Research Article

doi: 10.25081/jpc.2019.v47.i1.5528
Journal of Plantation Crops, 2019, 47(1):1-15

Gene expression analysis of *Coffea arabica* seeds processed under different post-harvest processing methods

Alberto Pallavicini*, Jeena Devasia[1], Martina Modonut, Paolo Edomi, Luciano Navarini[2] and Lorenzo Del Terra[2]

Università degli Studi di Trieste, Dipartimento di Biologia, Laboratorio di Genetica, Trieste -34100, Italy
[1]Central Coffee Research Institute, Coffee Board, Manasagangotri, Mysuru-570006, Karnataka, India
[2]illycaffè SpA, via Flavia 110, 34149 Trieste, Italy

(Manuscript Received: 23-02-2019, Revised: 22-03-2019, Accepted: 28-03-2019)

Abstract

The mode of coffee processing, either the wet or dry method, determines the characteristic flavour and establishes the differences in quality of the final green coffee produced. The present study focused mainly on identifying the differential gene expression in green coffee seeds of Brazilian arabica coffee *(Coffea arabica* L.) among samples prepared under three different post-harvest treatments (natural, washed and semi washed method) and grown in two different locations. Expression levels of 16 genes of interest were measured. These genes are involved in various cellular, metabolic and biochemical activities influencing levels of certain compounds, such as lipids, carbohydrates, caffeine and chlorogenic acid, associated with quality characteristics of the beverage. Microarray experiments were designed with cDNA probe sequences. Microarray data was analyzed to identify the differences in gene expression between two altitudes and between two variables: location and post-harvest treatment. Cluster analysis was carried out with samples showing similar patterns, which are characteristic to the group. With this approach, it was possible to identify the important genes in *C. arabica* seeds that have differential (increased or decreased) expression levels. It was also seen that between the location and treatments, location profoundly impacts the levels of gene expression in samples.

Keywords: Cluster analysis, *Coffea arabica,* cDNA, gene expression, microarray, post-harvest treatment

Introduction

Coffee is a preferred drink of choice for most people across the world. The taste and aroma of coffee has enthralled its connoisseurs from ages. The aroma of coffee in a freshly brewed cup is due to a series of changes and transformations from the green bean stage. Coffee aroma is determined by more than 1000 chemical constituents (Clarke, 2013). All the precursors, necessary to generate the aroma during the roasting process, are present in the green bean of coffee. However, the final quality of the cup depends upon several other factors including the post-harvest processing methods followed.

Coffee may be processed either by wet or dry methods to obtain the 'plantation/parchment' coffee and 'cherry coffee' respectively. These processes aim to remove the pericarp and the mucilage from the coffee cherry and to dry raw coffee seeds to the necessary moisture level (11±1% w/w). The wet or dry processing methods also determine the characteristic flavour causing typical difference in the quality of green coffee produced (Wintgens, 2004; Sivetz and Desrosier, 1979). In dry processing, the harvested fruits, also called coffee cherries, are dried to the necessary moisture level (sometimes up to several weeks) on drying yards, and are later hulled to obtain the coffee beans. In contrast, during wet processing, the pericarp or skin of berries are removed along with the mucilage by mechanical de-pulping followed by fermentative degradation of remaining mucilage before drying the coffee beans. An alternative method is the

*Corresponding Author: pallavic@units.it

semi-washed process, in which the fruits are de-pulped and then the beans are immediately dried without fermentation or washing steps. The dry-processed coffees have more body while the wet-processed coffees are found to have a better aroma, and have more acceptance among coffee consumers (Selmar *et al*., 2002).

The cultivation of arabica coffee at higher elevation is known to favourably influence the final quality of the beverage. The quantitative data, describing the effect of climatic conditions on chemical composition of the seed, however, is still lacking. Further, the chemical transformations, occurring during wet processing of the beans, known to affect the flavour, are also not fully understood.

In both processing methods, the freshly processed coffee beans exhibit active metabolic processes and remain viable (Bytof *et al.,* 2007). Reports are available on the metabolic activities present in green coffee beans, which influence during the course of processing (Selmar, *et al.*, 2002; 2006). The metabolic reactions which are mainly due to germination processes (Bytof *et al.,* 2007) and stress metabolism, are responsible for the significant changes in the chemical composition of coffee beans and hence for their changes in quality (Bytof *et al.*, 2005). It was also demonstrated that in the first phase of drying, numerous metabolic reactions are taking place to a great extent. The soluble carbohydrates, known for their role as relevant aroma precursors, represent one of the most important class of substances (Bradbury, 2001), the concentration of which is markedly influenced by the method of processing (Knopp *et al*., 2006).

The most widespread use of microarrays is in comparison of gene expression of a set of genes from a sample maintained in particular condition to the same set of genes from reference sample maintained in normal conditions (Pollack *et al*., 1999; Bumgarner, 2013). The identification of genes linked to the qualitative aspects of the coffee drink is one of the major objectives of the different research groups working on coffee. The present study involves an attempt to characterize some biological processes by careful quantitative evaluation of gene expression using real-time quantitative PCR (RT-qPCR). Analysis was also done to find differentially expressed genes in samples from two locations that were processed through same post-harvest treatments for a possible association between the changes in gene expression profile, the post-harvest processing and geographic influences on the coffee seed. Sixteen genes of interest, involved in both metabolic and biochemical activities associated with quality characteristics of the beverage for quantitative evaluation, were selected.

Materials and methods

Genotypes

Coffea arabica cv. Catuai seeds, collected in two plantations and treated with different techniques of post-harvest processes (natural, washed, semi-washed method) formed the material in the study. The locations of the plantations were Barra do Choca (latitude 14° 52'5''S, longitude 40° 34'44''W), State of Bahia (Brazil), situated at an altitude ranging from 860 to 900 MSL with semi-humid tropical climate, and Inhobim (latitude 15° 16'S, longitude 40° 57' W), State of Bahia (Brazil), altitude ranging from 600 to 840 MSL with semi-humid tropical climate.

Extraction of RNA from seeds

RNA was extracted from seeds using extraction buffer that composed of 100 mM of Tris-HCl (pH 8.0), NaCl (2.0 M), Spermidine (0.5 g L^{-1}), EDTA (25 mM), CTAB (2% p/v), PVP 30K (2% p/v), β-mercaptoethanol (2% p/v) and proteinase-K (0.5 mg mL^{-1}). Prior to this, the counters and centrifuges were cleaned with detergents such as RNase-specific ExitusPlus (Del Chimica).

DEPC treated H_2O was used to prepare all solutions. The final buffer was sterilized with 0.22 μM filter and stored at room temperature. LiCl (8 M) was prepared, sterilized by filtration and stored at room temperature. Around 18-19 grams of seeds, dried to the standard water content of commercial samples (11% w/w), were finely ground using Super Jolly Professional Espresso Grinder (Mazzer Luigi Srl). Hot extraction buffer (30 mL) was added to one gram of the resulting powder and incubated for two hours at 55 °C, then centrifuged at 7000 rpm for 15 min maintaining temperature at 10 °C, in order to precipitate the PVP. The supernatant (aqueous phase) was transferred into 50 mL falcon tubes and mixed with one volume of chloroform: isoamyl alcohol mixture (24:1) by inversion. The mixture was centrifuged for 25 min at 4000 rpm at 10 °C. This organic extraction step was

repeated twice. The supernatant aqueous phase was then transferred to 15 mL falcon tubes, 0.3 volume of 8 M LiCl was added, mixed by inversion, and allowed to precipitate overnight, at 4 °C. The following day, this mixture was centrifuged for 25 minutes at 7000 rpm at 10 °C. The aqueous phase was completely removed and the pellet was washed with 1 mL of 75 per cent ethanol. The pellet was resuspended in 100 µL of RNase free H_2O and then transferred to 1.5 mL Eppendorf tubes. The RNA extracted was quantified using spectrophotometric measurement of absorbance (A) of the sample at wavelengths of 230, 260, 280 and 310 nm. The quality of RNA extracted was tested by separation on one per cent agarose gel. Before being loaded, the samples were added to loading buffer and denatured at 65 °C for 3-5 minutes, then immediately placed in ice. The visual integrity of the sample was judged by presence and intensity of the two typical bands of ribosomal RNA 18 S and 28 S.

Microarray and probe selection

Each glass slide of microarray 90K was divided into two areas (45K) and in each area, 8,382 oligos were placed in five replicates. Appropriate probes, complementary to the target nucleotides, were selected from Expressed Sequence Tags (ESTs) of *Coffea arabica* present in a local database (sequences available on request). OligoWiz, a client server application, offered the detailed graphical interface and realtime user interaction, on the client side and a large collection of species, on the server side. Probes were selected according to five weighted scores: cross-hybridization, Δ*T*m, folding, position and low complexity and probes were placed with respect to sequence annotation using regular expressions. RNA extracted was converted to cDNA using reverse transcriptase enzyme and nucleotides were labelled using different fluorescent dyes.

RNA amplification

The Amino Allyl Message Amp™ II, aRNA amplification kit, was used for amplification of sRNA. The procedure consisted of reverse transcription with an oligo (dT) primer, bearing a T7 promoter, using ArrayScript™ reverse transcriptase (RT), engineered to produce higher yields of first-strand cDNA than wild-type enzymes. ArrayScriptRT, catalyzed the synthesis of virtually full-length cDNA. The cDNA then undergoes the sysnthesis of second strand and then becomes the template for *in vitro* transcription (IVT) with T7 RNA polymerase. To maximize the RNA yield, AmbionMEGAscriptR IVT technology was used to generate hundreds and thousands of antisense RNA copies of each mRNA in the sample. In this protocol the antisense amplified RNA, is referred to as aRNA. The IVT, was configured to incorporate the modified nucleotide, 5-(3-aminoallyl) -UTP (aaUTP) into the aRNA during *in vitro* transcription. aaUTP, contains a reactive primary amino group on the C5 position of uracil, which will be chemically coupled to N-hydroxysuccinimidyl ester-derivatized reactive dyes (NHS ester dyes), such as Cy™3 and Cy5. The labelled aRNA is suitable for use in commercially available, microarray gene expression systems.

The data generated was analysed by identification of the spots, and distinguishing them from spurious signals, followed by the determination of the spot area, to be surveyed and the determination of the local region, to estimate background hybridization. Finally, the summary statistics was reported for each spot, assigning the spot intensity, after subtracting for background intensity.

Background calculation

The quantitation software does not quantize the outer circles to estimate the background. The background is estimated by checking the intensity of the spots named as quality control (QC) or negative control (NC) of the chip as they should not make hybrids even if they have synthesized sequences. Quality spots gave specific signals, as these were of sequences from plant. The background signals was subtracted as Cy3 fluorophore was used, which emitted wavelengths in autofluorescence of the synthesized DNA. In general, the NC has a low intensity scan average maximum of 200.

The present study used the median value as the metric to represent the spot intensity, with background median value subtracted from it. The expression data was represented based on the absolute measurement, with each cell in the matrix representing the expression level of the gene in abstract units. The absolute measurement data was converted to discrete numbers, using binary expression matrix, of 1 and 0, where 1 denoted that

gene was expressed above the user defined threshold and 0 indicated the gene was expressed below the threshold. The expression profile of a gene was represented as a row vector and that of sample as column vector.

Multianalyzer viewer (MeV; released under terms of the Artistic License v2.0), was used for visualization and data-mining of large-scale genomic data analysis. Significance analysis of microarray (SAM), was used to identify the significant genes, in the set of microarray experiments (Tusher *et al*., 2001). In significance analysis of Microarrays (SAM), the relative difference d(i), was compared to the distribution of d(i), following random permutation of the sample categories. Genes having scores greater than a threshold were deemed potentially significant. To estimate the false discovery rate (FDR), nonsense genes were identified by analyzing permutations of the measurements. To identify smaller or larger sets of genes, the threshold was adjusted and FDRs were calculated for each set. Cluster analysis was carried out with samples showing similar patterns which were characteristic to the group.

Clustering

In the present study, the hierarchical divisive clustering (Alon *et al*., 1999)/ Non-hierarchical clustering with K-means/ self organizing maps (SOMs) were used. Divisive clustering was adopted wherein the whole set of genes was considered as a single cluster and was broken down iteratively into sub-clusters, with similar expression profiles until each cluster contained only one gene.

Analysis of gene expression with quantitative real-time PCR

To validate the gene expression data generated in Micro Array analysis, a set of 16 genes, involved in both metabolic and biochemical activities, associated with quality characteristics of the beverage, were selected along with two reference genes. The expressions of these genes were analyzed using real time PCR. All the selected sequences in seqformat, were aligned with the SeqMan program (DNaStar) and primers were designed targeting the zones comprising the largest number of the sequences with a good degree of pairing. All consensus alignments were analyzed with BLASTn against both the nucleotide database, and against the non-human non-mouse EST database, specific to *C. arabica,* to guarantee the uniqueness of the sequences. The software chosen for the design of oligonucleotides is the Primer3 online program (http:// frodo.wi.mit.edu).

Sample preparation for real time analysis

Approximately 1µg of total RNA, was treated with the Turbo DNaseI (Ambion) enzyme for elimination of possible contamination of genomic material. Total RNA 1 µg; Turbo DNase buffer 1X; Enzyme 1 unit; Water up to 10 µL. The reaction was carried out in a thermocycler DNA Engine, MJ Research, PTC-200 (Genenco) incubating at 37 °C for 30 minutes. Reaction was blocked by adding 0.3 µL of 0.5 m EDTA pH 8.0 in order to have the final concentration of 15 mM of EDTA, and incubated at 75 °C for 10 minutes.

RNA was now used in reverse transcription with the iQScrypt kit (BioRad) which already contains the right concentration of random primers and oligods in the buffer, according to following reaction: Reagent quantity final concentration: Total RNA 1 µg; 5 X iScript Buffer 4 µL 1 X; iScript reverse enzyme 1 µL; H_2O RNase-free up to a final volume of 20 µL. Reagents were mixed and amplified in a DNA Engine thermocycler, MJ Research, PTC-200 (Genenco): at incubation 25 °C for 5'; incubation at 42 °C for 30 '; incubation at 48 °C at 15' and final incubation at 85 °C for 5'. The samples were stored at -20 °C or used immediately.

Real time PCR quantification

All real-time reactions were done in triplicate, with a negative control (NTC), taking into account the error given by the standard deviation (data dispersion index). The reaction was prepared according to the IQ ™ SYBRGreenSupermix protocol (BioRad): cDNA 1 µL; iQSybr 2 X (100 mM) 7.5 µL 1 X; Primer For. 400 nM; Primer Rev. 400 nM H_2O RNase-free final volume 10 µL; KCl, 40 mM Tris-HCl pH 8.4, 0.4 mM dNTPs, iTaq DNA Polymerase 50 U mL^{-1}, 6 mM $MgCl_2$; SYBR Green I, 20 mM fluorescein, stabilizers. Reagents were mixed and centrifuged briefly, reaction was carried out in a C1000 Thermal cycler associated with CFX 96 Real-Time system (BioRad).

The raw fluorescence data was used to calculate the efficiency, to identify the best conditions for each

Table 1. Details of primer sequences designed for the reference genes and genes of interest for use in qPCR analysis

Reference gene	Forward primer (5'-3')	Reverse primer (5'-3')	Size of amplification product (bp)
Rpl 7	CATTCGAGGTATCAATGCTATGCA	TGTCTCAGGCGCAGAAGCT	66
S 24	GCCCAAATATCGGCTTATCA	TCTTCTTGGCCCTGTTCTTC	94
UBQ 10	TCAACCCTTCACTTGGTGCT	CAGACCAGCAGAGGTTGATC	169
GAPDH	CTTCCAGCCCTCAATGGTAA	ACTGTTGGAACTCGGAATGC	53
14-3-3	GCAGGCTGAGAGGTATGAGG	CGCTGTCCACTGTCTTAGCA	168
Isocitrate lyase	AAGCCAGGTGAATGTGGAAG	TGAACTGCAAATGTGCTGAA	108
GAD	GGGTTTATGTGGGACGAAGA	TAGAGTGAAGGCACCCATCC	126
β-tubuline	TTCCTCCGACTGGTTTGAAG	CTGCGGAACATAGCTGTGAA	10
Caffeine synthase	CGTCCCACCATTCAGATTTT	GGTAGAAGCTTGGCAGCAAC	83
p-coumaroyl 3 hydroxylase	CAAGGCAAAAGAGTGGCAAT	GTCCCAAAGGAGTCCAATGA	107
5-α steroid reductase	ATCTGGAGGGGGAATCAAAC	CGGGGAAGATCAGAAATGAA	113
Fatty acid synthase T	GGATGAGGATGGCAGAAAG	CATGCTGGTTCACATCCAAG	98
Pyruvate kinase	GAGCCCACTGGAGAGTCTTG	CCACCACGTGTCAGAACAAC	104
Geranyl transferase	CTTCTGCTTGTGCTGCTGAG	TCCCTCTCCTCAAATCATCG	87
Shikimate hydroxycinnamoyl transferase	ACATTTGAACGGGAGCACAT	AACCACCACCGCATGAATAG	60
Sucrose synthase	GGAGACCGAAGGAAGGAATC	GAATTGGCCGTTCAAGTTGT	89
Starch synthase	GGTGTTGACGGAGACGAAAT	TGCGAGGCAATAGGCTAACT	96
hexose transport protein	CTACGGTGCTGCAAAAATCA	AAGGAAGAGAGCCCCAAGAG	108
Aldolase	TCGTAACCTGAATGCCATGA	CAAGCCTTTAGGGTGCTCTG	100
Enoyl hydratase	TGAGATTGCTTTGGCTTGTG	ACGGGCAAGCTTCTGAGATA	99
Invertase	AGGTTTGTCAGCAGGTCCAC	CCTTCACACTTGGGGATGAT	11

pair of primers. The analysis was carried out with LinReg PCR, a program described by Ramakers *et al.* (2003) and by Ruijter *et al.* (2009) (version11.0, download: http://LinRegPCR.HFRC.nl).

The reference genes candidates were tested using the geNorm™ software (Vandesompele *et al.*, 2002) version 3.5, which allowed calculation of the stability of the gene expression of a specific one reference (value defined as M), as the geometric mean of a given gene with respect to the others. The genes showing the lowest values of M were considered to be the most stable. Relative changes in the expression were recorded as the ratio of the target gene on the reference one, in accordance with the mathematical model proposed by Pfaffel (2004) using the BioRad CFX Manager software version 1.1.

Statistical analysis

Data obtained from real-time reactions were analyzed through Xlstat package of excel. In particular the data normalized with the genes of specific reference for this work were used for multivariate ANOVA analysis, considering the treatment as a fixed effect, while the place and the place-treatment interaction as variables. The principal component analysis (PCA) was also performed to have an overall graphical representation of the correlation between genes, samples and treatments, using Pearson's correlation matrix.

Statistical analysis of genes was carried out using Significance Analysis of Microarrays (SAM), to study the treatment and location effect analysis, mainly to identify the genes, with statistically significant changes in expression.

Sensory analysis of "washed-natural" samples

The samples of green coffee were also tested for sensory analysis; the green seeds of coffee were all roasted at the same level in lab roaster (Probat) and tasted in espresso mode according to standard brewing procedures (Illy and Viani, 2005).

Results and discussion

The Significance Analysis of Microarrays (SAM) assigned score to each gene, based on changes in gene expression, relative to standard deviation of repeated measurements.

Post harvest treatment analysis

SAM two class unpaired was utilized, where samples fell in one of two groups, and the subjects were different between the two groups (analogous to a between subjects t-test).

Location Barra do Choca:

- Test 1 BaL *vs.* BaN (wet *vs.* dry treatment)
- Test 2 BaL *vs.* BaS (wet *vs.* semi-dry treatment)
- Test 3 BaN *vs.* BaS (dry *vs.* semi-dry treatment)

Table 2. Results of test1. BaL *vs.* BaN (wet *vs.* dry treatment) showing positive significant genes, font with grey background showing negative

Gene	Protein	Expected score (dExp)	Observed score (d)	Numerator (r)	Denominator (s+s0)
CAC05358_1	GlycosyltransferaseCAZy family 14	0.22	2.92	550.37	188.72
CAC02956_1	Tubulin folding cofactor B	-0.24	3.00	1450.46	483.78
CAC06743_1	Cytochrome	0.51	3.05	537.10	176.23
CAC00668_1	P450	-0.91	3.14	2812.63	894.51
CAC06913_1	class1 heat shock protein	0.56	3.19	807.13	253.02
CAC03238_1	NA	-0.18	3.21	918.98	285.89
CAC02546_1	dehydration responsive family protein	-0.32	3.25	2313.38	711.57
CAC08359_1	NBS-LRR resistance gene-like protein ARGH m10 (Fragment)	1.17	3.29	1293.80	393.20
CAC08585_1	small molecular heat shock protein 17.5	1.52	5.05	3055.42	605.16
CAC00207_1	CII small heat shock protein	-1.28	5.19	1105.16	212.74
CAC00668_2	NA	-0.91	6.08	2655.46	436.92
CAC03022_1	Heat shock protein 90-2	-0.22	6.21	796	128.21
CAC02786_1	NA	-0.27	-6.01	-1658.00	275.88

Table 3. Results of test 2 BaL *vs.* BaS (wet *vs.* semi-dry treatment) showing positive significant genes, font with grey background showing negative

Gene	Protein	Expected score (dExp)	Observed score (d)	Numerator (r)	Denominator (s+s0)
CAC02095_1	Maturase	-0.43	3.68	5417.05	1472.13
CAC04233_1	Diphtamide biosynthesis protein	0.01	3.77	774.00	205.52
CAC05061_1	Dehydrin	0.16	3.81	1120.08	294.08
CAC01204_1	DH1Aa	-0.68	4.16	1574.19	378.76
CAC00668_1	Putative uncharacterized protein	-0.92	5.35	5509.79	1030.35
CAC03022_1	class1 heat shock protein	-0.22	5.73	910.13	158.90
CAC08585_1	heat shock conjugate 70 kDa	1.53	6.10	3356.72	550.52
CAC00668_2	CII small heat shock protein	-0.92	9.91	3240.33	327.04
CAC05613_1	Heat shock protein class I 17.5 kDa	0.27	10.47	6296.73	601.25
CAC04936_1	YMR049Cp-like protein	0.14	-10.68	-4263	399.18
CAC00320_1	cytochrome P450 like protein	-1.16	-7.38	-1002.5	135.93
CAC02786_1	Dead Box ATPD dependant RNA helicase	-0.27	-7.00	-2348.62	335.68
CAC01124_1	NA	-0.71	-6.91	-3850.28	557.11
CAC07646_1	Seed maturation protein	0.79	-6.62	-9220.48	1392.52
CAC06549_1	NA	0.47	-6.05	-1774.4	293.53
CAC00324_1	Peptidyl-tRNA hydrolase II like protein	-1.15	-5.12	-701.56	137.09
CAC00481_1	Proteasome subunit alpha type	-1.03	-4.79	-1320.39	275.43
CAC01455_1	Fructose bisphosphate aldolase, putative	-0.60	-4.74	-2432.35	513.08
CAC02797_1	Acyl-CoA synthetase	-0.27	-4.72	-1896.95	401.49
CAC00350_1	NA	-1.13	-4.30	-382.814	88.93
CAC06663_1	Putative uncharacterized protein	0.50	-4.30	-1375.40	319.61
CAC03052	NA	-0.22	-4.28	-643.78	150.32
CAC00730_1	NA	-0.88	-4.24	-532.02	125.47
CAC07112_1	Acetolactate synthase (Fragment)	0.62	-4.23	-1273.52	301.26
CAC04586_1	putative peptydil cis-trans isomerase	0.07	-4.19	-1753.10	418.07
CAC05084_1	NA	0.17	-4.06	-886	218.44
CAC06359_1	NAD-dependent isocitrate dehydrogenase	0.43	-4.05	-5881.89	1451.38
CAC08678_1	Cytochrome c oxidase subunit 1 (Fragment)	2.07	-3.96	-1257.03	317.08
CAC06492_1	NA	0.46	-3.88	-542.17	139.62
CAC05804_1	NADH ubiquinone oxidoreductase chain 2 (Fragment)	0.31	-3.84	-1222.08	318.64
CAC03238_1	NA	-0.18	-3.79	-1259.26	332.69
CAC03312_1	NA	-0.17	-3.70	-1501.57	405.76
CAC04015_1	Protease C56, putative	-0.04	-3.64	-428.13	117.48
CAC03899_1	NA	-0.06	-3.63	-896.68	246.99
CAC00159_1	NA	-1.37	-3.61	-9825.87	2718.36
CAC03060_1	Prefoldin 6	-0.22	-3.60	-529.68	146.95
CAC04031_1	Peroxidase	-0.03	-3.52	-350.87	99.57
CAC07306_1	UDP-glucose glucosyl transferase	0.68	-3.52	-953.48	271.13
CAC03727_1	Phosphoprotein phosphatase, putative	-0.09	-3.50	-971.23	277.13
CAC03866_1	NA	-0.06	-3.42	-876.45	256.48
CAC04054_1	Protease C56, putative	-0.03	-3.36	-1059.34	315.40

Table 3. Results of Test 3 BaN *vs.* BaS (dry *vs.* semi-dry treatment) showing positive significant genes, font with grey background showing negative

Gene	Protein	Expected score (dExp)	Observed score (d)	Numerator (r)	Denominator (s+s0)
CAC05358_1	Glycosyltransferase CAZy family 14	0.22	2.92	550.37	188.72
CAC02956_1	Tubulin folding cofactor B	-0.24	3.00	1450.46	483.78
CAC06743_1	Cytochrome	0.51	3.05	537.10	176.23
CAC00668_1	P450	-0.91	3.14	2812.63	894.51
CAC06913_1	class1 heat shock protein	0.56	3.19	807.13	253.02
CAC03238_1	NA	-0.18	3.21	918.98	285.89
CAC02546_1	dehydration responsive family protein	-0.32	3.25	2313.38	711.57
CAC08359_1	NBS-LRR resistance gene-like protein ARGH	1.17	3.29	1293.80	393.20
CAC08585_1	m10 (Fragment)	1.52	5.05	3055.42	605.16
CAC00207_1		-1.28	5.19	1105.16	212.74
CAC00668_2	small molecular heat shock protein 17.5	-0.91	6.08	2655.46	436.92
CAC03022_1	CII small heat shock protein	-0.22	6.21	796	128.21
CAC02786_1	NA	-0.27	-6.01	-1658.00	275.88

Location Inhobim:

- Test 4 InL *vs.* InN (wet vs. dry treatment)
- Test 5 InL *vs.* InS (wet *vs.* semi-dry treatment)*
- Test 6 InN *vs.* InS (dry *vs.* semi-dry treatment)*

(* tables available as supplementary material)

Dry and semi-dry methods represented an increased expression of heat shock proteins and dehydration-related genes. The overexpressed proteins are (1) Small molecular heat shock protein 17.5 kDa, CII small heat shock protein, Heat shock protein 90 and (2) Heat shock conjugate 70 kDa. This was explicable because coffee cherries treated with the dry method are left to dry in the sun, and so are subject to heat, sunlight and dehydration for an extended period. Tubulin folding cofactor B is also overexpressed and it might have a role in degradation of α microtubules in senescence, as already shown in *A. thaliana* (Pankaj *et al*., 2006).

The wet and dry method showed an up-regulation of OLE-2. Oleosins are small molecular mass proteins which are located on the surface of oil bodies. They form a barrier to prevent the phospholipid layer from contacting, and coalescing with the phospholipid layers of adjacent oil bodies (Huang, 1996). The lipase binding and lipolysis during the seed germination is facilitated by the large surface area of oil bodies maintained as small droplets (Kim *et al*., 2002). Tissue-specific expression of oleosins, during embryogenesis and seed development, requires high transcriptional activity. The promoters of oleosins are absolutely inactive, in vegetative tissues such as roots, stems, and leaves (Plant *et al*., 1994).

The wet method also showed an increased expression of storage globulin 11S. Reserve proteins of the 11S family are the most abundant globulins in coffee seeds. From industrial point of view, they act as a nitrogen source during roasting and as precursors of flavor and aroma compounds. Glycosyltransferase Caz family had a decreased expression level in wet method: some of the roles of plant glycosyltransferases are the stabilization of the pigments, the regulation of the plant growth factors, and an increase in aglycone solubility (Jones and Vogt, 2001).

SAM two class unpaired

Table 4. Results of test 4 InL *vs.* InN showing positive significant genes, font with grey background showing negative

Gene	Protein	Expected score (dExp)	Observed score (d)	Numerator (r)	Denominator (s+s0)
CAC08163_1	Formin-like protein	0.973	3.44	816.85	237.54
CAC02022_1	NA	-0.41	3.61	645.81	179.08
CAC07166_1	Predicted proteinAspartate aminotransefare	0.59	3.65	869	238.17
CAC05055_1	Putative suppressor of ty	0.15	3.69	583.82	158.20
CAC02107_1	Putative glycosyl transefarse	-0.39	3.76	641.39	170.78
CAC05358_1	CAZ family14	0.20	3.88	645.65	166.44
CAC05116_1	Extensin like protein	0.16	3.95	906.68	229.34
CAC06515_1	Aspartic proteinase nephenestin-1percursor	0.42	3.97	727.52	183.16
CAC06901_1	Glutaredoxin, grx, putative	0.52	3.98	17493.45	4400.19
CAC08133_1	Putative uncharacterized protein	0.96	4.10	9769.55	2381.86
CAC04314_1	Putative uncharacterized protein	0.02	4.18	2772.96	663.66
CAC04711_1	MaltoseO	0.09	4.42	914.35	206.68
CAC04572_1	Acetyltransferase Predicted	0.06	5.07	1159.50	228.55
CAC00797_1	Predicted protein	-0.79	5.13	25361.77	4947.60
CAC04422_1	Putative uncharacterize d protein	0.04	5.67	1286.44	227.02
CAC01032_1	SUI1proteinSuccinyl diaminopimelate	-0.70	-10.29	-5100.65	495.59
CAC05661_1	desuccinylase Chemotaxis	0.25	-8.65	-4093.57	473.31
CAC02496_1	responseregulator, CheY4	-0.30	-7.22	-9600.52	1330.50
CAC00117_1	11S Storage globulin	-1.40	-6.88	-2682.05	389.59
CAC00668_1	Class I heat shock protein	-0.86	-6.25	-1175.48	188.10
CAC03881_1	Pseudo response regulator7	-0.05	-5.87	-576.07	98.12
CAC00625_1	Seed maturation protein	-0.88	-5.72	-4767.95	833.93

In the semi-dry method two genes were found that probably interact together: Dehydrin dh1 and LEA4 (late embryogenic abundant), dehydrins (or group 1 late embryogenic abundant proteins) are hydrophilic, Gly-rich proteins, that are induced in vegetative tissues in response to dehydration, elevated salt, and low temperature, in addition to being expressed during the late stages of seed maturation. Expression of dehydrin gene is associated with osmotic stress and is widely perceived to participate with other LEA proteins, in dehydration process that occurs during late stages of seed maturation, by assisting the acclimatization of seed tissues to the lower water content found in mature seeds (Close, 1996; Nylander, 2001).

The dehydrins synthesized in seeds during maturation are presumed to continue to stabilize the associated cellular structures during seed quiescence. Recently, it has been proposed that dehydrins may also possess a radical-scavenging capability (Hara *et al*., 2003) and have metal-binding properties (Alsheikh *et al*., 2003), both characteristics likely to be useful during long periods of seed storage.

Analysis between locations

SAM analysis as two class unpaired data carried out was as follows:

- Test 7 BaL *vs.* InL (Barra do Choca *vs.* Inhobim, wet treatment)*
- Test 8 BaN *vs.* InN (Barra do Choca *vs.* Inhobim, dry treatment)*
- Test 9 BaS *vs.* InS (Barra do choca *vs.* Inhobim, semi-dry treatment)*

(* Tables available as supplementary materials)

SAM two classes unpaired

In the results generated in Location Inhobim, all three treatments showed an up-regulation of SUI1 gene. Accumulation of *CaSUI1* mRNA in mature coffee bean, a prerequisite for intense transcription and protein synthesis, is required for the further germination of coffee seeds (Giorgini, 1988). It was also observed that sequences similar to *C. arabica* were observed in "Robusta", *i.e*., somatic embryos from *C. canephora*. As reported by Lashermes *et al.*, 1999, *C. canephora* is one of the progenitor species for *C. arabica*. Also, during evolution, high conservation of the SUI1-coding sequence is expected, due to its essential role in the initiation of translation process. The SUI1 protein, besides initiating transcription, may also be involved in repairing impaired mRNAs (Cui *et al*., 1999).

Presence of sequences homologous to *CaSUI1* sequence was revealed in cDNA libraries from all tissues during screening of coffee ESTs from the Brazilian Coffee Genome Project during early decade of the millennium. This confirmed the presence of housekeeping function of the SUI1 protein (Fields and Adams, 1994). Comparison of the intensities of bands revealed, after probing with total RNA loaded on the same gel, that the *CaSUI1* gene was highly expressed in mature beans while, the expression of β-galactosidase and storage-protein (11S) encoding genes was not detected (Marraccini *et al*.,1999, 2001).

In the present study, all three treatments in location Inhobim showed an up-regulation of SUI1 gene and down regulation of 11S storage globulin. This may be related to differences in elevation, which is known to affect the final quality of the coffee beverage favourably (Joet *et al*., 2009). The location effects also suggest that environmental factors have a stronger influence than the effects of post-harvesting treatments. The number of differentially expressed genes included significant genes from treatment analysis and others. These samples had to be considered separately to reduce the potentially confounding effects from differences between different samples.

SAM multiclass

The multiclass response indicated more than two groups, containing different experimental units each. This is a generalization of the unpaired setup, to more than two groups. SAM Multiclass calculates, the standardized mean difference between the gene's expression in one class, versus its over all mean expression. Multiclass in this case was not useful because of possible confounding effects.

Since the seeds were treated differently and grown in different locations, resulting genes could be called significant because of the location or treatment effects, but MeV does not give the information about variable effects. Multiclass was useful for divisive HCL. It clusters samples and genes calculating distances between them using Pearson correlation. All the genes and samples are those present from previous test (SAM two classes unpaired).

Sample clusters

Two major branches that divide samples in two clusters "Ba" and "In" are observed. The "Ba" branch shows that the dry (BaN, replicates 1 and 5) and wet methods (BaL, replicates 5 and 4) are clustered together, and so are the dry (BaN, replicates 2, 3, and 4) with semi-dry treated samples (BaS, replicates 1, 2 and 3). The dry method could be placed in between the other 2 methods. The "In" branch has two minor clusters, but the three different methods are clustered separately. In this case the dry process is separated from wet and semi-dry method but it is much closer to wet method. Wet and semi-dry samples could not be clustered together; SAM 2 class unpaired test also showed that BaL *vs.* BaS and InL *vs.* InS tests have the biggest number of differentially expressed genes.

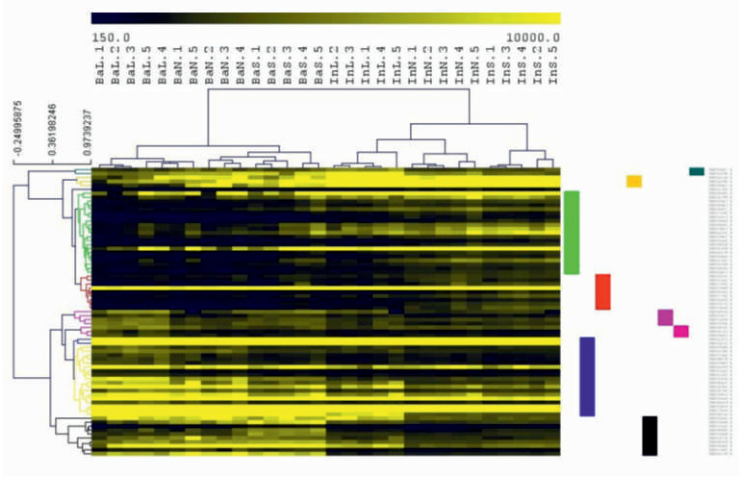

Fig. 1. Heat map generated using hierarchical divisive clustering method with genes showing similar expression in the different conditions tested

Details of Gene clusters:

Cluster 1

CAC00867_1 Predicted protein

CAC01257_1 Non-canonical ubiquitin conjugating enzyme, putative

CAC01584_1 Metal-dependent phosphohydrolase HD sub domain protein

CAC02107_1 Suppressor of ty, putative

CAC02334_1 NA

CAC02799_1 Adenosine monophosphate kinase

CAC03653_1 Predicted protein

CAC04314_1 Putative uncharacterized protein

CAC04543_1 CETS1

CAC04927_1 Putative uncharacterized protein (Fragment)

CAC05116_1 Extensin-like protein

CAC05221_1 HTH-type transcriptional regulator kipR

CAC05308_1 Chloroplast isoprene synthase

CAC05358_1 glycosyltransferase CAZy family 14

CAC05869_1 Cytochrome b (Fragment)

CAC06478_1 Chlorophyll a/b-binding protein

CAC06515_1 Aspartic proteinase nephenestin-1 percursor

CAC06615_1 NA

CAC06913_1 NA

CAC07490_1 NA

CAC08508_1 Hyp-rich glicoprotein

Cluster 2

CAC00324_1 Proteasome subunit alpha type

CAC00880_1 60S ribosomal protein L24

CAC01032_1 SUI 1 protein

CAC01124_1 Predicted protein

CAC01756_1 Nucleotidyl transferase

CAC02064_1 small nuclear ribonucleoprotease

CAC02303_1NADH dehydrogenase subunit D (Fragment)

CAC02786_1 NA

CAC02953_1 CTLMA2 (Fragment)

CAC03312_1 mavicyanine

CAC04586_1 Predicted protein

CAC04606_1Phosphonates metabolism transcriptionalregulatorPhnF

CAC04936_1 Cytochrome P450 likeNAD-dependent isocitrate dehydrogenase

CAC06549_1 Peptidyl t-RNA hydrolase II like protein

CAC07646_1 Putative uncharacterized protein

CAC08678_1Cytochrome c oxidase subunit 1 (Fragment)

CAC00320_1 dead box ATP-dependent RNA helicase

Cluster 3

CAC00625_1 Seed maturation polypeptide

CAC00668_1 Heat shock protein class I

CAC00668_2 Heat shock protein 90-2

CAC03022_1 Heat shock conjugate 70 kDa

CAC03071_1 Senescence associated protein

CAC03384_1 Heat shock proteinZ

CAC05661_1Class I low-molecular-weight heat shockprotein

CAC06738_1 Heat shock protein, putative

CAC08359_1 Small molecular heat shock protein 17.5kDa

CAC08585_1 C II heat shock protein

Cluster 4

CAC02045_1 NA

CAC02219_1 NADH-ubiquinone oxidoreductase chain 4

CAC03646_1 Predicted protein (Fragment)

CAC03928_1 NA

CAC04422_1 Putative uncharacterized protein

CAC04572_1 Putative uncharacterized protein

CAC05460_1 Pentatricopeptide containing protein

CAC07094_1 Cyclophilin-like peptydil-prolyl cis-trans isomerase

CAC07792_1 Putative uncharacterized protein

Cluster 5

CAC00207_1 NA

CAC00481_1 Fructose-bisphosphate aldolase, putative

CAC01455_1 Acyl-CoA synthetase

CAC02546_1NBS-LRR resistance gene-like protein ARGH m10 (Fragment)

CAC03238_1 NA

CAC06663_1 NA

CAC06720_1 Putative uncharacterized protein

Cluster 6

CAC02095_1 Maturase

CAC02436_1 Pc13g15380 protein

CAC05613_1 ATP synthase subunit d

To summarise, of the total 18 genes of *Coffea arabica*, 16 target genes are involved in various metabolic pathways and two were reference genes. Significant differences were observed for samples treated with different post-harvest methods, specially for the samples from location of Barra do Choça, while expression profiles of Inhobim location was similar in all the three treatments. The interaction with external factors such the activities of the operator, the duration of the various steps, the method of storage and environmental effects during shipping of samples, would have been the cause.

Quantitative PCR highlighting the behaviour of different transcripts as enzymes markers of germination of seed (isocitrate lyase and β-tubulin) have been described in the literature as the most expressed in the washed coffee (Selmar *et al*., 2006; Bytof *et al*., 2007). In the present study, samples from the station Barra do Choça followed trend described for isocitratelyase. The β-tubulin instead, showed a contrasting expression pattern, with samples of semi washed Inhobim presenting the maximum value of the enzyme. This lack of variability in the expression profile may be related with the water content of the seed which, if it goes down to levels below 25 per cent is observed to affect normal metabolic activity and hence blocks the germination.

The GAD enzyme, associated with the stress response of the seed (Bytof *et al*., 2005) was expressed in the natural samples grown in location Inhobim, where samples underwent longer periods of dehydration and water stress and hence respond by producing a greater amount of the metabolite GABA, which in turn is produced subsequent to decarboxylation by the enzyme GAD. Samples collected in Barra do Choça, however showed no significant difference in the level of expression.

Caffeine synthase and enzymes involved in the metabolism of chlorogenic acids were mainly expressed in the washed and semi washed treated samples. The enzymes involved in the metabolism of plant hormones (5-α reductase steroid and geranyl transferase) were expressed more in natural samples than in washed and semi washed samples.

The enzymes involved in fatty acids metabolism, namely in the formation of triacylglycerols (pyruvate kinase) and in β-oxidation (enoyl CoA hydratase), were mainly expressed in semi washed samples collected from Barra do Choça. The lipid concentration in the seed is very important to create the flavour profile and the body of the drink. Enzymes pyruvate kinase and aldolase of glycolysis and gluconeogenesis analyzed in the present study showed higher levels of expression respectively in natural coffee from the station Barra do Choça (pyruvate kinase) and washed Barra do Choça (aldolase). Samples from Inhobim however, did not show significant differences between the three treatments.

Among enzymes associated with the metabolism and transport of sugars such as sucrose synthase, invertase and starch synthase (which contributes to the formation of starch, alternate source of carbohydrates), sucrose synthase is expressed more in the samples treated with the washed method from Inhobim location while the transport proteins of hexoses and invertase in semi washed and washed samples. This included starch synthase which was expressed predominantly in the washed samples. It is well known that the saccharose content, in the collected seeds, is one of the factors which contribute to determining the quality of a roasted coffee with respect to another (Campa *et al*., 2004; Ky *et al*., 2000); aromatic characteristics such as sweet, chocolate and caramel are important positive qualities to the drink of coffee. An analysis of gene expression profile of proteins involved in the transport, synthesis and degradation of sucrose and hexoses, stands out in the washed and semi-washed samples and these reflect in positive characteristics in the cup.

These samples were also subject to sensory analysis and the natural coffee were found to possess lower quality compared to washed and semi washed which did not show any appreciable differences between themselves. In this test the location of crops from Barra do Choçaor Inhobim was not influential for quality on the sensory profile of the drink.

Conclusion

The analysis of 16 genes was crucial to the understanding of some differences between the post-harvest methods (natural method, washed and semi washed), but given the wide variability in gene expression between samples grown in Barra do Choça and Inhobim, though it was not possible to generate a profile. The analysis of samples from the two geographical locations and treatments through PCA highlighted that the geographical locations indeed determined the values of gene expression and similarly significant differences were also observed between the three post-harvest treatments in samples from Barra do Choça. The same experimental design was proposed for the analysis of a specially developed microarray. The data are still under development and processing and probably the enrichment of the expression profile with the thousands of genes studied, could better describe and highlight the differences between the various methods of post-harvest processing that are employed on the seeds of green coffee. The technique of quantitative PCR was also used for analysis of gene expression profiles in green coffee samples collected in different times of the drying stage.

References

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D. and Levine, A.J. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences* **96** (12): 6745-6750.

Alsheikh, M.K., Heyen, B.J. and Randall, S.K. 2003. Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. *Journal of Biological Chemistry* **278**: 40882-40889.

Bradbury, A.G.W. 2001. Chemistry I: Non-volatile compounds. Chapter 1. In: *Coffee: Recent Developments.* (Eds.) Clarke, R. J. &Vitzhum, O. G. Oxford: Blackwell Science.

Bumgarner, R. 2013. Overview of DNA microarrays: types, applications, and their future. *Current Protocols in Molecular Biology,* doi:101002/0471142727.mb2201s101.

Bytof, G., Knopp, S.E., Kramer, D., Breitenstein, B., Bergervoet, J.H., Groot, S.P. and Selmar, D. 2007. Transient occurrence of seed germination processes during coffee post-harvest treatment. *Annals of Botany* **100**(1): 61-66.

Bytof, G., Knopp, S.E., Schieberle, P., Teutsch, I. and Selmar, D. 2005. Influence of processing on the generation of γ-aminobutyric acid in green coffee beans. *European Food Research and Technology* **220:** 245-250.

Campa, C., Ballester, J.F., Doulbeau, S., Dussert, S., Hamon, S. and Noirot, M. 2004. Trigonelline and sucrose diversity in wild *Coffea* species. *Food Chemistry*. **88**: 39-44.

Clarke, R.J. 2013. *Coffee Volume 1, Chemistry.* New York: Spinger. ISBN 978-9401086936.

Close, T. 1996. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiologia Plantarum* **97**: 795-803.

Cui, Y., Gonzalez, C.I., Kinzy, T.G., Dinman, J.D. and Peltz, S.W. 1999. Mutations in the MOF2/SUI1 gene affect both translation and nonsense-mediated mRNA decay. *RNA* **5**(6): 794-804.

Fields, C. and Adams, M.D. 1994. Expressed sequence tags identify a human isolog of the SUIl translation initiation. *Biochemical and Biophysical Research Communications* **198**: 288-291.

Giorgini, J.F. 1988. Ribonucleic acid characterization and synthesis in germinating coffee seed endosperm. *Brazilian Journal of Medicine and Biology Research* **21:** 811-824.

Hara, M., Terashima, S., Fukaya, T. and Kuboi, T. 2003. Enhancement of cold tolerance and inhibition of lipid peroxidation by citrus dehydrin in transgenic tobacco. *Planta* **217:** 290-298.

Huang, A.H.C. 1996. Oleosins and oil bodies in seeds and other organs. *Plant Physiology* **110:** 1055-1061.

Illy, A. and Viani R. 2005. *Espresso Coffee, the Science of Quality*. Elsevier Academic Press.

Joët, T., Laffargue, A., Descroix, F., Doulbean, S., Bertrand, B., de Kochko, A. and Dussert S. 2009. Influence of enviromental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. *Food Chemistry* **118**(3): 693-701.

Jones, P. and Vogt, T. 2001. Glycosyl transferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta* **213**: 164-174

Kim, H.U., Hsieh, K., Ratnayake, C and Huang, A.H. C. 2002. Expression of *Arabidopsis oleos* in genes and characterization of their encoded oleosins. *Journal of Biological Chemistry* **277**: 22677-22684.

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

Ky, C.L., Doulbeau, S., Guyot, B., Akaffou, S., Charrier, A., Hamon, S., Louarn, J., Noirot, M. 2000. Inheritance of coffee bean sucrose content in the interspecific cross *Coffea pseudozanguebariae* x *Coffea liberica* 'dewevrei'. *Plant Breeding* **119**: 165-168.

Lashermes, P., Combes, M. C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F. and Charrier, A. 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics* **261**(2): 259-266.

Marraccini, P., Allard, C., André, M-L., Courjault, C., Gaborit, C., Lacoste, N., Meunier, A., Michaux, S., Petit, V., Priyono, P., Rogers, W.J. and Deshayes, A. 2001. Update on coffee biochemical compounds, protein and gene expression during bean maturation and in other tissues. In: *Proceedings of the 19th International Scientific Colloquium on Coffee, Trieste, Italy*, Abstract B214 CD-rom ISBN 2-90012-18-9.

Marraccini, P., Deshayes, A., Pétiard, P. and Rogers, W. J. 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.

Nylander, M., Svensson, J., Palva, E. T. and Welin, B. V. 2001. Stress-induced accumulation and tissue specific localization of dehydrins in *Arabidopsis thaliana*. *Plant Molecular Biology* **45**: 263 279.

Pankaj, D., Bargmann, B.O. and Gadella Jr, T.W. 2006. Arabidopsis tubulin folding cofactor B interacts with α-tubulin *in vivo*. *Plant and Cell Physiology* **47**(10): 1406-1411.

Pfaffel, M.W. 2004. Chapter 3: Quantification strategies in real-time PCR A-Z quantitative PCR Editor Bustin S.A.

Plant, A.L., van Rooijen, G.J., Anderson, C.P. and Moloney, M.M. 1994. Regulation of an Arabidopsis oleosin gene promoter in transgenic *Brassica napus*, *Plant Molecular Biology* **25:**193-205.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D. and Brown, P.O.1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genetics* **23**(1): 41-46.

Ramakers, C., Rujiter, J. M., Deprez, R.H.L., Moorman, A.F.M. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data *Neuroscience Letters* **339**: 62 66.

Ruijter, J.M., Ramakers, C., Hoogaars, W.M.H., Karlen, Y., Bakker, O., Van den Hoff, M. J. B. and Moorman, A.F.M. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research* **37**(6): e45-e45.

Selmar, D., Bytof, G. and Knopp, S. E. 2002. New aspects of coffee processing: the relation between seed germination and coffee quality. In: Proceedings of the "19emeColloque Scientifique In-ternational sur le Caf ́e": ASIC Paris.

Selmar, D., Bytof, G., Knopp, S-E. and Breitenstein, B. 2006. Germination of coffee seeds and its significance for coffee quality. *Plant Biology* **8**: 260-264.

Sivetz, M. and Desrosier, N.W. 1979. *Coffee Technology*. AVI publishing company. Westport. Connecticut. Pp. 99-109.

Tusher, V. G., Tibshirani, R. and Chu, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(9), 5116–5121. doi:10.1073/pnas.091062498.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe B., Van Roy, N., De Paepe, A. and Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes *Genome Biology* **3**(7): research 0034.1-0034.12.

Wintgens, J.N. 2004. Coffee: Growing, Processing, Sustainable Production. A guide for growers, traders, and researchers.WILEY-VCH Verlag GmbH and Co.KGaA, Weinheim, Germany.