ 357
Vv225450440           GPTLVVTQINHAIGAEGIVSMECKEVVSQYGNMMWDLLVSGVLPSKVCSQIGLCMAS-VL 354
At15221141            GPTTIITMINHAIGAAGVVSQQCKTVVDQYGQTILDLLLSETQPKKICSQIGLCTFDGTR 354
Pt224144963           PNQAHYNVYMKEIEVGGTV-LELPTDVFDSGDRRGTIIDSGTTLAYLPEVVYDSMMNEIR 262
Al297840891           PNQMHYNVMLMGMDVDGTA-LDLPPSIMRNG---GTIVDSGTTLAYFPKVLYDSLIETIL 324
CcAP1                 ----------------------------------SIISQLASAGKVKKIF--SHMLKII 195
                                                       :: .      . .    .

                             370       380       390       400       410       420
                             ....|....|....|....|....|....|....|....|....|....|....|
CaAP2                 HENAYIKSVVDEENKEEASVG-ESPMCTACEMAVVWMQNQLKQQGTKEKVLAYVNQLCES 414
CcAP2                 HENAYIKSVVDEENKEEASGG-ESPMCTACEMAVVWMQNQLKQQGTKEKVLAYVNQLCES 412
St50540937            HVSSNIKTVVERET-EGSSVG-EAPLCTACEMAVVWMQNQLKQEGTKEKVLEYVNQLCEK 401
Sl350535356           HVSSNIRTVVERET-EGSSVG-EAPLCTACEMAVVWMQNQLKQEQTKEKVLEYVNQLCEK 413
St82623417            HVSSNIRTVVERET-EGSSVG-EAPLCTACEMAVVWMQNQLKQAGTKEKVLEYVNQLCEK 413
Capsicum171854659     HVSSNIRTVVEREN-EGSSVG-EAPLCTVCEMAVVWIQNQLKQQGTKERVLEYVDQLCEK 413
Nt294440430           HVSSNIRTVVERET-EGSSVG-EAPLCTACEMAVVWMQNQLKQKETKERVLEYVNQLCEK 413
Gm351725345           SKSAGIEMVTEKEQ-EELAAR-DTPLCSSCQMLVLWIQNQLKQKATKDRVFNYVNQLCES 415
Vv225450440           WCSPGIRTVVEKEKMESVEEVGDVVFCNACEMIAVWIQSQLKQMKTKDKVLRYVTELCGS 414
At15221141            GVSMGIESVVDKEN-AKLSNGVGDAACSACEMAVVWIQSQLRQNMTQERILNYVNELCER 413
Pt224144963           SQQPGLSLHTVEEQFICFKYS-GNVDDGFPDIKFHFKDSLTLTVYPHDYLFQISEDIWCF 321
Al297840891           ARQP-VKLHIVEDTFQCFSFS-ENVDVAFPPVSFEFEDSVKLTVYPHDYLFTLEKELYCF 382
CcAP1                 ASQSNLKIHIVENQFKCFVYS-GNVD-GFPVVXFEFEDSLSLTVYPHEYLFDLHDDQWCI 253
                        .  :     .:                  :   : :.      .:: ::    :

                             430       440       450       460       470       480
                             ....|....|....|....|....|....|....|....|....|....|....|
CaAP2                 IPSPMGESIIDCNSLSTLPNVSFTIGGKSFELTPKQYVLQTGEGFAEVCISGFMAMDVPP 474
CcAP2                 IPSPMGESIIDCNSLSTLPNVSFTIGGKSFELTLKEYVLRTGEGFAEVCISGFMAMDVPP 472
St50540937            IPSPMGESAIDCNNISSMPDITFTIKDKAFVLTPEQYILKTGEGVATICVSGFAALDVPP 461
Sl350535356           IPSPMGESAIDCNRISSMPDITFTIKDTAFVLTPEQYILKTGEGVATICVSGFAALDVPP 473
St82623417            IPSPMGESTIDCNSISSMPDISFTIKDKAFVLTPEQYILKTGEGVATICVSGFAALDVPP 473
Capsicum171854659     LPSPMGESVVDCNSISSLPNITFTIKDKAFVLTPEQYILKTGEGIASICISGFAAFDVPP 473
Nt294440430           LPSPMGESVIDCSMISAMPNITFTIKDKAYVLTPEQYILKTGEGITTICMSGFAALDVPP 473
Gm351725345           LPSPSGESVISCNSLSKMPNITFTIGNKPFVLTPEQYILRTGEGITEVCLSGFIAFDVPP 475
Vv225450440           LPSPMGESVIDCTSVANMPNITFIIGDKAFDLTPDQYILRTGDGSATVCLSGFTALDVPP 474
At15221141            LPSPMGESAVDCAQLSTMPTVSLTIGGKVFDLAPEEYVLKVGEGPVAQCISGFIALDVAP 473
Pt224144963           GWQNGGMQSKDGRDMTLLGDLVLSNKLVLYDIENQAIGWTEYNCKYHLLFL--------- 372
Al297840891           GWQAGGLTTGERTEVILLGDLVLSNKLVVYDLENEVIGWADHNCSSSIKIKDG-SGGVYS 441
CcAP1                 GWQNKGMQIRDGREVTLLCDLVLANKLVSYDLENQTIGWAEYNCSSSIKLRDEKSGNVYA 313
                        .  *    .   :  :  : :      : :  .       :      .
```

```
                              490       500       510
                     ....|....|....|....|....|....|....|.
CaAP2                PRGPIWVLGDVFMGVYHTVFDYGNLRMGFAKAA--- 507
CcAP2                PRGPIWVLGDVFMGVYHTVFDYGNLRMGFARAA--- 505
St50540937           PRGPLWILGDVFMGPYHTVFDYGKSQVGFAEVHNSL 497
Sl350535356          PRGPLWILGDVFMGPYHTVFDYGKSQVGFAEAA--- 506
St82623417           PRGPLWILGDVFMGPYHTVFDYGKSQVGFAEAA--- 506
Capsicum171854659    PRGPLWILGDVFMGPYHTVFDYGNSQVGFAEAA--- 506
Nt294440430          PRGPLWILGDVFMGVYHTVFDYGNSRLGFAEAV--- 506
Gm351725345          PKGPLWILGDVFMRAYHTVFDYGNLQVGFAEAV--- 508
Vv225450440          PKGPLWILGEIFMGVYHTVFDFGDLRIGFAEAA--- 507
At15221141           PRGPLWILGDVFMGKYHTVFDFGNEQVGFAEAA--- 506
Pt224144963          -------------------------------------
Al297840891          VGADNLSSAPPL---LMITKLLTILSPLIAVALLH- 473
CcAP1                VGSEIISSARGLNAGKALRFLLLIITSLLHALLIP- 348
```

Figure 6. Multiple alignment of the deduced amino acid sequences of the *Coffea arabica* aspartic proteinase. Protein domains (pfam – http://pfam.sanger.ac.uk/) are indicated by highlights as follows: □ = eukaryotic aspartyl proteinase domain (pfam – PF00026).

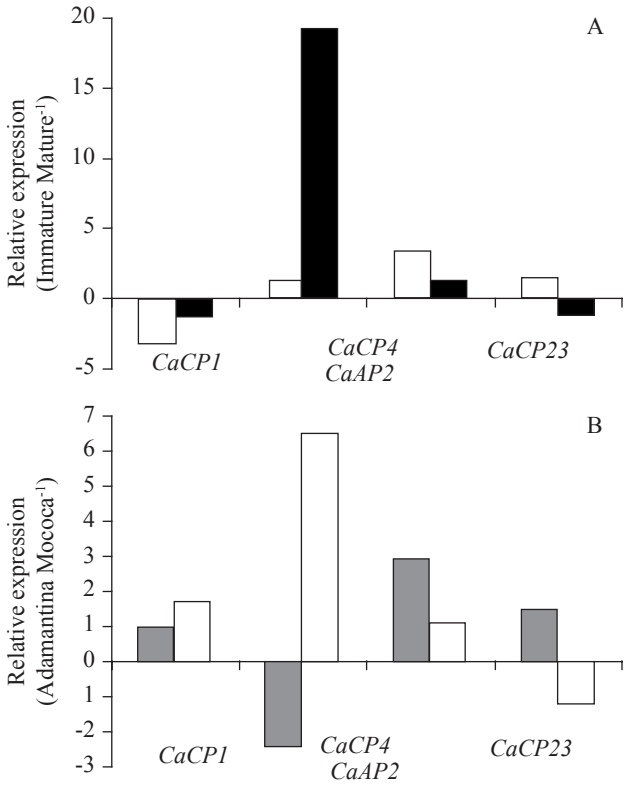


Figure 7. Relative expression analysis of cysteine (*CaCP1*, *CaCP4* and *CaCP23*) and aspartic (*CaAP2*) proteinases between endosperms from immature and mature coffee fruits (A) and between Adamantine and Mococa (B). The data shown are the means of two biological replicates, each composed of three technical replicates. In (A) ■ = Mococa, and in (B) □ = mature fruits.

Vincent 1985), and significant metabolic changes may occur in the living seed during the processing steps (Bytof et al. 2000, Knopp et al. 2006). Thus, genetic factors, management techniques and post-harvesting processing all contribute to the complex changes in the biochemical composition of coffee seeds. The composition and concentrations of the compounds are then further chemically changed during roasting, which causes many chemical interactions among compounds, such as the Maillard reaction among N compounds (amino acids and proteins) and sugars (Leloup et al. 1995, Montavon et al. 2003, Bertrand et al. 2012).

In addition to these factors, the environmental conditions in which coffee trees are grown have an impact on coffee beverage quality (Decazy et al. 2003, Silva et al. 2005, Bertrand et al. 2012). In Brazil, the time required for coffee fruit maturation changes drastically, differing by up to two months depending on the environment in which coffee is grown (Ortolani et al. 2000, 2001, Silva et al. 2005). Maturation is slower in regions with lower temperatures, and this property has been related to higher coffee quality (Ortolani et al. 2000, Bertrand et al. 2012). Although never tested, it has been proposed that slower maturation provides the time required for the biosynthesis of the adequate proportions of all compounds involved in coffee quality (Silva et al. 2005). Similarly, it has been suggested that coffee trees cultivated under the shadows of other trees undergo slower and more uniform ripening, which allows the fruits to be developed in a more balanced manner and leads to better coffee quality (Muschler 2001). For both of the sites selected in this study, the difference in fruit ripening time is at least one month but can be greater depending on weather variations (Ortolani et al. 2000, 2001). Adamantina produces lower-quality coffee when compared with Mococa (Ortolani et al. 2000, 2001, Silva et al. 2005).

The mean temperatures in Adamantina and Mococa during 2004 and 2005 showed a marked difference that was even more consistent when the 15-year average was calculated (1992 to 2007). This temperature difference is likely related to altitude, as Adamantina is located at 443 m and Mococa is located at 663 m. Beans from coffee

trees grown at higher altitudes in Honduras produced a beverage with superior quality (Decazy et al. 2003), which could be explained by the inverse relationship between altitude and temperature.

The annual precipitation (rainfall) in 2004 and 2005 was 1377 and 1171 mm in Adamantina and 1892 and 1389 mm in Mococa, respectively. Both locations have a dry winter season, with nearly no precipitation. For this reason, the coffee plants were kept under irrigation. However, plants from the same plots in Adamantina and Mococa were used in a previous study (Silva et al. 2005) that showed that temperature had greater influence than water availability on beverage quality. This study also confirmed the findings of previous studies (Ortolani et al. 2000, 2001) that showed that coffee beans from Mococa were of higher quality than those from Adamantina. Moreover, the using main component analysis showed a large difference in the amount of nitrogen compounds and proteinase activity between these two sites (Silva et al. 2005).

Free amino acids are among the most important compounds in defining the final quality of coffee (Macrae 1985). Amino acids react with reducing sugars in the Maillard reaction during roasting to produce compounds related to aroma and flavour (Arnoldi 2004). This study showed that the highest concentration of free amino acids occurred in immature coffee samples from Adamantina; the levels of free amino acids were reduced such that the contents in the mature endosperm were similar in seeds from the two sites tested. Silva et al. (2005) used the same experimental coffee plots in Mococa and Adamantina that we used here and also did not find any difference in the free amino acid contents of seeds from mature fruits (sun dry-processed). We suggest that the lower amino acid content in endosperm from mature fruits may simply indicate a decrease in the metabolic rate that reflects a more advanced developmental stage (Sano *&* Kawashima 1982, Emmanuel 1983). However, the higher content in immature endosperms from Adamantina may indicate a faster metabolism, although the fruits collected from both sites had endosperm of a similar size (dry weight mass).

The analysis of the amino acid profile showed marked differences between the samples from Adamantina and Mococa. However, for most of the amino acids, individual variations are difficult to explain in the light of the knowledge available on the physiology of coffee maturation. The available information is usually related to alterations that occur during roasting. Thus, depending on the roasting conditions, Asn and Gln may undergo spontaneous deamination as free amino acids or as protein residues (Reineccius 1995). Asn is found in higher amounts in samples from Adamantina. In the presence of heat, e.g., during the roasting process, Asn and Gln deamination leads to the formation of Glu and Asp to release ammonia. This free, reactive ammonia may undergo a series of reactions that result in its combination with products from sugar degradation and lead to the formation of pyrazines (Ho et al. 1993), which are among the compounds responsible for the roasted coffee aroma. However, high amounts of pyrazines interfere with beverage flavour (Ho et al. 1992). The aroma produced by pyrazines is complex and depends on their concentration and also on the influence of synergic and antagonistic effects (Dart *&* Nursten 1985). Thus, our data suggest that the increase in the amount of Asn in samples from Adamantina compared with the amount found in Mococa could lead to pyrazine production and interfere with the quality of the beverage produced by coffee from this region.

Although it was not among the most abundant amino acids and was present at similar levels in immature endosperm from both sites, GABA showed a marked increase in the endosperm of mature fruits from Adamantina. This non-protein amino acid has been studied largely because it accumulates in plant tissues under biotic and abiotic stress conditions (Shelp et al. 1999). Coffee endosperm accumulates GABA during dry processing because of the decrease in water content (Bytof et al. 2004). Kramer et al. (2010) noted complex changes in GABA levels after harvesting and during dry-processing; these changes are related to dehydrin gene expression, which is associated with water stress in coffee beans. These data are supported by the finding that heat seems to cause a considerable increase in the concentration of GABA in plants relative to other types of stress (Shelp et al. 1999, Kinnersley *&* Turano 2000).

Soluble protein analysis indicated a lack of differences between immature endosperms, but protein accumulation was significantly higher in the endosperm of mature fruits from Mococa. Coffee from the Mococa region produces a better beverage than that from Adamantina (Ortolani et al. 2000, 2001, Silva et al. 2005), and one hypothesis for this observation, as suggested by Silva et al. (2005), is that the higher proteinase activity in fruits from Adamantina breaks down proteins that are important for the development of coffee aroma and flavour.

Proteinase assays demonstrated higher activity in endosperm from immature fruits when compared with mature fruits, which partially explains the higher amount of free amino acids present in immature endosperms. It

might be expected that high proteinase activity should be negatively correlated with protein content and vice versa, as observed in endosperm from Mococa. However, although the proteinase activity and amino acid content of endosperms from Adamantina fit this relationship, immature and mature endosperm showed similar contents. Silva et al. (2005) also observed an inverse correlation between proteinase activity and protein content in dry-processed beans from Mococa and Adamantina but did not analyse immature fruits. We did not perform bi-dimensional electrophoresis, but it is possible that the protein profiles of samples from each site were different despite the similar protein contents of the samples.

Endoproteinases are the first proteinases to act on proteins, producing smaller peptides that are then hydrolysed to free amino acids by exopeptidases (Callis 1995). Coffee protein extracts were analysed to characterise endoproteinase classes using polyacrylamide gels co-polymerised with BSA and then incubated with specific inhibitors for each proteinase type: pepstatin A for aspartic proteinases, PMSF for serine proteinases, EDTA for metalloproteinases and iodoacetamide for cysteine proteinases. Enzymatic activity was characterised by the presence of negative bands (no colour development), which indicates that the co-polymerised BSA was degraded by enzymatic action. Consistent with the spectrophotometry results, the proteinase activity in the endosperm of mature coffee fruits was very low and could not be properly detected by electrophoresis. Thus, characterisation was performed in extracts from immature endosperms.

In gel activity assays identified cysteine proteinases to be the predominant type of endopeptidase in coffee endosperm, but serine proteinases were also detected. In contrast, although we detected *in gel* serine proteinase activity, no serine proteinase sequence was identified in our search of the CafEST database. However, this search was performed before EST data for *C. canephora* fruits (Leroy et al. 2005) had been incorporated into the CafEST (Mondego et al. 2011). Thus, we did not expect to find homologous proteinase genes from *C. canephora* (McCarthy et al. 2007, Lepelley et al. 2012).

The CafEST database search for proteinases returned a large number of sequences (reads), which were grouped by similarity to form contigs. Contigs with a high proportion of reads from fruit cDNA libraries were chosen for further studies. We identified a massive predominance of cysteine proteinases, followed by aspartic proteinases; the CafEST database yielded only one contig for serine proteinase that was formed only by two sequences (reads) from a leaf cDNA library. The analysis of ESTs from *C. arabica* and *C. canephora* together showed that the most widely expressed contigs in *C. arabica* encode a papain-like cysteine proteinase (Mondego et al. 2011). However, these data should be interpreted cautiously because we selected only contigs generated from fruit cDNA libraries. One possible reason for the lack of correspondence between the *in gel* characterisation and the absence of serine proteinase genes is that the fruits collected here and those used to produce the cDNA libraries in the coffee EST database may not have been at the same stage of development. Additionally, the CafEST included only two libraries generated from fruits of *C. arabica*, and the stage at which these fruits were collected was not specified (Vieira et al. 2006). In wheat grains, the appearance of different proteinase activities is controlled temporally; serine proteinases are detected at early stages, whereas aspartic proteinases and metalloproteinases are present at later stages (Dominguez *&* Cejudo 1996). Furthermore, proteinase inhibition may be substrate-dependent, as observed in many microorganisms (Coradi *&* Guimarães 2006), and metalloproteinase activity may have been observed if substrates other than BSA had been tested. However, based on the activity characterisation and the number of contigs encoding cysteine proteinases, it is evident that this class of endopeptidases is the most abundant in coffee endosperm.

Our search for proteinase genes in fruit cDNA libraries revealed four sequences, and a comparison with sequences deposited on the NCBI database indicated that these four sequences corresponded to three cysteine proteases (*CaCP1*, *CaCP4* and *CaCP23*) and one aspartic proteinase (*CaAP2*). Two cysteine proteinase genes (*CcCP-1* and *CcCP-4*) and an aspartic proteinase gene (*CcAP-2*) isolated from *C. canephora* (McCarthy et al. 2007, Lepelley et al. 2012) encode proteins with 97 to 99% similarity to the two cysteine proteinases and aspartic proteinase of *C. arabica* identified here, which strongly suggests that these genes are orthologous. For this reason, the orthologous genes from *C. arabica* (*CaCP1, CaCP4* and *CaAP2*) were named after the *C. canephora* genes.

To better understand the relationships among the proteinases studied, we built a phylogenetic tree that included the predicted amino acid sequences of coffee cysteine proteinases and similar sequences deposited in the NCBI protein database. This tree allowed us to categorise each gene that we identified into well-defined subgroups and showed that proteinases containing domains from the cysteine family, particularly from the C1A subfamily (papain proteinases), contained

differences in other regions that enabled their classification into distinct groups. The alignment of the predicted amino acid sequences of the identified contigs with some complete sequences that had similarity to coffee proteinases sequences allowed us to better visualise these different regions.

The papain family includes enzymes with a large range of activities, including large-spectrum endoproteinases such as papain, aminopeptidases, dipeptide peptidases and enzymes with exo- and endopeptidase activities (Rawlings et al. 2008). The phylogenetic tree clearly shows that a single plant species may express proteinases from different subgroups [e.g., *Phaseolus vulgaris* (groups 1 and 3), *Vitis vinifera* (groups 2 and 3), *Lycopersicum esculentum* (groups 2 and 4) and *Glycine max* (groups 1 and 2)]. In *G. max*, for instance, the sequences grouped with the *Ca*CP1 and *Ca*CP4 proteinases present very dissimilar functions. The *G. max* sequences homologous to *Ca*CP1 are induced by water stress (Nong et al. 1995), whereas the sequence homologous to *Ca*CP4 is related to senescence (Ling et al. 2003).

*Ca*CP23 belongs to a group that contains the granulin domain, which is not found in the other groups). This domain may have several biological functions, is most likely activated by post-translational proteolytic processing and is most often found in metazoans (Bhandari et al. 1992). In plants, this domain is usually related to the C-terminal region of cysteine proteinases belonging to the papain subfamily. The gene *tdi-65*, isolated from the tomato, is phylogenetically related to *CaCP23* and is induced by water stress but not by ABA, which suggests a relationship with senescence and cell death (Harrak et al. 2001).

The subgroup that includes the protein *Ca*CP4 (subgroup 1, figure 4A) presents a conserved KDEL sequence at the C-terminus, which controls the transport of the protein from the endoplasmic reticulum to protein storage vacuoles (Toyooka et al. 2000). This same subgroup contains some previously characterised genes, such as *SH-EP* (sulphydryl endoproteinase), which encodes a cysteine proteinase (Vm445927) that is involved in the degradation of storage globulin in germinating seeds and in the early development of ripening stage fruits in *Vigna mungo* (Yamauchi et al. 1992). This finding is particularly interesting because the 11S globulin is the main storage protein in coffee seeds (Acuña et al. 1999, Marraccini et al. 1999, Rogers et al. 1999). The electrophoretic profile of protein extracts shows that this storage protein is predominant in several cultivars and coffee species (Baú et al. 2001).

Although we studied only four contigs, i.e., those with the highest proportion of reads from cDNA libraries, our qPCR expression analysis showed good agreement with the proteinase activity assays because the highest gene expression was observed in endosperm from green fruits. *CaCP4* was the most highly expressed gene in the immature endosperm and in fruits from Mococa. *CaCP1* was not expressed at similar levels in Adamantina and Mococa, and it was expressed at higher levels in mature endosperm than in immature endosperm. *CaCP23*, which has not been described in previous reports, was most highly expressed in immature endosperms from Adamantina. *CaCP4* was the most differentially expressed gene. Lepelley et al. (2012) observed that *CcCP1* was more highly expressed than its orthologue *CcCP4* throughout the fruit maturation process in *Coffea canephora*, in contrast to our data, with highest expression of both genes at the mature stage (red fruits). Our *in gel* activity also showed that the protease activity was highest in immature endosperm. These differences may reflect differences between the coffee species (*C. arabica* and *C. canephora*) used in these studies as well as the regions in which the coffee trees were grown. The fruits used by Lepelley et al. (2012) were harvested from plants in the field in Ecuador, and no further information was given.

The *CaAP2* gene is orthologous to the *CcAP2* gene from *C. canephora*, and its expression level was similar in coffee roots, pericarp and endosperm, but its expression in the mature fruit pericarp was higher than that in other tissues and developmental stages (McCarthy et al. 2007). *CaAP2* and *CcAP2* belong to the peptidase A1 family, and some of their homologs have been reported as responsive to wounding or pathogen infection and are expressed in different tissues such as leaves, roots and seeds (Asakura et al. 1995, Schaller *&* Ryan 1996, Terauchi et al. 2004, Guevara et al. 2005, Huang et al. 2009). *CaAP2* was not significantly differentially expressed between mature and immature endosperm or between plants grown at different sites.

It is clear from our and others' results that comparisons between gene expression and enzymatic activity for coffee proteinases must be analysed carefully because some proteinase genes may not have been identified, whereas activity data reflect the activity of all expressed genes. Nevertheless, the joint analysis of enzymatic activity data and gene expression demonstrates that the activities of these proteinases are quite complex. The finding that some of the proteinase genes identified in coffee are similar to proteinase genes that are modulated by stress in other plants suggests that further studies

may define the role of each enzyme according to the environmental variations that occur during each phase of coffee fruit development and maturation. It is important to consider that immature fruits were harvested in January (Adamantina) and March (Mococa), which are the months in which the highest temperatures were registered and thus may lead to heat stress conditions. Therefore, our and others' results (Decazy et al. 2003, Lepelley et al. 2012) suggest that the increased activity and proteinase expression during these months may be related to heat stress because the plants were irrigated and thus were not water-stressed.

Environmental factors have a strong influence on coffee beverage quality by changing the amino acid and protein profile, which is an important factor in coffee quality. Proteinase activity is higher in coffee beans from warmer places, which can change the protein and amino acid composition and thus change the quality. Therefore, a relationship seems to exist among temperature, proteinase activity and coffee quality. This information, together with other data on temperature and coffee composition, helps to explain why coffee from regions with low temperatures has higher quality.

It is believed that the high quality of coffee produced in colder climates is related to the time required for complete maturation, which enables the biosynthesis of substances that are important for the development of aroma and flavour. However, there is no concrete proof for this assumption, and the different composition of coffee beans from locations with different mean temperatures is the only and indirect evidence supporting this theory. In this work, we showed that temperature modulates proteinase activity and expression in endosperm of coffee fruits, which leads to variations in the amounts and composition of amino acids and proteins. More detailed proteomics studies will show whether differences in proteinase activity among locations significantly alter the protein profile of coffee endosperm or the types of protein cleaved by these enzymes, and whether the amino acid compositions of the peptides generated would lead them to take part in the Maillard reaction and consequently alter the beverage quality of coffee.

Acknowledgements – Hellen M.C. Abreu & Paula Y. Yamamoto received fellowships from Capes-Brazil; Paula M. Nobile & Milton M. Shimizu received post-doctoral fellowships from Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café; and P. Mazzafera & Carlos A. Colombo received researcher fellowships from CNPq-Brazil (CNPq). This work was sponsored by the Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café.

REFERENCES

Acuña R, Bassuner R, Beilinson V, Cortina H, Cadena-Gómez G, Montes V, Montes V, Nielsen NC. 1999. Coffee seeds contain 11S storage proteins. Physiologia Plantarum 105:122-131.

Alègre C. 1959. Climats et caféiers d'Arabie. Agronomy Tropical 14:23-58.

Arnoldi A. 2004. Factors affecting the Maillard reaction. *In* Understanding and measuring the shelf-life of food (R Steele, ed.). Woodhead Publishing, Cambridge, p.111-127.

Asakura T, Watanabe H, Abe K, Arai S. 1995. Rice aspartic proteinase, oryzasin, expressed during seed ripening and germination, has a gene organization distinct from those of animal and microbial aspartic proteinases. European Journal of Biochemistry 232:77-83.

Barsalobres-Cavallari C, Severino F, Maluf M, Maia I. 2009. Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. BMC Molecular Biology 10:1.

Baú SMT, Mazzafera P, Santoro LG. 2001. Seed storage proteins in coffee. Brazilian Journal of Plant Physiology 13:31-40.

Bertrand B, Boulanger R, Dussert S, Ribeyre F, Berthiot L, Descroix F, Joët T. 2012. Climatic factors directly impact the volatile organic compound fingerprint in green Arabica coffee bean as well as coffee beverage quality. Food Chemistry 135:2575-2583.

Bhandari V, Palfree RG, Bateman A. 1992. Isolation and sequence of the granulin precursor cDNA from human bone marrow reveals tandem cysteine-rich granulin domains. Proccedings of the National Academy of Sciences of the United States of America 89:1715-1719.

Bielesk RL, Turner NA. 1966. Separation and estimation of amino acids in crude plant extracts by thin layer electrophoresis and chromatography. Analytical Biochemistry 17:278-282.

Bradford MM. 1976. Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Analytical Biochemistry 72:248-254.

Bytof G, Knopp SE, Schieberle P, Teutsch I, Selmar D. 2004. Influence of processing on the generation of γ-aminobutyric acid in green coffee beans. European Food Research and Technology 220:245-250.

Bytof G, Selmar D, Schieberle P. 2000. New aspects of coffee processing: how do the different post harvest treatments influence the formation of potential flavour precursors? Journal of Applied Botany 74:131-136.

Callis J. 1995. Regulation of protein degradation. Plant Cell 7:845-857.

Camargo AP, Santinato R, Cortez JG. 1992. Aptidão climática para qualidade da bebida nas principais regiões cafeeiras de Arábica no Brasil. *In* Anais do XVIII Congresso Brasileiro de Pesquisas Cafeeiras (Instituto Brasileiro do Café, ed.), Araxá, p.70-74.

Carvalho A. 1988. Principles and practices of coffee plant breeding for productivity and quality factors: *Coffea arabica*. *In* Coffee: Agronomy (RJ Clarke, R Macrae, eds.). Elsevier Applied Science, London, p.129-165.

Clarke RJ. 1985. Green coffee processing. *In* Coffee: botany, biochemistry and production of beans and beverage (MN Clifford, KC Wilson, eds.). Avi Publishing Company, Inc., Westport, Connecticut, p.230-250.

Clifford MN. 1985. Chemical and physical aspects of green coffee and coffee products. *In* Coffee: botany, biochemistry and production of beans and beverage (MN Clifford, KC Wilson, eds.). AVI Publishing Company, Inc., Westport, Connecticut, p.305-374.

Clifford MN. 1985. Chlorogenic acids. *In* Coffee: chemistry (RJ Clarke, R Macrae, eds.). Elsevier Applied Science, London, v.1, p.153-202.

Cocking EC, Yemm EW. 1954. Estimation of amino acids by ninhydrin. Analyst 80:209-213.

Coradi S, Guimarães S. 2006. *Giardia duodenalis*: protein substrates degradation by trophozoite proteases. Parasitology Research 99:131-136.

Cruz F, Kalaoun S, Nobile P, Colombo C, Almeida J, Barros L, Romano E, Grossi-de-Sá M, Vaslin M, Alves-Ferreira M. 2009. Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. Molecular Breeding 23:607-616.

Damatta FM, Ramalho JDC. 2006. Impacts of drought and temperature stress on coffee physiology and production: a review. Brazilian Journal of Plant Physiology 18:55-81.

Dart SK, Nursten HE. 1985. Volatile components. *In* Coffee: chemistry (RJ Clarke, R Macrae, eds.). Elsevier Applied Science, London, v.1, p.223-265.

Decazy F, Avelino J, Guyot B, Perriot JJ, Pineda C, Cilas C. 2003. Quality of different Honduran coffees in relation to several environments. Journal of Food Science 68: 2356-2361.

Dominguez F, Cejudo FJ. 1996. Characterization of the endoproteases appearing during wheat grain development. Plant Physiology 112:1211-1217.

Emmanuel OA. 1983. Accumulation and quality of storage protein in developing cowpea, mung bean and soya bean seeds. Journal of the Science of Food and Agriculture 34:1351-1357.

Frohmann MA, Dush MK, Martin GR. 1988. Rapid production of full-length cDNAs from rare transcripts by amplification using a single genetic specific oligonucleotide primer. Proccedings of the National Academy of Sciences of the United States of America 85:8998-9002.

Geromel C, Ferreira LP, Cavalari AA, Pereira LFP, Guerreiro SMC, Vieira LGE, Leroy T, Pot D, Mazzafera P, Marraccini P. 2006. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. Journal of Experimental Botany 57:3243-3258.

Geromel C, Ferreira LP, Davrieu F, Guyot B, Ribeyre F, Scholz MBS, Protasio Pereira LF, Vaast P, Pot D, Leroy T, Androcioli Filho A, Esteves Vieira LG, Mazzafera P, Marraccini P. 2008. Effects of shade on the development and sugar metabolism of coffee (*Coffea arabica* L.) fruits. Plant Physiology and Biochemistry 46:569-579.

Guevara MG, Almeida C, Mendieta JR, Faro CJ, Veríssimo P, Pires EV, Daleo GR. 2005. Molecular cloning of a potato leaf cDNA encoding an aspartic protease (StAsp) and its expression after *P. infestans* infection. Plant Physiology and Biochemistry 43:882-889.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Harrak H, Azelmat S, Baker EN, Tabaeizadeh Z. 2001. Isolation and characterization of a gene encoding a drought-induced cysteine protease in tomato (*Lycopersicon esculentum*). Genome 44:368-374.

Ho CT, Hwang HI, Yu TH, Zhang J. 1993. An overview of the Maillard reactions related to aroma generation in coffee. *In* Proccedings of XV International Conference on Coffee Science (ASIC, ed.). Association Scientifique Internationale du Café (ASIC), Montpellier, p.519-527.

Ho CT, Zhang Y, Oh YC, Shu CK. 1992. Peptides as flavor precursors in model Maillard reactions. ACS Symposium Series 490:193-202.

Huang G-J, Huang S-S, Chen HJ, Chang YS, Chang SJ, Chang H-Y, Hsieh P-C, Chand M-J, Lin Y-C, Lin Y-H. 2009. Cloning and expression of aspartic proteinase cDNA from sweet potato storage roots. Botanical Studies 50:149-158.

Kinnersley AM, Turano FJ. 2000. Gamma aminobutyric acid (GABA) and plant responses to stress. Critical Reviews in Plant Sciences 19:479-509.

Knopp S, Bytof G, Selmar D. 2006. Influence of processing on the content of sugars in green Arabica coffee beans. European Food Research and Technology 223: 195-201.

Kramer D, Breitenstein B, Kleinwächter M, Selmar D. 2010. Stress metabolism in green coffee beans (*Coffea arabica* L.): expression of dehydrins and accumulation of GABA during drying. Plant Cell Physiology 51: 546-553.

Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5:150-163.

Leloup V, Louvrier A, Liardon R. 1995. Degradation mechanisms of chlorogenic acids during roasting. *In* Proccedings of XVI International Scientifique Colloquium on Coffee (ASIC, ed.). Association Scientifique International du Café (ASIC), San Francisco, p.92-198.

Lepelley M, Amor M, Martineau N, Cheminade G, Caillet V, Mccarthy J. 2012. Coffee cysteine proteinases and related inhibitors with high expression during grain maturation and germination. BMC Plant Biology 12:31.

Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira L, Jourdan I, Pot D, Andrade A, Glaszmann J, Vieira L, Piffanelli P. 2005. Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes Theoretical and Applied Genetics 111: 1032-1041.

Leroy T, Ribeyre F, Bertrand BI, Charmetant P, Dufour M, Montagnon C, Marraccini P, Pot D. 2006. Genetics of coffee quality. Brazilian Journal of Plant Physiology 18:229-242.

Ling J, Kojima T, Shiraiwa M, Takahara H. 2003. Cloning of two cysteine proteinase genes, CysP1 and CysP2, from soybean cotyledons by cDNA representational difference analysis. Biochimica et Biophysica Acta 1627:129-139.

Livak KJ, Schmittgen D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCt. Methods 25:402-408.

Ludwig E, Lipke U, Raczek U, Jager A. 2000. Investigations of peptides and proteases in green coffee beans. European Food Research and Technology 211:111-116.

Macrae R. 1985. Nitrogenous Compounds. *In* Coffee: chemistry (RJ Clarke, R Macrae, eds.). Elsevier Applied Science, London, v.1, p.115-142.

Marcone MF. 2004. Composition and properties of Indonesian palm civet coffee (Kopi Luwak) and Ethiopian civet coffee. Food Research International 37:901-912.

Marraccini P, Deshayes A, Pétiard V, Rogers WJ. 1999. Molecular cloning of the complete 11S seed storage protein gene of *Coffea arabica* and promoter analysis in transgenic tobacco plants. Plant Physiology and Biochemistry 37:273-282.

Marur CJ, Sodek L, Magalhães ACN. 1994. Free amino acids in leaves of cotton plants under water deficit. Brazilian Journal of Plant Physiology 6:103-108.

Mccarthy J, Amor MB, Petiard V, Tanksley SD, Lin C. 2007. Modulation of coffee flavour precursors levels in green coffee grains. US Patent – US 2007/0022489 A1.

Mondego J, Vidal R, Carazzolle M, Tokuda E, Parizzi L, Costa G, Pereira L, Andrade A, Colombo C, Vieira L, Pereira G, Consortium BCGP. 2011. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. BMC Plant Biology 11:30.

Montavon P, Duruz E, Rumo G, Pratz G. 2003. Evolution of green coffee protein profiles with maturation and relationship to coffee cup quality. Journal of Agricultural and Food Chemistry 51:2328-2334.

Montavon P, Mauron A-F, Duruz E. 2003. Changes in green coffee protein profiles during roasting. Journal of Agricultural and Food Chemistry 51:2335-2343.

Muschler RG. 2001. Shade improves coffee quality in a sub-optimal coffee-zone of Costa Rica. Agroforestry Systems 85:131-139.

Nong VH, Becker C, Müntz K. 1995. cDNA cloning for a putative cysteine proteinase from developing seeds of soybean. Biochimica et Biophysica Acta 1261:435-438.

Ortolani AA, Cortez JG, Pedro Junior MJ, Camargo MBP, Thomaziello RA, Alfonsi RR. 2000. Clima e qualidade natural de bebida do café arábica no estado de São Paulo. *In* Anais do 1º Simpósio de Pesquisa dos Cafés do Brasil (Embrapa Café, ed.) Embrapa, Poços de Caldas, p.662-664.

Ortolani AA, Pedro Junior MJ, Camargo MBP, Cortez JG, Pallone Filho WJ. 2001. Regionalização da época de maturação e qualidade natural de bebida do café arábica no Estado de São Paulo. *In* Anais do XII Congresso Brasileiro de Agrometeorologia (Sociedade Brasileira de Agrometeorologia, ed.). Sociedade Brasileira de Agrometeorologia, Fortaleza, p.53-54.

Prado H. 2003. Solos do Brasil: gênese, morfologia, classificação levantamento e manejo. 3ª ed. Embrapa, Piracicaba.

Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ. 2008. MEROPS: the peptidase database. Nucleic Acids Research 36:320-325.

Reineccius GA. 1995. The Maillard reaction and coffee flavor. *In* Proccedings of XVI International Conference on Coffee Science (ASIC, ed.). Association Scientifique International du Café (ASIC), Paris, p.249-257.

Rogers WJ, Bézard G, Deshayes A, Meyer I, Pétiard V, Marraccini P. 1999. Biochemical and molecular characterization and expression of the 11S-type storage protein from *Coffea arabica* endosperm. Plant Physiology and Biochemistry 37:261-272.

Rogers WJ, Michaux S, Bastin M, Bucheli P. 1999. Changes to the content of sugars, sugar alcohols, myo-inositol, carboxylic acids and inorganic anions in developing grains from different varieties of robusta (*Coffea canephora*) and arabica (*C. arabica*) coffees. Plant Science 149:115-123.

Russo Junior M. 1984. Dados climáticos auxiliares para planejamento e projeto de sistemas de irrigação. Companhia Energética de São Paulo, São Paulo.

Sano M, Kawashima N. 1982. Changes in the contents of free and protein amino acids in tobacco seeds and placentae during maturation. Plant Cell Physiology 23:1245-1250.

Schaller A, Ryan CA. 1996. Molecular cloning of a tomato leaf cDNA encoding an aspartic protease, a systemic wound response protein. Plant Molecular Biology 31:1073-1077.

Shelp BJ, Bown AW, Mclean MD. 1999. Metabolism and functions of gammaaminobutyric acid. Trends in Plant Science 4:446-452.

Shimizu MM, Mazzafera P. 2000. Compositional changes of proteins and amino acids in germinating coffee seeds. Brazilian Archives of Biology and Technology 43: 259-265.

Silva EA. 2004. Influência de distintas condições edafoclimáticas e do manejo de irrigação no florescimento, produção e qualidade de bebida do café (*Coffea arabica* L.). Tese de doutorado, Universidade Estadual de Campinas, Campinas.

Silva EA, Mazzafera P, Brunini O, Sakai E, Arruda FB, Mattoso LHC, Carvalho CRL, Pires RCM. 2005. The influence of water managment and environmental conditions on the chemical composition and beverage quality of coffee beans. Brazilian Journal of Plant Physiology 17:229-238.

Staff SS. 1999. Soil taxonomy: a basic system of soil classification for making and interpreting soul surveys. 2nd ed. USDA, Natural Resources Conservation Service, Agricultural Handbook 436, US Government Printing Office, Washington DC.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony method. Molecular Biology and Evolution 28:2731-2739.

Terauchi K, Asakura T, Nishizawa N, Matsumoto I, Abe K. 2004. Characterization of the genes for two soybean aspartic proteinases and analysis of their different tissue-dependent expression. Planta 218:947-957.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins D. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 24:4876-4882.

Toyooka K, Okamoto T, Minamikawa T. 2000. Mass transport of proform of a KDEL-tailed cysteine proteinase (SH-EP) to protein storage vacuoles by endoplasmic reticulum–derived vesicle is involved in protein mobilization in germinating seeds. Journal of Cell Biology 148:453-463.

Vieira LGE, Andrade AC, Colombo CA, Moraes AHDA, Metha Â, Oliveira ACD, Labate CA, Marino CL, Monteiro-Vitorello CDB, Monte DDC, Giglioti É, Kimura ET, Romano E, Kuramae EE, Lemos EGM, Almeida ERPD, Jorge ÉC, Albuquerque ÉVS, Silva FRD, Vinecky F, Sawazak HE, Dorry HFA, Carrer H, Abreu IN, Batista JAN, Teixeira JB, Kitajima JP, Xavier KG, Lima LMD, Camargo LEAD, Pereira LFP, Coutinho LL, Lemos MVF, Romano MR, Machado MA, Costa MMDC, Sá MFGD, Goldman MHS, Ferro MIT, Tinoco MLP, Oliveira MC, Van Sluys M-A, Shimizu MM, Maluf MP, Eira MTSD, Guerreiro Filho O, Arruda P, Mazzafera P, Mariani PDSC, Oliveira RLBCD, Harakava R, Balbao SF, Tsai SM, Mauro SMZD, Santos SN, Siqueira WJ, Costa GGL, Formighier EF, Carazzolle MF, Pereira GAG. 2006. Brazilian coffee genome project: an EST-based genomic resource. Brazilian Journal of Plant Physiology 18:95-108.

Vincent JC. 1985. Green coffee processing. *In* Coffee: technology (RJ Clarke, R Macrae, eds.) Elsevier Applied Science, London, v.2, p.1-33.

Xiaoqiu H. 1992. A contig assembly program based on sensitive detection of fragment overlaps. Genomics 14:18-25.

Yamauchi D, Akasofu H, Minamikawa T. 1992. Cysteine endopeptidase from *Vigna mungo*: gene structure and expression. Plant Cell Physiol. 33:789-797.