Mol Breeding (2009) 23:607–616
DOI 10.1007/s11032-009-9259-x

Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR

Fernanda Cruz · Samara Kalaoun · Paula Nobile · Carlos Colombo · Juliana Almeida · Leila M. G. Barros · Eduardo Romano · Maria Fátima Grossi-de-Sá · Maité Vaslin · Marcio Alves-Ferreira

Received: 27 August 2008 / Accepted: 15 January 2009 / Published online: 3 February 2009

Abstract Accuracy in quantitative real-time polymerase chain reaction (qPCR) requires the use of stable endogenous controls. Normalization with multiple reference genes is the gold standard, but their identification is a laborious task, especially in species with limited sequence information. Coffee (*Coffea* ssp.) is an important agricultural commodity and, due to its economic relevance, is the subject of increasing research in genetics and biotechnology, in which gene expression analysis is one of the most important fields. Notwithstanding, relatively few works have focused on the analysis of gene expression in coffee. Moreover, most of these works have used less accurate techniques such as northern blot assays instead of more accurate techniques (e.g., qPCR) that have already been extensively used in other plant species. Aiming to boost the use of qPCR in studies of gene expression in coffee, we uncovered reference genes to be used in a number of different experimental conditions. Using two distinct algorithms implemented by geNorm and Norm Finder, we evaluated a total of eight candidate reference genes (*psaB*, *PP2A*, *AP47*, *S24*, *GAPDH*, *rpl39*, *UBQ10*, and *UBI9*) in four different experimental sets (control versus drought-stressed leaves, control versus drought-stressed roots, leaves of three different coffee cultivars, and four different coffee organs). The most suitable combination of reference genes was indicated in each experimental set for use as internal control for reliable qPCR data normalization. This study also provides useful guidelines for reference gene selection for researchers working with coffee plant samples under conditions other than those tested here.

Keywords Reference genes · Coffee · Drought stress · Development · Gene expression · Real-time PCR · qPCR

Electronic supplementary material The online version of this article (doi:10.1007/s11032-009-9259-x) contains supplementary material, which is available to authorized users.

F. Cruz · S. Kalaoun · M. Alves-Ferreira (✉)
Department of Genetics, Center of Health Sciences, Federal University of Rio de Janeiro (UFRJ), Ilha do Fundão, Caixa Postal 68011, Rio de Janeiro, RJ CEP 21944-970, Brazil
e-mail: alvesfer@biologia.ufrj.br

P. Nobile · C. Colombo
Center of Genetics, Agronomic Institute of Campinas, Caixa Postal 28, Campinas, SP CEP 13012-970, Brazil

J. Almeida · L. M. G. Barros · E. Romano · M. F. Grossi-de-Sá
EMBRAPA Recursos Genéticos e Biotecnologia, PO Box 02372, Brasília, DF CEP 70770-900, Brazil

M. Vaslin
Microbioloy Institute, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brazil

Introduction

Coffee ranks among the five most valuable agricultural exports in developing nations, and its production

has great worldwide economic relevance (FAOSTAT 2008). Although coffee has been successfully improved by conventional approaches in recent years, coffee breeding has also proved to be a lengthy, laborious, and capital-consuming process (Etienne et al. 2002). The recent release of *Coffea* expressed sequence tag (EST) databases (Lin et al. 2005; Poncet et al. 2006; Vieira et al. 2006) has greatly prompted the study of genes involved in important agronomic traits such as sucrose and chlorogenic acid accumulation, relevant for flavor and aroma quality (Lepelley et al. 2007; Privat et al. 2008; Salmona et al. 2008). Coffee gene expression associated with the carotenoid biosynthetic pathway was studied, showing a redirection of the carotenoid flux towards the xanthophyll cycle, which implies an adaptation mechanism to drought (Simkin et al. 2008). Therefore, gene expression studies constitute a crucial step for functional characterization of coffee ESTs.

Quantitative real-time polymerase chain reaction (qPCR) is a robust method to study gene expression changes. The high sensitivity and specificity of qPCR analysis, even when limited amounts of RNA are available, have boosted its use among plant biologists (Gachon et al. 2004). Although the use of qPCR for gene expression studies in model species such as *Arabidopsis* and rice has been widespread, its use in coffee has been limited (Ganesh et al. 2006; Lepelley et al. 2007; Pre et al. 2008; Privat et al. 2008; Salmona et al. 2008; Simkin et al. 2006, 2008). Similar to other techniques aimed at evaluating gene expression, qPCR analysis normalization is required to allow accurate comparisons among samples. Normalization is necessary for the correction of technical variations such as differences in the quantity and quality of RNA, and in the efficiencies of reverse transcription and PCR (Udvardi et al. 2008). A number of strategies have been proposed to normalize qPCR data but normalization remains one of the most important challenges concerning this technique (Huggett et al. 2005).

Reference genes are frequently used to normalize qPCR analysis. The expression levels of these genes should be stable in all samples under investigation and experimental conditions evaluated (Vandesompele et al. 2002). Recent reports have demonstrated that some of the most well-known and frequently used reference genes are inappropriate for normalization in qPCR analysis due to expression variability (Czechowski et al. 2005; Remans et al. 2008; Schmittgen and Zakrajsek 2000). The importance of reference genes for plant qPCR analysis has been recently emphasized, even though the identification of these genes is quite laborious (Brunner et al. 2004; Reid et al. 2006). Microarray datasets can also be a rich source of information for selecting qPCR reference genes (Czechowski et al. 2005), but unfortunately this tool is still not available for most plant species, including coffee.

Several statistical methods for evaluating reference genes were proposed such as geNorm (Vandesompele et al. 2002), Norm Finder (Andersen et al. 2004), and Best Keeper (Pfaffl et al. 2004). These methods are based on different statistical algorithms, using multiple reference genes as the best strategy for normalization of qPCR results. geNorm applet provides a measure of gene expression stability (M), which is the mean pairwise variation between an individual gene and all other tested control genes (Vandesompele et al. 2002). Genes with the lowest M value have the most stable expression. In addition, geNorm performs a stepwise exclusion of the gene with the highest M value, resulting in the best combination of two reference genes with the most stable expression profiles. The Norm Finder approach focuses on finding the two genes with the smaller intra- and intergroup expression variation, where groups are defined as different experimental conditions (e.g., control and drought-stressed plants; Andersen et al. 2004). Since the stability value is a combination of intra- and intergroup variation parameters, Norm Finder ranks the genes that present minimal variation instead of candidates with correlated expression.

To date several studies have used qPCR to assess coffee gene expression profiles in different organs and tissues, e.g., flowers, leaves, stems, branches, roots, and fruits (pericarps and grains; Ganesh et al. 2006; Lepelley et al. 2007; Pre et al. 2008; Privat et al. 2008; Salmona et al. 2008; Simkin et al. 2006, 2008), and only one of these has investigated the expression stability of seven coffee reference genes during different stages of seed development. Among the seven candidate genes tested, polyubiquitin (*UBQ10*) showed the most stable expression pattern and was considered an acceptable reference gene for studies on coffee seed development (Salmona et al. 2008). Aiming to identify more adequate reference

genes for qPCR gene expression studies in coffee we evaluated eight coffee candidate reference genes (*GAPDH*, *S24*, *UBI9*, *UBQ10*, *rpl39*, *psaB*, *PP2A*, and *AP47*; Czechowski et al. 2005; Fernandez et al. 2004; Salmona et al. 2008; Simkin et al. 2006), and the expression stability of these genes was subsequently tested in different coffee tissues, cultivars, and growth conditions using both geNorm (Vandesompele et al. 2002) and Norm Finder (Andersen et al. 2004) algorithms.

Materials and methods

Plant material

Experiments were performed using 6-month-old *Coffea arabica* plants, cultivars "Catuaí Vermelho IAC44," "Mundo Novo 379/19," and "Bourbon Amarelo IAC J10". Plants were grown under controlled temperature (21 ± 2°C) and natural photoperiod, in trays containing 20 plants each. Plants were watered with 500 ml water at 1-day intervals. Water potential (Ψ_w) of each plant was measured at predawn using a Scholander-type pressure chamber, and control plants presented a Ψ_w of around −0.2 MPa. Drought stress was induced by withholding normal watering for 10 days. Under our growth conditions drought-stressed plants presented a Ψ_w of −4.45 MPa. Totally expanded leaves (third pair), stems, and lateral root samples were harvested from five different coffee plants, at two different times of the same year (February and July/Fall 2007). Mature cherry fruits were harvested from 10-year-old coffee trees grown in the field, in May (Fall 2008). Two fruits from each of five different plants of *Coffea arabica* cv "Catuaí Vermelho IAC 77" were collected in two different areas, which were 10 m apart. Fruits from each group (a total number of ten) were collected and pooled together. All samples were immediately frozen in liquid nitrogen, and stored at −80°C until needed for RNA extraction.

Total RNA isolation and cDNA synthesis

Frozen samples were ground to a fine powder in liquid nitrogen with a pestle and mortar. About 100 mg of this powder was resuspended in 500 µl cold (4°C) Concert^{TM} plant RNA reagent (Invitrogen) according to the manufacturer's instructions. After 5 min incubation at room temperature (RT), samples were centrifuged at RT, for 2 min, at 12,000×*g*. The clear supernatant was transferred to a new tube and 100 µl 5 M NaCl and 300 µl chloroform were added to samples, mixing thoroughly by inversion. Samples were subsequently centrifuged at 4°C for 10 min at 12,000×*g*. The aqueous phase was recovered and total RNA was precipitated with an equal volume of ice-cold isopropyl alcohol for 10 min at RT, followed by a centrifugation step at 4°C for 10 min at 12,000×*g*. The pellet was washed with ice-cold 75% ethanol, air-dried, and dissolved in 30 µl RNAse-free water. To avoid any DNA contamination, samples were treated with RNAse-free DNAseI (Invitrogen) at 37°C for 15 min, followed by two phenol:chloroform:isoamyl alcohol (25:24:1) extractions and precipitation with 3 M sodium acetate and cold 100% ethanol. RNA concentration and purity were determined before and after DNAseI treatment using a NanoDrop^{TM} spectrophotometer ND-1000 (Thermo Scientific), and RNA integrity was verified in 1% agarose gel electrophoresis.

cDNAs were synthesized by adding 50 µM Oligo (dT_{24}) primer and 10 mM each deoxyribonucleoside 5′-triphosphate (dNTPs) to 1 µg total RNA. This mixture was incubated at 65°C for 5 min, and briefly chilled on ice. First Strand Buffer, 20 mM dithiothreitol, and 200 units superscript III (Invitrogen) were added to the prior mixture and the total volume (20 µl) was incubated at 50°C for 1 h following manufacturer's instructions. Inactivation of the reverse transcriptase was done by incubating the mixture at 70°C for 15 min.

Quantitative real-time PCR and data analysis

Six of the eight putative coffee reference genes evaluated in this work, *AP47* (clathrin-associated protein), *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase), *PP2A* (protein phosphatase), *S24* (ribosomal protein), *UBI9* (ubiquitin-like protein), and *psaB* (photosystem subunit), were selected from the online Sol Genomics Network (Mueller et al. 2005) and HarvEST Coffee (Lin et al. 2005) databases according to the level of DNA sequence similarity to genes from *Arabidopsis thaliana* and *Coffea arabica* (Table 1; Czechowski et al. 2005; Fernandez et al. 2004). BlastN was used for this

Table 1 Coffee candidate reference genes used in this study

Gene abbreviation[a]	Coffee source gene[b]	Ortholog locus	Plant	Annotation	Similarity (*E*-value)	Identity (%)	Gene size[c]	Blast alignment
psaB	Contig 14672[d]	ATCG00340	*A. thaliana*	D1 subunit of photosystem I and II reaction centers	0.00	92	2,201	477
PP2A	SNG-U358147	AT1G13320	*A. thaliana*	Protein phosphatase 2A subunit A3	3e-84	82	1,764	533
AP47	SNG-U359260	AT5G46630	*A. thaliana*	Clathrin adaptor complexes medium subunit	1e-104	82	1,317	615
GAPDH[e]	SNG-U347734	AT1G13440	*A. thaliana*	Glyceraldehyde-3-phosphate dehydrogenase C-2	1e-139	81	1,017	968
S24[e]	SNG-U349723	AT3G04920	*A. thaliana*	40S ribosomal protein S24	2e-75	84	402	389
UBI9	SNG-U350425	NA	*C. arabica*	Ubiquitin-like protein UBI9	NA	NA	NA	NA

NA Not applicable

[a] All coffee sequences were named according to the most similar ortholog locus (*psaB*, *PP2A*, *AP47*, *GAPDH* and *S24* from *Arabidopsis thaliana*; *UBI9* from coffee)

[b] Accession number of the most similar contig according to the SOL Genomics Network (*PP2A*, *AP47*, *GAPDH*, *S24*, and *UBI9*; Mueller et al. 2005)

[c] Size in base pairs (bp) of the coding sequence of the ortholog locus in *A. thaliana*

[d] HarvEST ID for the coffee gene *psa*, this gene was not found in the SOL Genomics Network (Lin et al. 2005)

[e] Coffee genes previously used as reference in RT-PCR analysis (C. Colombo, personal communication) based on their similarity to *Malus domestica* genes (Defilippi et al. 2005)

comparison with a default setting and only genes of HarvEST Coffee database with a similarity higher than 1e-139 (*E*-value) were considered as putative orthologous to the *Arabidopsis* genes. Primers were designed with Primer 3 software (Rozen and Skaletsky 2000) using as a criterion to amplify products from 80 to 100 bp with a Tm around 60°C (primer sequences are shown in Supplemental Table 1). *UBQ10* (polyubiquitin 10) and *rpl39* (large ribosomal subunit 39) were amplified using primers previously described in the literature (Salmona et al. 2008; Simkin et al. 2006; Supplemental Table 1). Candidate reference genes were amplified from cDNA. Melting curve analysis of the amplification products and gel electrophoresis analysis confirmed that the primers amplified only a single product (data not shown). Primer sets efficiencies were estimated for each experimental set by Miner software (Zhao and Fernald 2005), and the values were used in all subsequent analyses (Supplemental Table 1). Miner software pinpoints the starting and ending points of PCR exponential phase from raw fluorescence data, and estimates primer set amplification efficiencies through a nonlinear regression algorithm without the need of a standard curve.

Polymerase chain reactions were carried out in an optical 96-well plate with a Chromo 4 real-time PCR detector (BioRad) sequence detection system, using SYBR®Green to monitor dsDNA synthesis. Reactions mixture contained 10 µl diluted cDNA (1:50), 0.2 µM of each primer, 50 µM of each dNTP, 1× PCR buffer (Invitrogen), 3 mM $MgCl_2$, 1 µl SYBR®-Green I (Molecular Probes) water diluted (1:10,000), and 0.25 units Platinum Taq DNA polymerase (Invitrogen), in a total volume of 20 µl. Reaction mixtures were incubated for 5 min at 94°C, followed by 40 amplification cycles of 15 s at 94°C, 10 s at 60°C, and 15 s at 72°C. PCR efficiencies and optimal cycle threshold (Ct) values were estimated using the online real-time PCR Miner tool (Zhao and Fernald 2005). For all reference genes studied, two independent biological samples of each experimental condition were evaluated in technical triplicates.

Cycle threshold values were converted in qBase software v1.3.5 (Hellemans et al. 2007) into nonnormalized relative quantities (Q), corrected by PCR efficiency, using the formula $Q = E^{\Delta CT}$, where E is the efficiency of the gene amplification and ΔCT is the sample with the lowest expression in the dataset

minus the Ct value of the sample in question. Efficiencies values were manually inputted in qBase for each independent experimental set. These quantities were imported to geNorm v3.4 (Vandesompele et al. 2002) and Norm Finder (Andersen et al. 2004) analysis tools, which were used as described in their manuals. Data of biological replicates were analyzed separately in both programs.

Results

In order to compare the expression levels of target genes in different tissues at the same time, it is crucial to normalize all the samples by the same set of reference genes. In the present study, eight coffee candidate reference genes were evaluated for gene expression stability in four different experimental sets. The first and second experimental set were composed of leaves and roots of control and drought-stressed "Catuaí Vermelho" cultivar, respectively. The third experimental set was composed of leaves of three different *C. arabica* cultivars that are extensively cultivated in Brazil and South America ("Catuaí Vermelho," "Mundo Novo," and "Bourbon Amarelo"). Finally, in the fourth experimental set, different organs of "Catuaí Vermelho" cultivar were compared (leaves, stems, roots, and mature fruits in the cherry phase).

Three criteria were used to select the genes evaluated in this work: traditional coffee reference genes (*rpl39* and *UBQ10*; Lepelley et al. 2007; Pre et al. 2008; Privat et al. 2008; Salmona et al. 2008; Simkin et al. 2006, 2008); coffee homologues to the top 100 reference genes of *Arabidopsis* selected by gene comparison in HarvEST Coffee database (*psaB*, *PP2A* and *AP47*; Czechowski et al. 2005), and genes previously tested in coffee by RT-PCR analysis (*S24* and *GAPDH*; C. Colombo, personal communication). *S24*, named here according to its similarity with the *Arabidopsis thaliana* gene, and *GAPDH* were initially selected by homology with reference genes used in qPCR analysis in apple (Defilippi et al. 2005). *psaB* and *PP2A* are, according to microarray analysis, ranked among the most stable genes in shoot abiotic stress series, while *PP2A* and *AP47* are ranked among the most stable genes in root abiotic stress series (Czechowski et al. 2005). These abiotic stress series included samples of a wide range of environmental conditions, e.g., cold, osmotic, salt, drought, genotoxic, oxidative, ultraviolet B (UV-B), wounding, and heat stress time courses. In addition, we selected genes belonging to different functional classes, based on *Arabidopsis* sequence information, reducing the chances of the occurrence of coregulated expression among these genes (Table 1).

Primer efficiencies for all primer combinations were higher than 0.90 (90%) in all experimental sets. However, the same primer pair showed different efficiencies with different samples. *PP2A*, for example, was amplified with an efficiency of 0.99 in the third experimental set and showed an efficiency of 0.92 in the fourth experimental set (Supplemental Table 1). Ct values were in the range of 14.86 and 30.04 (Table 2). *psaB* and *UBQ10* showed the lowest Ct values in all sets, suggesting that these genes are highly expressed, followed by *rpl39* and *GAPDH*. *AP47* together with *PP2A* presented the lowest RNA levels in all samples, while *S24* and *UBI9* presented an intermediate expression.

According to geNorm, *UBI9* ranked as one of the two most stable genes in almost all experimental sets (Table 3). *UBI9* and *S24* were indicated as the two most stable genes in the first and fourth experimental sets, when leaves of drought-stressed versus control plants and different organs of "Catuaí Vermelho" were compared, respectively. The genes *UBI9* and *PP2A* came out as the most stable reference genes when comparing roots of drought-stressed roots versus control plants. Together with *AP47*, *PP2A* was also ranked as one of the two most stable genes in the third experimental set. In order to determine the optimal number of reference genes in each experimental set, pairwise variation ($V_{n/n+1}$) was calculated by geNorm. The value of $V_{2/3}$ for all experimental sets was smaller than the cut-off threshold of 0.15, below which the inclusion of another reference gene has no significance, indicating that the use of two reference genes is sufficient for normalization in all experimental data sets tested (Table 3; Vandesompele et al. 2002). However, Vandesompele and collaborators recommend the use of at least three reference genes whenever this result obtained in our analysis is observed (Vandesompele et al. 2002).

geNorm and Norm Finder results matched in the analysis of leaves of drought-stressed and control plants (Tables 3, 4). In addition, two out of the three

Table 2 Average cycle threshold (Ct) ± standard deviation (±SD) values of the reference genes in biological replicates of the listed treatment/organ combination

Cultivar/treatment/organ	Ct (±SD)							
	psaB	*PP2A*	*AP47*	*GAPDH*	*S24*	*UBI9*	*UBQ10*	*rpl39*
"*Catuaí Vermelho IAC44*"/Control/Leaves	16.11 (±0.56)	23.63 (±0.66)	25.07 (±0.73)	18.95 (±0.48)	20.90 (±0.32)	20.64 (±0.48)	17.73 (±0.52)	19.98 (±0.38)
"*Catuaí Vermelho IAC44*"/Control/Stems	17.24 (±0.38)	23.15 (±1.52)	23.86 (±1.08)	18.54 (±0.75)	19.74 (±0.45)	19.92 (±0.69)	18.90 (±1.47)	19.13 (±0.67)
"*Catuaí Vermelho IAC44*"/Control/Roots	19.03 (±1.91)	21.43 (±0.43)	23.20 (±0.83)	17.99 (±0.40)	19.48 (±0.46)	19.32 (±0.36)	16.90 (±0.60)	19.04 (±1.88)
"*Catuaí Vermelho IAC44*"/Control/Fruits	20.05 (±0.83)	26.48 (±3.56)	26.04 (±1.70)	22.03 (±2.60)	21.42 (±0.77)	21.26 (±0.76)	20.63 (±2.52)	20.06 (±1.23)
"*Catuaí Vermelho IAC44*"/Drought/Leaves	18.18 (±0.21)	23.48 (±0.54)	24.62 (±0.16)	20.42 (±0.16)	21.40 (±0.22)	21.13 (±0.50)	16.98 (±0.37)	20.26 (±0.17)
"*Catuaí Vermelho IAC44*"/Drought/Roots	20.11 (±0.44)	25.05 (±0.99)	25.48 (±0.28)	21.58 (±1.53)	22.26 (±1.34)	22.35 (±0.89)	19.03 (±1.62)	19.31 (±0.52)
"*Bourbom Amarelo IAC J10*"/Control/Leaves	15.61 (±0.75)	22.60 (±0.98)	23.96 (±0.86)	19.09 (±0.87)	20.60 (±0.86)	19.61 (±0.40)	17.75 (±0.48)	19.18 (±0.36)
"*Mundo Novo 379/19*"/Control/Leaves	15.97 (±0.44)	22.41 (±0.17)	23.93 (±0.07)	19.36 (±0.14)	20.60 (±0.50)	20.15 (±0.19)	16.83 (±0.73)	19.24 (±1.06)

most stable genes ranked by both programs matched in the second and third experimental data sets, *S24*/*UBI9* and *AP47*/*GAPDH*, respectively. However, when different organs of "Catuaí Vermelho" were evaluated, only *UBI9* was among the three most stable genes selected by geNorm and Norm Finder (Table 4).

Norm Finder takes into account intra- and inter-group variations for normalization factor (NF) calculations, thus the best combination of reference genes (to reach the appropriate NF) is not necessarily the one containing the most stable genes according to a gene-by-gene comparison. This was observed in the first, third, and fourth experimental sets, in which the best gene combinations recommended by Norm Finder were *GAPDH*/*UBQ10*, *AP47*/*GAPDH*, and *UBQ10*/*S24*, respectively, (Table 4). Although these genes were not the same recommended by geNorm, all genes are above geNorm's cut-off limit of 1.5 (Vandesompele et al. 2002), which supports their use as reference genes (Table 3).

Discussion

Differences were observed in geNorm and Norm Finder evaluation of the best sets of reference genes for each experimental condition tested, although this inconsistency between the two methods was already expected given that they are based on distinct statistical algorithms. We elected the Norm Finder as the preferential method for the selection of the best references genes since it considers intra- and intergroup variations for the NF. When one or two of the genes indicated by Norm Finder to compose the best combination of genes were not among the three most stable genes ranked by geNorm and/or Norm Finder, we considered that the inclusion of a third gene to the set would give additional support to the NF. We used the expression stability calculated by geNorm and Norm Finder to select the additional gene (Table 5). Although, it is important to highlight that the use of the additional genes is optional.

Our analysis has shown that each experimental condition tested demands a specific set of reference genes. This result emphasizes the importance of reference genes validation for each experimental condition, especially when samples belong to very

Table 3 Candidate genes ranked according to their expression stability estimated using the geNorm algorithm after stepwise exclusion of the least stable reference gene

Drought-stressed leaves		Drought-stressed roots		Leaves different cultivars		Different organs	
Ranking	Stability value (*M*)	Ranking	Stability value (*M*)	Ranking	Stability value (*M*)	Ranking	Stability value (*M*)
S24	0.227	*PP2A*	0.113	*AP47*	0.209	*S24*	0.224
UBI9	0.227	*UBI9*	0.113	*PP2A*	0.209	*UBI9*	0.224
rpl39	0.338	*S24*	0.433	*GAPDH*	0.381	*rpl39*	0.389
PP2A	0.566	*UBQ10*	0.440	*S24*	0.441	*AP47*	0.685
UBQ10	0.659	*GAPDH*	0.499	*psaB*	0.495	*GAPDH*	1.250
AP47	0.651	*AP47*	1.134	*UBI9*	0.545	*UBQ10*	1.164
GAPDH	1.172	*psaB*	2.295	*UBQ10*	0.918	*psaB*	1.357
psaB	1.113	*rpl39*	2.469	*rpl39*	0.918	*PP2A*	1.424
$V_{2/3}$	0.110	$V_{2/3}$	0.144	$V_{2/3}$	0.126	$V_{2/3}$	0.124

Stability values are listed from most stable to least stable

Table 4 Candidate genes ranked according to their expression stability as determined by Norm Finder

Water-stressed leaves		Water-stressed roots		Leaves different cultivars		Different organs	
Ranking	Stability value	Ranking	Stability value	Ranking	Stability value	Ranking	Stability value
S24	0.134	*AP47*	0.285	*AP47*	0.125	*GAPDH*	0.387
UBI9	0.220	*S24*	0.487	*psaB*	0.132	*UBI9*	0.395
rpl39	0.240	*UBI9*	0.503	*GAPDH*	0.136	*AP47*	0.408
PP2A	0.450	*PP2A*	0.543	*UBI9*	0.171	*S24*	0.436
UBQ10	0.495	*UBQ10*	0.725	*PP2A*	0.192	*rpl39*	0.503
AP47	0.534	*GAPDH*	0.922	*S24*	0.217	*UBQ10*	0.523
GAPDH	0.539	*psaB*	1.117	*UBQ10*	0.288	*PP2A*	0.672
psaB	0.714	*rpl39*	1.446	*rpl39*	0.430	*psaB*	0.725
Best combination	Stability value	Best combination	Stability value	Best combination	Stability value	Best combination	Stability value
GAPDH/UBQ10	0.110	*AP47/S24*	0.232	*AP47/GAPDH*	0.095	*UBQ10/S24*	0.240

Stability values are listed from most stable to least stable

Table 5 Best combination of genes based on geNorm and Norm Finder expression stability values

Experimental sets			
I Drought-stressed leaves	II Drought-stressed roots	III Different cultivars	IV Different organs
GAPDH	*AP47*	*AP47*	*UBQ10*
UBQ10	*S24*	*GAPDH*	*S24*
S24	*UBI9*		*UBI9*

different groups, e.g., different organs. When leaves of control and drought-stressed plants were compared, *GAPDH*, *UBQ10*, and *S24* were considered the most appropriate reference genes. *GAPDH* and *UBQ10* should avoid error transferences since they were chosen by Norm Finder as the best combination of genes, while *S24* was ranked by Norm Finder as the most stable gene and by geNorm as one of the two most stable genes. *AP47*/*S24* and *UBI9* were assigned as good reference genes when roots of drought-stressed and control plants were tested. *AP47* and *S24*

were chosen by Norm Finder as the best combination of two genes, while *UBI9* (and also *S24*) was ranked by both programs among the three most stable genes, conferring higher robustness to the NF. Our analyses of different coffee cultivar leaves revealed that *AP47* and *GAPDH* are the most acceptable genes for gene expression normalization, since *AP47* and *GAPDH* are ranked by both algorithms as either the first or the third most stable genes, respectively. They also represent the best combination of genes considered by Norm Finder to improve NF. Finally, *UBI9*/*S24* and *UBQ10* were considered the best combination to the fourth experimental set. *UBI9* and *S24* were considered the most stable genes according to geNorm, and *S24*/*UBQ10* was chosen as the best combination of genes by Norm Finder.

psaB, *PP2A*, and *AP47* were identified as novel references genes in *A. thaliana* through microarray experiments. When compared with *GAPDH* and *UBQ10*, two traditional reference genes for several plant species including coffee, *PP2A* and *AP47* presented superior stability values calculated by geNorm (Czechowski et al. 2005). Although the stability of *psaB* has only been evaluated by microarray analysis in *Arabidopsis*, its putative coffee homologous gene showed a higher stability in leaves of distinct cultivars than common reference genes such as *UBI9*, *UBQ10*, and *rpl39*. However, the use of *psaB*, which encodes for a photosystem protein, as a reference gene, should be restricted only to qPCR analysis of green tissue samples.

Salmona et al. (2008) showed that *UBQ10* is the most stable reference gene among the seven candidate reference genes tested for normalization of qPCR experiments when coffee samples during seed development are evaluated. We consider that our results combined with the previous results from Salmona et al. (2008) can provide a comprehensive set of reference genes for qPCR analysis in coffee. *rpl39* has been extensively used as an internal control when comparing gene expression profiles of target genes among different coffee organs, e.g., leaves, stems, fruits, branches, roots, and flowers (Lepelley et al. 2007; Pre et al. 2008; Privat et al. 2008; Simkin et al. 2006, 2008) and also when comparing control and drought-stressed leaves (Simkin et al. 2008). Our work revealed that *rpl39* is not the most accurate reference gene when comparing control and drought-stressed leaves and also different coffee organs.

The two programs employed in our study to evaluate reference genes (geNorm and Norm Finder) use the same input data, i.e., nonnormalized relative quantities, and Ct values need to be transformed considering primer pairs efficiencies. In our experience, it is crucial to evaluate primer pairs efficiencies for each sample tested since the same primer pair could display different efficiencies. The importance of this step can be well illustrated by the primer efficiency variation of *PP2A* in leaves of different cultivars compared with different organs (Supplemental Table 1). The values of Ct presented herein should not be considered alone, but they may help in the selection of best combination of reference genes when there is previous data about target gene expression levels. Similar expression levels of the reference and target genes are considered an important issue regarding qPCR normalization (Vandesompele et al. 2002). Indeed, references genes with excessively high/low expression levels compared with target genes can trigger problems for data analysis (Frost and Nilsen 2003; Robinson et al. 2007). As suggested by Remans et al. (2008), biological replicates were submitted to geNorm and Norm Finder as independent samples. This procedure increased the credibility of the most suitable coffee reference genes because it takes into account possible variations in reference gene expression that are not due to different treatments, but intrinsic to the gene itself.

We consider that the genes evaluated in this study will be very useful for future evaluations of gene expression analysis when studying drought-stressed leaves and roots, leaves of different cultivars, and when comparing different organs of coffee. Moreover, this study provides useful guidelines for reference gene selection for researchers working with coffee.

Acknowledgements We are indebted to Fatima Barbosa, Luiz Frade, and Sarah Muniz Nardeli for technical assistance; to Bruna Matta M.Sc. for help with data management and analysis; to Adriana Martinelli Ph.D., Erica D.,Silveira M.Sc., and Ute Achenbach Ph.D. for comments on the manuscript; and to Eduardo Arcoverde Ph.D. for help with the Scholander-type pressure chamber. This work is part of Fernanda Cruz's Ph.D. thesis in Department of Genetics, Federal University of Rio de Janeiro, Brazil. This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPQ (grants no. 310254/2007-8 to M.A.-F.), the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro—FAPERJ (grant no. E-26/102.861/2008 to M.A.-F.), the International Foundation of Science (grant no.C/3962-1 to M.A.-F.), and the International Basic Science Program (grant no. IBSP/UNESCO-3-BR-28 to M.A.-F.).

References

Andersen CL, Jensen JL, Orntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64:5245–5250. doi:10.1158/0008-5472.CAN-04-0496

Brunner A, Yakovlev I, Strauss S (2004) Validating internal controls for quantitative plant gene expression studies. BMC Plant Biol 4:14. doi:10.1186/1471-2229-4-14

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol 139:5–17. doi:10.1104/pp.105.063743

Defilippi BG, Kader AA, Dandekar AM (2005) Apple aroma: alcohol acyltransferase, a rate limiting step for ester biosynthesis, is regulated by ethylene. Plant Sci 168:1199–1210. doi:10.1016/j.plantsci.2004.12.018

Etienne H, Anthony F, Dussert S, Fernandez D, Lashermes P, Bertrand B (2002) Biotechnological applications for the improvement of coffee (*Coffea arabica* L.). In Vitro Cell Dev Biol Plant 38:129–138. doi:10.1079/IVP2001273

FAOSTAT (2008) http://faostat.fao.org/site/406/default.aspx. Accessed date 7 Oct 2008

Fernandez D, Santos P, Agostini C, Bon MC, Petitot AS, Silva MC, Guerra-Guimaraes L, Ribeiro A, Argout X, Nicole M (2004) Coffee (*Coffea arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*). Mol Plant Pathol 5:527–536. doi:10.1111/j.1364-3703.2004.00250.x

Frost P, Nilsen F (2003) Validation of reference genes for transcription profiling in the salmon louse, Lepeophtheirus salmonis, by quantitative real-time PCR. Vet Parasitol 118:169–174. doi:10.1016/j.vetpar.2003.09.020

Gachon C, Mingam A, Charrier B (2004) Real-time PCR: what relevance to plant studies? J Exp Bot 55:1445–1454. doi:10.1093/jxb/erh181

Ganesh D, Petitot AS, Silva MC, Alary R, Lecouls AC, Fernandez D (2006) Monitoring of the early molecular resistance responses of coffee (*Coffea arabica* L.) to the rust fungus (*Hemileia vastatrix*) using real-time quantitative RT-PCR. Plant Sci 170:1045–1051. doi:10.1016/j.plantsci.2005.12.009

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8:R19. doi:10.1186/gb-2007-8-2-r19

Huggett J, Dheda K, Bustin S, Zumla A (2005) Real-time RT-PCR normalisation; strategies and considerations. Genes Immun 6:279–284. doi:10.1038/sj.gene.6364190

Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J (2007) Chlorogenic acid synthesis in coffee: an analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. Plant Sci 172:978–996. doi:10.1016/j.plantsci.2007.02.004

Lin CW, Mueller LA, Mc Carthy J, Crouzillat D, Petiard V, Tanksley SD (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theor Appl Genet 112:114–130. doi:10.1007/s00122-005-0112-2

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond. plant physiol 138:1310–1317. doi:10.1104/pp.105.060707

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: bestkeeper—excel-based tool using pair-wise correlations. Biotechnol Lett 26:509–515. doi:10.1023/B:BILE.0000019559.84305.47

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. Mol Genet Genomics 276:436–449. doi:10.1007/s00438-006-0153-5

Pre M, Caillet V, Sobilo J, McCarthy J (2008) Characterization and expression analysis of genes directing galactomannan synthesis in coffee. Ann Bot (London) 102:207–220. doi:10.1093/aob/mcn076

Privat I, Foucrier S, Prins A, Epalle T, Eychenne M, Kandalaft L, Caillet V, Lin C, Tanksley S, Foyer C, McCarthy J (2008) Differential regulation of grain sucrose accumulation and metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) revealed through gene expression and enzyme activity analysis. New Phytol 178:781–797. doi:10.1111/j.1469-8137.2008.02425.x

Reid KE, Olsson N, Schlosser J, Peng F, Lund ST (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol 6:27. doi:10.1186/1471-2229-6-27

Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A (2008) Normalisation of real-time RT-PCR gene expression measurements in Arabidopsis thaliana exposed to increased metal concentrations. Planta 227:1343–1349. doi:10.1007/s00425-008-0706-4

Robinson TL, Sutherland IA, Sutherland J (2007) Validation of candidate bovine reference genes for use with real-time PCR. Vet Immunol Immunopathol 115:160–165. doi:10.1016/j.vetimm.2006.09.012

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Salmona J, Dussert S, Descroix F, de Kochko A, Bertrand B, Joet T (2008) Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. Plant Mol Biol 66:105–124. doi:10.1007/s11103-007-9256-6

Schmittgen TD, Zakrajsek BA (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods 46:69–81. doi:10.1016/S0165-022X(00)00129-9

Simkin AJ, Qian TZ, Caillet V, Michoux F, Ben Amor M, Lin CW, Tanksley S, McCarthy J (2006) Oleosin gene family of *Coffea canephora*: quantitative expression analysis of

five oleosin genes in developing and germinating coffee grain. J Plant Physiol 163:691–708. doi:10.1016/j.jplph.2005.11.008

Simkin AJ, Moreau H, Kuntz M, Pagny G, Lin C, Tanksley S, McCarthy J (2008) An investigation of carotenoid biosynthesis in *Coffea canephora* and *Coffea arabica*. J Plant Physiol 165:1087–1106. doi:10.1016/j.jplph.2007.06.016

Udvardi MK, Czechowski T, Scheible W-R (2008) Eleven golden rules of quantitative RT-PCR. Plant Cell 20:1736–1737. doi:10.1105/tpc.108.061143

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:7. doi:10.1186/gb-2002-3-7-research0034

Vieira LGE, Andrade AC, Colombo CA, AHdA Moraes, Metha Â, ACd Oliveira, Labate CA, Marino CL, CdB Monteiro-Vitorello, DdC Monte, Giglioti É, Kimura ET, Romano E, Kuramae EE, Lemos EGM, ERPd Almeida, Jorge ÉC, Albuquerque ÉVS, FRd Silva, Vinecky F, Sawazaki HE, Dorry HFA, Carrer H, Abreu IN, Batista JAN, Teixeira JB, Kitajima JP, Xavier KG, LMd Lima, LEAd Camargo, Pereira LFP, Coutinho LL, Lemos MVF, Romano MR, Machado MA, MMdC Costa, MFGd Sá, Goldman MHS, Ferro MIT, Tinoco MLP, Oliveira MC, Van Sluys M-A, Shimizu MM, Maluf MP, MTSd Eira, Guerreiro Filho O, Arruda P, Mazzafera P, Mariani PDSC, RLBCd Oliveira, Harakava R, Balbao SF, Tsai SM, SMZd Mauro, Santos SN, Siqueira WJ, Costa GGL, Formighieri EF, Carazzolle MF, Pereira GAG (2006) Brazilian coffee genome project: an EST-based genomic resource. Braz J Plant Physiol 18:95–108

Zhao S, Fernald RD (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. J Comput Biol 12:1047–1064. doi:10.1089/cmb.2005.12.1047