

Establishing references for gene expression analyses by RT-qPCR in *Theobroma cacao* tissues

T.T. Pinheiro¹, C.G. Litholdo Jr.¹, M.L. Sereno¹, G.A. Leal Jr.¹,
P.S.B. Albuquerque² and A. Figueira¹

¹Centro de Energia Nuclear na Agricultura,
Universidade de São Paulo, Piracicaba, SP, Brasil

²ERJOH, Comissão Executiva do Plano da Lavoura Cacaueira,
Marituba, PA, Brasil

Corresponding author: A. Figueira
E-mail: figueira@cena.usp.br

Genet. Mol. Res. 10 (4): 3291-3305 (2011)

Received February 17, 2011

Accepted April 19, 2011

Published November 17, 2011

DOI <http://dx.doi.org/10.4238/2011.November.17.4>

ABSTRACT. Lack of continuous progress in *Theobroma cacao* (Malvaceae) breeding, especially associated with seed quality traits, requires more efficient selection methods based on genomic information. Reverse transcript quantitative PCR (RT-qPCR) has become the method of choice for gene expression analysis, but relative expression analysis requires various reference genes, which must be stable across various biological conditions. We sought suitable reference genes for various tissues of cacao, especially developing seeds. Ten potential reference genes were analyzed for stability at various stages of embryo development, leaves, stems, roots, flowers, and pod epicarp; seven of them were also evaluated in shoot tips treated either with hormones (salicylate; ethefon; methyl-jasmonate) or after inoculation with the fungus *Moniliophthora perniciosa* (Marasmiaceae *sensu lato*). For developing embryos, the three most stable genes were actin (ACT), polyubiquitin (PUB), and ribosomal protein L35 (Rpl35). In the analyses of various tissues, the most stable genes were malate dehydrogenase (MDH), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and acyl-carrier protein B (ACP B). GAPDH, MDH and

tubulin (TUB) were the most appropriate for normalization when shoot apices were treated with hormones, while ACT, TUB and Rpl35 were the most appropriate after inoculation with *M. perniciosa*. We conclude that for each plant system and biological or ontogenetical condition, there is a need to define suitable reference genes. This is the first report to define reference genes for expression studies in cacao.

Key words: Cocoa; Cocoa butter; Fatty acid biosynthesis; Gene expression; Sterculiaceae

INTRODUCTION

Theobroma cacao L. (cacao; Malvaceae *sensu lato*) is an important tropical crop as its seeds are the unique source of raw materials for chocolate manufacturing (Pires et al., 1998). Moreover, cocoa butter is a specialty fat obtained from cacao seeds with highly specific physicochemical attributes, valued by the pharmaceutical and cosmetic industries (Silva and Figueira, 2005). The average rate of increase in demand for cacao seeds has been ca. 3%/year for the last 100 years (World Cocoa Foundation, 2010). However, increasing threats of disease outbreaks and lack of continuous progress in cacao breeding, specially associated with seed quality traits (such as cocoa butter content and hardness, and cocoa flavor), require the implementation of more efficient methods of selection (Bennett, 2003). New approaches based on genomic information offer the potential to be adopted in precision breeding schemes. The availability of genomic information has increased drastically, with the genome sequence completed (Argout et al., 2011; <http://www.cacaogenomedb.org/>), and over 150,000 expressed sequence tags (EST) available at public databases, representing 42,000 unique sequences (Verica et al., 2004; Leal Jr. et al., 2007; Argout et al., 2008). A major challenge is to assign functions to these sequences, which will require functional studies involving analysis of gene expression and gene manipulation by complementation, over-expression or silencing (Guiltinan et al., 2008). Accordingly, gene expression has been investigated under distinct biological conditions and tissues of *T. cacao* (Leal et al., 2007; Bae et al., 2008; Mondego et al., 2008).

The development of effective methods to quantify gene expression has allowed the investigation of many genes in various biological systems (Donson et al., 2002), and quantitative amplification of reversed transcripts (RT-qPCR) has become the method of choice for analyses of expression, due to the sensitivity, specificity, and relatively speediness (Gachon et al., 2004). However, quantification of gene expression by RT-qPCR requires a series of standardizations, concerning the amount, integrity and quality of the original RNA sample, the efficacy of reverse transcription and amplification, among others (Bustin et al., 2009). For relative gene expression, it is required to adopt internal gene references, which must be stable across various biological and ontogenetic conditions, to normalize test conditions (Bustin et al., 2009). Reference genes frequently employed for relative quantification by RT-qPCR are those encoding proteins responsible for cellular structure and basic metabolism ('housekeeping genes'), such as tubulin (Jeong et al., 2006); actin (ACT; Liu et al., 2010); ubiquitin (Brunet et al., 2009); elongation factors (Bomal et al., 2008); glyceraldehyde 3-phosphate dehydrogenase (GAPDH; Magneschi et al., 2009); ribosomal proteins (Calsa Jr. and Figueira, 2007), among others. The expression of housekeeping genes should not drastically vary among tissues or changes in

the environment, ontogenetical stages, or under various biological treatments. However, the stability of gene expression can fluctuate considerably, and thus, it is advisable to investigate genes with the least variation in expression for each experimental condition to define suitable references. The study of reference gene stability has been conducted for many species under various biological conditions, demonstrating the need for specific evaluation for each type of experiment, especially in actively growing tissues, such as developing embryos (Reid et al., 2006; Hong et al., 2008; Barsalobres-Cavallari et al., 2009; Carvalho et al., 2010).

Various statistical approaches have been developed to choose the most suitable genes for relative expression by RT-qPCR, and softwares such as geNorm (Vandesompele et al., 2002) and NormFinder (Andersen et al., 2004) have been widely adopted to establish suitable reference genes. geNorm is a freeware developed for Microsoft Excel[®], which determines the most stable reference gene among various candidates, and the ideal number of reference genes for each experiment. For all candidate reference genes, a pairwise variation is estimated, analyzing the deviation in expression for all possible combination of gene pairs. Gene stability is expressed as an 'M value', which is inversely proportional to the variation in expression for a given gene (Vandesompele et al., 2002). NormFinder estimates the variation in expression within and among groups of experimental conditions, which refers to the source of biological variation among treatments, expressing stability values for each candidate reference gene. More stable gene expression is indicated by lower average expression stability values (Andersen et al., 2004).

To our knowledge, a systematic selection of the most appropriate reference gene for *T. cacao* in RT-qPCR experiments has not yet been conducted. Therefore, the objective of this study was to establish reference genes for various embryo developmental stages and tissues of *T. cacao*, evaluating 10 candidates, including ACT; elongation factor-1 α (EF-1 α); GAPDH; malate dehydrogenase (MDH); polyubiquitin (PUB); ribosomal protein L35 (Rpl35); β -tubulin 5 (TUB), and three isoforms of the acyl-carrier protein (ACP A; ACP B; ACP C). The stability of the 10 reference genes was investigated in various stages of embryo development (site of cocoa butter synthesis and other quality attributes), leaves, stems, roots, flowers, and pod epicarp, while seven of them were further tested in shoot tips treated with hormones [salicylate (AS); ethephon (ET); methyl-jasmonate (MJ)] or after inoculation with *Moniliophthora perniciosa* (causal agent of the witches' broom disease).

MATERIAL AND METHODS

Plant material

Embryos were obtained from pods derived from controlled manual pollination from the accession 'PA 16', collected 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160 days after pollination in Marituba, PA, Brazil. Cacao seedlings from an undefined genotype collected at ESALQ/USP, Piracicaba, SP, Brazil, were grown under nursery conditions for around 5 months to collect leaves at various stages of development, stem, roots, and shoot apex. Samples of flowers and pods to collect epicarp were obtained from field-grown trees from an undefined accession at ESALQ/USP, Piracicaba. For the experiment using application of hormones, shoot tips from genotype 'Pound 7' were collected 1 h after individual treatment consisting of spraying aqueous solutions of 7.25 mM AS; 0.1 mM ET, or 0.12 mM MJ. For

fungal inoculation, 30 μ L of a 10^5 /mL basidiospore suspension of *M. pernicioso* was applied to shoot tips from clonal plants of the 'CAB214' genotype. After inoculation, plants were kept in a humid chamber for 24 h, and samples were collected 6 and 48 h after inoculation. The collected tissue samples were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction.

Search for reference gene sequences and primer design

Sequences for the candidate reference genes coding for ACP A (GenBank CU536770); ACP B (CU536770); ACP C (CU536770); EF-1 α (CU534208); MDH (CU609279), and PUB (CU525140) were obtained from the cacao EST collection as described by Argout et al. (2008). The identity of all the recovered sequences was firstly confirmed by BLASTn and tBLASTn analyses. Primers for GAPDH were designed based on the 'TC (Tentative Consensus) 385' sequence from *T. cacao* Gene Index database (<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=cocoa>). Primers for ACT were as proposed by Leal et al. (2007), whereas primers for Rpl35 and TUB have been previously used in our laboratory (Leal GA Jr. and Figueira A, unpublished data). Primers for EF-1 α , MDH, PUB, Rpl35, and TUB were designed based on a single-read sequence. The recovered sequences of ACP A, ACP B and ACP C were multi-aligned with homologues from other species using Bioedit (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>) and ClustalW, and specific primers were developed for each gene located at conserved regions. Primers were developed using Primer3 (Rozen and Skaletsky, 2000), with quality analyzed using NetPrimer (<http://www.premierbiosoft.com/netprimer>).

RNA extraction and cDNA synthesis

Total RNA was extracted from developing embryos and from all other cacao tissues (expanding and mature leaves; stem; root; shoot apex; flowers; pod epicarp) using the protocol as described by Verica et al. (2004), with a few modifications. Briefly, 100 mg cacao tissues (including embryos) was ground under liquid nitrogen and extracted with buffer [2% CTAB (w/v); 2 M NaCl; 100 mM Tris, pH 8.0; 25 mM EDTA, pH 8.0; 2% PVP 10,000 (w/v); 2% β -mercaptoethanol (v/v), and 0.5 g/L spermidine (w/v)]. Samples were then extracted with chloroform:isoamyl alcohol, and total RNA was precipitated by the addition of 8 M LiCl for 12 h at 4°C . After centrifugation at 8000 g for 30 min at 4°C , RNA was resuspended in DEPC water, and further purified using the Mini-Invisorb Spin Plant RNA Kit (Invitek, Berlin, Germany). RNA was quantified in an Ultrospec 1100 pro-spectrophotometer (Amersham Biosciences, Piscataway, NJ, USA); RNA integrity was examined by gel electrophoresis, and the absence of genomic DNA contamination was confirmed by PCR using primers to the ACT gene. Before cDNA synthesis, total RNA was treated with DNase I (Fermentas Life Sciences, Burlington, Canada) at 37°C for 30 min. cDNA was synthesized by adding 500 ng 18mer-oligo-dT to 1 μ g RNA and incubating for 5 min at 70°C , followed by cooling to 4°C for 5 min. Reverse transcription was conducted in a reaction containing 200 U Improm-II Reverse Transcriptase (Promega, Madison, WI, USA); 3 mM MgCl_2 ; 5 μ M dNTPs; 40 U RNase Out (Invitrogen, Carlsbad, CA, USA) in a thermocycler programmed for 25°C for 5 min; 42°C for 60 min, and 70°C for 15 min. Primers were tested for specificity by determining the amplicon size from enzymatic amplification (RT-PCR) of a pool of cDNA from embryos of various

stages of development, and using individual cDNAs derived from the various tissues and treatments. The identity of the amplified fragments from primers specific for ACP A, ACP B, and ACP C was confirmed after cloning and sequencing (data not shown).

Quantitative amplification of reversed transcripts

Quantitative amplification of reversed-transcribed sample RNAs was conducted in 10- μ L reactions containing 1 μ L cDNA 1:10 (v/v) dilution; 0.5 μ M of each transcript-specific primers, and 5 μ L 2X Platinum SYBR-Green RT-qPCR SuperMix-UDG (Invitrogen). Amplifications were performed in a RotorGene 3000 thermocycler (Corbett Life Science, Sidney, Australia) in triplicates, with initial incubation at 50°C for 2 min and 95°C for 2 min, followed by 40 cycles of 95°C for 15 s; 60°C for 15 s, and 72°C for 20 s, with fluorescence detection at the end of the extension cycles. After the final cycle, melting curves for each amplicon were determined between 72° and 95°C. Experiments included a negative control (no template DNA), and amplification efficiency of each primer pair was determined by standard curve using serial dilutions [10^{-1} , 10^{-2} , 10^{-3} or 1:10, 1:20, 1:40 (v/v), depending on the anticipated amount of transcripts present] of a cDNA pool from developing embryos. Real-time data acquisition was performed by the RotorGene Real-Time Analysis 6.0 (Corbett Research, Mortlake, Australia), which provided the values of cycle threshold (C_T) and PCR efficiency (E).

Reference gene stability analyses

To establish the ideal reference gene among diverse biological conditions, gene expression data were analyzed using two statistical approaches embodied into the geNorm and NormFinder softwares. Both programs use C_T values transformed into relative quantities (Q), estimated by $Q = E^{(C_{T \text{ minimum}} - C_{T \text{ sample}})}$, where E is the amplification efficiency and $C_{T \text{ minimum}}$ is the lowest C_T value (sample with more expression among all samples) (Vandesompele et al., 2002). The geNorm program estimates the mean expression stability, providing an 'M value' and the variation of stability ('V') from pairwise comparisons of multiple reference genes, gradually excluding the least suitable genes, with the largest 'M values'. On the other hand, NormFinder identifies the best reference gene by ranking all candidates to reference genes, according to their expression stability in a given sample set and experimental conditions. NormFinder calculates not only the overall variation of candidate normalization genes, but also the variation between subgroups from the sample set under investigation.

RESULTS

Selection of candidate reference genes

Based on previous study in other plant systems, the genes coding for ACT, EF-1 α , GAPDH, MDH, PUB, Rpl35, and TUB were chosen to be tested as reference. In addition, genes encoding three isoforms of ACP, indispensable co-factors for many metabolic pathways that synthesize fatty acids, phospholipids, signaling molecules, and metabolic co-factors (Byers and Gong, 2007) were selected as reference candidates as they might present basal expression levels in tissues with intense synthesis of membrane or storage lipids, such as cacao

embryos. ACT, GAPDH, Rpl35, and TUB have been previously used as reference genes in our laboratory (Leal et al., 2007). Sequences putatively coding for the other five reference genes tested (ACP B; ACP C; EF-1 α ; MDH, and PUB) were obtained from public EST databases, while the sequence for ACP A was originally identified in a cDNA library constructed in our laboratory (Silva, 2005).

Each pair of primers specific for the candidate reference genes (Table 1) was then tested by amplifying a cDNA pool derived from a mixture of RNA from developing embryos to confirm fragment size and specificity as a single amplicon. Later, all primers were tested amplifying cDNA samples derived from the various tissues. All primer combinations amplified a single fragment with the anticipated size. Fragments of ACP A, ACP B and ACP C obtained by RT-PCR were also cloned and sequenced, with identity confirmed (data not shown). Amplification efficiency for each primer pair was then determined by standard curve using serial dilutions [10^{-1} , 10^{-2} , 10^{-3} or 1:10, 1:20, 1:40 (v/v), depending on the expected amount of each specific transcript present based on C_T values] of a cDNA pool from developing embryos. The estimated efficiency for each primer pair ranged from 1.74 for MDH and PUB to 1.99 (ACP B), with all estimated R^2 values above 0.98 (Table 1).

Table 1. Candidate reference genes with primer sequences, expected fragment size (base pairs), actual amplification efficiency, and R^2 obtained from RT-qPCR.

Gene	Primer sequences	Fragment size (bp)	Efficiency	R^2
ACT	TCCTCTTCCAGCCAICTCTC TCTCCTTGCTCATTGGTCT	171	1.91	0.99
EF-1 α	AGGTCCACCAACCTTGACTG TTGGGCTCGTTAATCTGGTC	71	1.98	0.99
GAPDH	GATGCTCCTATGTTTGTGG TCTTCTCCTCTCCAGTCCTT	222	1.96	0.99
MDH	AAAATGGAGITGGTGGATGC AACCATGACTGCGATGTTGA	102	1.74	0.99
PUB	TTCAGGACAAGGAGGGGATT AGGACAAGATGAAGGGTGGA	124	1.74	0.99
Rpl35	CCTAACAAGCTTACCAAGATAAAGG TTATTCTTAATAGCCTCCCTCAGTG	107	1.85	0.98
TUB	ATTCCCCGTCTTCACTTCT TCTGCTCATCAACCTCTTTGG	212	1.95	0.99
ACP A	AAAGCTTCCCAAAAGTCGAT ATTCAACGAGACACCCTTGC	236	1.96	0.99
ACP B	GCAGACAAGATCAGCACA AAATCAAAGGGCAGACT	195	1.99	0.98
ACP C	TCTTCCCACACTCTTTTCG AGGCTGCATGGAGATTGAT	118	1.89	0.98

ACT = actin; EF-1 α = elongation factor-1 α ; GAPDH = glyceraldehyde 3-phosphate dehydrogenase; MDH = malate dehydrogenase; PUB = polyubiquitin; Rpl35 = ribosomal protein L35; TUB = β -tubulin 5; ACP = acyl-carrier protein.

Expression profile of candidate reference genes

The analyses of raw gene expression levels estimated by C_T across all samples detected large variation among reference genes (Figure 1A-D), indicating that no single reference gene had a constant expression level in the many distinct conditions investigated. Mean C_T values fluctuated from 14.6 for GAPDH to 28.9 for TUB both in samples from developing

embryos (Figure 1A). In general, PUB displayed the lowest C_T for RT-qPCR using cDNA samples from the various tissues/organs of cacao (Figure 1B) and from shoot apices treated either with hormones (AS, ET, MJ) (Figure 1C) or inoculated with *M. perniciosa* (Figure 1D). ACP C, ACT and Rpl35 showed the highest C_T values, respectively, for the same experiments (Figure 1B, C, and D).

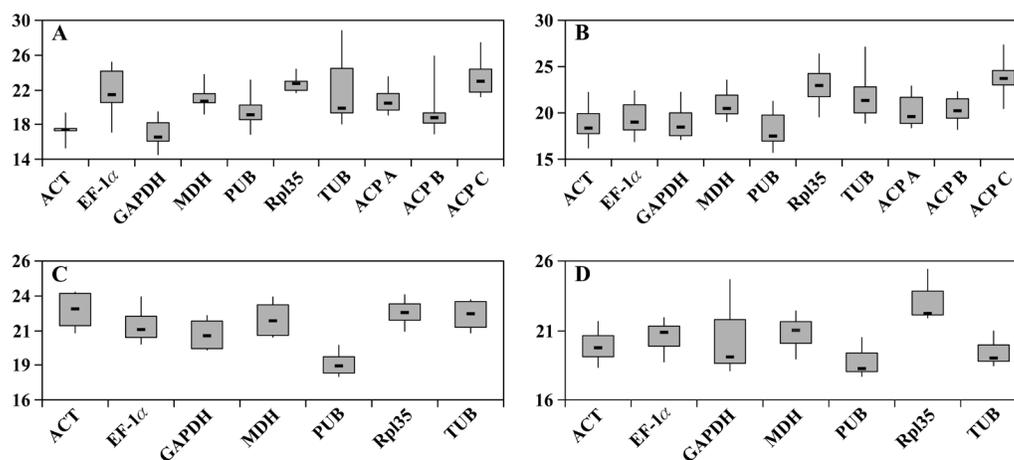


Figure 1. Box-whisker plots showing C_T variation for each candidate reference gene estimated by RT-qPCR for distinct samples. **A.** Developing embryos. **B.** Tissues/organs. **C.** Shoot tips treated with salicylate, ethephon or methyl-jasmonate. **D.** Cacao shoot tips inoculated with *Moniliophthora perniciosa*. A line across the box is depicted as median, whereas the box indicates the 25th and 75th percentiles. Whiskers represent the maximum and minimum values. For abbreviations, see legend to Table 1.

Analysis of gene expression stability among *T. cacao* tissues and conditions

Each primer pair specific for a gene was then tested for stability using all tissue samples. cDNAs obtained from RNA extracted from developing embryos and the other tissues were diluted to 1:10 (v/v) and analyzed by RT-qPCR. Stability of the reference genes was then calculated using geNorm and NormFinder based on the observed C_T values, transformed into relative quantities. For geNorm, the gene stability values were measured based on non-normalized level of expression for each experiment (developing embryos; various tissues/organs; shoot apex treated with hormones or inoculated with *M. perniciosa*). This software calculates all variations from pairwise comparison among the genes, defining the 'M value' as the mean variation in relation to all other candidates. The ideal reference gene with the greatest stability would display the lowest M value.

Initially, stability of the 10 genes was evaluated for cacao developing embryos using the geNorm algorithm (Figure 2A). The results indicated that ACT and Rpl35 were the most stable pair, with an average expression stability of 0.64 (Figure 2A). The least stable gene for similar biological conditions was TUB, with an average expression stability of 2.15 (Figure 2A). Gene stability was then evaluated for the other cacao tissues/organs (expanding and mature leaves; stem; root; shoot apex; flowers; pod epicarp, and seeds). In that case, geNorm indicated GAPDH and ACP A as the most stable pair of genes, with an average expression

stability of 0.54, while ACP C as the least stable gene, with an average expression stability of 1.52 (Figure 2B). When samples derived from cacao shoot tips collected 1 h after treatment with the hormones AS, ET and MJ were used to test reference gene stability, geNorm indicated that the most stable pair of genes was MDH and GAPDH, with an average expression stability of 0.10, while EF-1 α was the least stable gene with average M = 0.61 (Figure 2C). For cacao shoot apex collected from seedlings inoculated with *M. perniciosa*, MDH and EF-1 α was the most stable gene pair, with an average expression stability of 0.23 (Figure 2D). In this case, GAPDH was the least stable gene with an average expression stability of 1.15 (Figure 2D). In samples derived from cacao shoot apex treated with hormones or inoculated with fungal pathogen, the genes coding for ACP A, ACP B, and ACP C were not evaluated as reference.

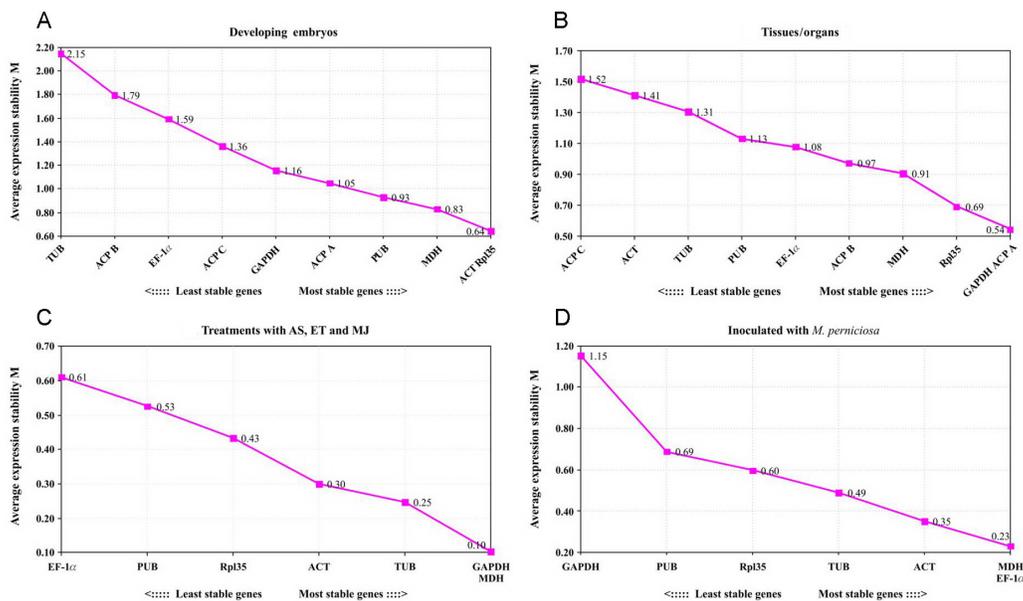


Figure 2. Average expression stability values (M) and ranking of candidate reference genes as calculated by geNorm analysis for: **A.** cacao developing embryos; **B.** tissues/organs; **C.** shoot tips treated with salicylate (AS), ethephon (ET) or methyl-jasmonate (MJ); **D.** shoot tips inoculated with *Moniliophthora perniciosa*. For abbreviations, see legend to Table 1.

It is recommended that for each expression experiment a minimum number of reference genes must be used to normalize expression of the target gene, using, at least, the three most stable references (Vandesompele et al., 2002; Huggett et al., 2005; Bustin et al., 2009). geNorm indicates the optimal number of reference genes required for accurate normalization by pairwise variation. The pairwise variation (V_n/V_{n+1}) is calculated between consecutively ranked normalization factors obtained from stepwise inclusion of the next more stable reference gene to determine the effect of adding the next reference gene in normalization (Figure 3). geNorm recommends a pairwise variation cut-off value of 0.15, below which the inclusion of an additional reference gene is not necessary (Vandesompele et al., 2002).

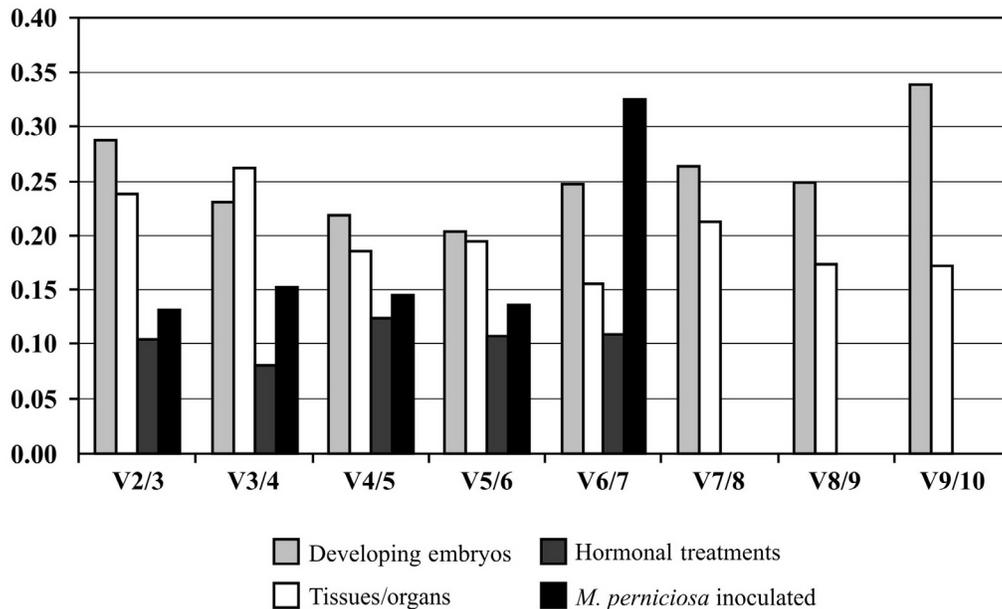


Figure 3. Determination of the optimal number of reference genes for relative expression calculated by geNorm. $V_{2/3}$ is the pairwise variation when the number of reference genes is increased from two to three. Stepwise inclusion of less stable gene generates the next data point.

For developing embryos and samples from various tissues/organs of cacao, the pairwise variation from stepwise inclusion of the next more stable reference gene was always greater than 0.15 (Figure 3). The calculated normalization factor for increasing number of reference genes up to the limit of $M < 1.5$ (Vandesompele et al., 2002) showed that the inclusion of additional genes up to seven (ACT/Rpl35, MDH, PUB, ACP A, GAPDH, and ACP C) for developing embryos, and up to nine genes (GAPDH/ACP A, Rpl35, MDH, ACP B, EF-1 α , PUB, TUB, and ACT) for tissues/organs, still significantly contributed to the variation of the normalization factor ($V_{n/n+1} > 0.15$) (Figure 3). Conversely, the pairwise variations for reference genes for samples from shoot apex treated with hormones or inoculated with *M. perniciosa* were all lower than 0.15 (Figure 3), indicating that based on the normalization factor, two genes ($V_{2/3}$) would be enough for normalizing gene expression.

The stability of the reference genes was also analyzed using NormFinder. Similar to geNorm, this program allows the estimation of expression stability of reference genes for different samples, but also between sample subgroups. Therefore, samples derived from cacao developing embryos, tissues, and from shoot apex treated with AS, ET or MJ, or inoculated with *M. perniciosa* were analyzed individually, but also as two subgroups that considered “non-treated tissues” (developing embryos and tissues/organs) and “treated tissues” (treatment with AS, ET or MJ, and inoculated with *M. perniciosa*). The analysis using comparison between subgroups offered the best gene-pair combination (Table 2).

Table 2. Candidate reference genes expression stability and stability ranking (between parentheses) according to NormFinder.

	GAPDH	Rp135	ACT	TUB	MDH	EF-1 α	PUB	ACPA	ACPB	ACPC	Best combination
Developing embryos	0.85 (6)	0.45 (3)	0.20 (1)	2.34 (10)	0.59 (5)	1.45 (8)	0.30 (2)	0.50 (4)	1.69 (9)	1.18 (7)	
Various tissues/organs	0.41 (2)	0.60 (5)	1.12 (8)	1.13 (9)	0.32 (1)	0.73 (6)	0.81 (7)	0.50 (4)	0.46 (3)	1.17 (10)	
Treatments with hormones (MJ, AS, ET)	0.04 (1)	0.38 (5)	0.29 (4)	0.13 (3)	0.04 (2)	0.52 (7)	0.50 (6)	-	-	-	
Inoculated with <i>M. perniciosa</i>	1.58 (7)	0.16 (3)	0.09 (1)	0.15 (2)	0.52 (5)	0.61 (6)	0.47 (4)	-	-	-	
Non-treated	0.46 (6)	0.51 (7)	0.42 (4)	0.73 (9)	0.41 (3)	0.83 (10)	0.43 (5)	0.23 (1)	0.37 (2)	0.60 (8)	ACP A/ACP B
Treated	0.71 (7)	0.09 (1)	0.22 (4)	0.25 (5)	0.11 (2)	0.12 (3)	0.26 (6)	-	-	-	Rp135/MDH

MJ = methyl-jasmonate; AS = salicylate; ET = ethephon. For other abbreviations, see legend to Table 1.

According to NormFinder, for developing embryos, the most stable reference gene was ACT, with estimated stability of 0.20, followed by the gene coding for PUB (0.30), and Rpl35 (0.45) (Table 2). The least stable gene for this condition was TUB, with stability of 2.34 (Table 2). When the same genes were evaluated for other tissues and organs (expanding and mature leaves, stem, root, shoot apex, flowers, and pod epicarp), NormFinder indicated that the most stable gene was MDH, with stability of 0.32, followed by GAPDH (0.41) and ACP B (0.46) (Table 2). The least stable gene identified for these cacao tissues by NormFinder was ACP C, with stability of 1.17 (Table 2). In the case of cacao shoot apices collected 1 h after treatment with AS, ET or MJ, the most stable genes were GAPDH and MDH, both with stability equal to 0.04, followed by TUB (0.13) (Table 2). For shoot apices inoculated with *M. pernicioso*, the analysis with NormFinder identified ACT as the most stable gene, with stability of 0.09, followed by TUB (0.15) and Rpl35 (0.16) (Table 2). The least stable genes were EF-1 α (0.52) for treatment with AS, ET or MJ, and GAPDH (1.58) when inoculated with *M. pernicioso* (Table 2). The most stable gene for the subgroup “non-treated tissues” was ACP A with stability of 0.23, followed by ACP B (0.37) and MDH (0.41) (Table 2), with ‘ACP A/ACP B’ as the best gene pair combination. When considering the subgroup “treated tissues”, the three most stable genes were Rpl35 (0.09), MDH (0.11) and EF-1 α (0.12), and ‘Rpl35/MDH’ was the best gene pair combination.

DISCUSSION

RT-qPCR has become the method of choice for gene expression, but evaluating expression in relative terms requires a series of normalizations by the adoption of stable reference genes across various biological and ontogenetic conditions (Gachon et al., 2004; Bustin et al., 2009). The genes encoding proteins associated with cell structure and basal metabolism have been preferentially used as reference for relative gene expression (Jeong et al., 2006; Bomal et al., 2008; Brunet et al., 2009; Magneschi et al., 2009; Liu et al., 2010). However, the stability of these genes can fluctuate considerably, and for a robust gene expression quantification analysis, the most stable reference genes must be identified and adopted (Vandesompele et al., 2002). For each experiment, a minimum number of reference genes must be used, and the target genes should be normalized against at least the three most stable genes (Vandesompele et al., 2002; Huggett et al., 2005; Bustin et al., 2009).

The expression stability of candidate reference genes was analyzed here for *T. cacao* using two specific softwares, geNorm and NormFinder. While geNorm provides the most stable pair of genes and the ideal number of reference genes for each experiment, NormFinder indicates the most stable gene and the best pair combination, considering the inter- and intra-group differences for stability under specific experimental conditions. The results obtained here from both softwares were similar, but with specific differences. When analyzing samples from developing embryos at various stages, ACT, Rpl35, MDH, PUB, and ACP A were the five most stable reference genes identified by geNorm or NormFinder (Figure 2A; Table 2), suitable for normalization of gene expression for this specific biological condition. However, the ranking of gene expression stability differed depending on the software used. While ACT, Rpl35, MDH were the three first most stable genes based on geNorm, ACT, PUB and Rpl35 were the first recommended by NormFinder.

When the various cacao tissues/organs were analyzed by both approaches, the recom-

mended most stable reference genes were again similar. geNorm indicated GAPDH, ACP A, Rpl35, MDH, and ACP B (Figure 2B) as the best five reference genes for the various tissues, while NormFinder designated MDH, GAPDH ACP B, ACP A, and Rpl35 as the most stable (Table 2). ACP C, ACT and TUB appeared to be the less stable genes for these cacao tissues/organs for both algorithms (Figure 2B; Table 2). Both softwares showed the same ranking for all the candidate reference genes for shoot apices treated with hormones (Figure 2C; Table 2). However, for samples obtained from shoot apices inoculated with *M. pernicioso*, only ACT was ranked among the three best reference genes for both algorithms (Figure 2D; Table 2).

Among all genes tested, ACT and Rpl35 were considered two of the most stable reference genes for investigating gene expression in cacao developing embryos and various tissues by geNorm and NormFinder (Figure 2A,B; Table 2). Reid et al. (2006) found four genes suitable for normalization in development studies of the grape berry, and actin was one of them. Genes encoding ribosomal proteins are extensively used as reference for relative expression studies in various plant systems, such as coffee (Lepelley et al., 2007; Pre et al., 2008). Rpl35 was identified as one of the most stable genes among various tissues and biological conditions in sugarcane based on Serial Analysis of Gene Expression (SAGE) tag counts, and successfully adopted for RT-qPCR (Calsa Jr. and Figueira, 2007). The gene encoding the ribosomal protein L25 was recommended by Schmidt and Delaney (2010) as one of the three most stable genes for normalization of expression in developmentally distinct tobacco tissues, but actin showed the worst stability value in the same experiment. Here, PUB was also considered one of the most stable reference genes to investigate gene expression in developing cacao embryos by geNorm and NormFinder, but it ranked between the less stable genes for the other three experiments. Genes encoding PUB have been recommended as reference for studies of gene expression during seed development in coffee (Salmona et al., 2008), and for various tissues of *Brachypodium* under hormonal treatment (Hong et al., 2008).

ACPs are indispensable co-factors for many metabolic pathways that synthesize fatty acids, phospholipids, signaling molecules, and metabolic co-factors (Byers and Gong, 2007). When genes encoding ACP isoforms were evaluated as reference gene in cacao tissues, ACP A showed to be the most stable pair gene with GAPDH to normalize gene expression in diverse tissues and organs (Figure 2B), whereas ACP B ranked between the best five reference genes for the same samples (Table 2; Figure 2B). When the tissues/organs and developing embryo samples were evaluated together as the “non-treated tissues” subgroup, NormFinder suggested ACP A/ACP B as the best gene pair combination for normalization of gene expression under this condition (Table 2). To our knowledge, this is the first report about the stability of ACP genes among various biological conditions and their suitability as reference gene.

Genes coding for TUBs have normally been used as reference in many plant systems (Jeong et al., 2006). However, these genes were neither considered as good reference for cacao developing embryos nor for the various tissues/organs by both statistical approaches. Similar discrepancy occurred for β -tubulin, considered the least stable reference gene for many tissues in *Lolium perenne* (Martin et al., 2008), and α -tubulin in flax plant tissues (Huis et al., 2010).

The gene encoding GAPDH was designated by both softwares as one of the most stable reference for analysis of gene expression in various cacao tissues. GAPDH was also considered the ideal reference among various tissues in sugarcane (Iskandar et al., 2004) and coffee (Barsalobres-Cavallari et al., 2009). In this study, GAPDH was also pointed out as the best candidate reference gene for shoot apex treated with hormones by both softwares. On

the other hand, EF-1 α , usually recommended as reference gene in studies with plant tissues treated with hormones (Migocka and Papierniak, 2010), was the least stable for our experiments. For shoot apices inoculated with the fungal pathogen *M. perniciosa*, ACT was elected one of the best references among the tested candidates by both softwares (Table 2; Figure 2D) in concert with the EF-1 α /MDH gene pair (Figure 2D). Similar results were obtained for *Vicia faba* inoculated with *Uromyces viciae-fabae* (rust) (Gutierrez et al., 2010). EF-1 α was also one of the most stable genes indicated for normalization during late blight infection in potato experiments, but ACT was the least stable one (Nicot et al., 2005).

The MDH catalyzes the inter-conversion of malate and oxaloacetate, participating in the tricarboxylic acid (TCA) cycle (Park et al., 1995). According to NormFinder, MDH displayed the lowest stability value in the experiments with various tissues/organs and hormonal treatment of cacao (Table 2), and for geNorm, it was considered the most stable pair with GAPDH after hormonal treatment, whereas for the *M. perniciosa* inoculation experiment, it formed with EF-1 α the most stable pair (Figure 2C and D). In addition, when the “treated” subgroup was evaluated by NormFinder, Rpl35/MDH was the best pair combination suggested, indicating that it can be considered a suitable reference gene for *T. cacao* under these biological conditions.

The results obtained from geNorm and NormFinder were not identical as both are based on distinct algorithms. geNorm is more sensitive to co-regulated genes or genes with similar expression patterns, which could indicate not the most stable gene pair, but the pair with similar expression profiles (Andersen et al., 2004).

CONCLUSIONS

Based on these arguments, the results from NormFinder, which were corroborated by geNorm, were selected, not considering the pair of genes. For studies of gene expression in developing embryos, the genes ACT, PUB and Rpl35 were considered the most suitable as references. When all *T. cacao* tissues/organs were investigated, the three most stable genes were MDH, GAPDH and ACP B. For shoot tips treated with hormones, the genes elected were GAPDH, MDH and TUB, while ACT, TUB and Rpl35 were recommended for shoot tips from plants inoculated with *M. perniciosa*.

Gene expression stability should be assessed, especially if comparisons among tissues are needed, which tend to reduce the constancy of expression pattern. The data obtained in this study confirmed that for each plant system and biological or ontogenetical conditions, there is a requirement to define a suitable reference gene. This is the first report to define suitable reference genes for expression studies in cacao, which underlined the importance to establish stable references across various biological conditions.

ACKNOWLEDGMENTS

The authors thank FAPESP for financial support (#2007/07175-0) and fellowship to C.G. Litholdo Jr. and CEPLAC for technical support. T.T. Pinheiro and M.L. Sereno were recipients of CAPES fellowships, and P.S.B. Albuquerque, G.A. Leal Jr., and A. Figueira were recipients of CNPq fellowships. Technical assistance by Edivaldo Pimentel was greatly appreciated.

REFERENCES

- Andersen CL, Jensen JL and 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.
- Argout X, Fouet O, Wincker P, Gramacho K, et al. (2008). Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. *BMC Genom.* 9: 512.
- Argout X, Salse J, Aury JM, Guiltinan MJ, et al. (2011). The genome of *Theobroma cacao*. *Nat. Genet.* 43: 101-108.
- Bae H, Kim SH, Kim MS, Sicher RC, et al. (2008). The drought response of *Theobroma cacao* (cacao) and the regulation of genes involved in polyamine biosynthesis by drought and other stresses. *Plant Physiol. Biochem.* 46: 174-188.
- Barsalobres-Cavallari CF, Severino FE, Maluf MP and Maia IG (2009). Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. *BMC Mol. Biol.* 10: 1.
- Bennett AB (2003). Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends Plant Sci.* 8: 561-563.
- Bomal C, Bedon F, Caron S, Mansfield SD, et al. (2008). Involvement of *Pinus taeda* MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. *J. Exp. Bot.* 59: 3925-3939.
- Brunet J, Varrault G, Zuily-Fodil Y and Repellin A (2009). Accumulation of lead in the roots of grass pea (*Lathyrus sativus* L.) plants triggers systemic variation in gene expression in the shoots. *Chemosphere* 77: 1113-1120.
- Bustin SA, Benes V, Garson JA, Hellems J, et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55: 611-622.
- Byers DM and Gong H (2007). Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. *Biochem. Cell Biol.* 85: 649-662.
- Calsa T Jr and Figueira A (2007). Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C4 metabolism and putative antisense transcripts. *Plant Mol. Biol.* 63: 745-762.
- Carvalho K, de Campos MK, Pereira LF and Vieira LG (2010). Reference gene selection for real-time quantitative polymerase chain reaction normalization in "Swingle" citr melo under drought stress. *Anal. Biochem.* 402: 197-199.
- Donson J, Fang Y, Espiritu-Santo G, Xing W, et al. (2002). Comprehensive gene expression analysis by transcript profiling. *Plant Mol. Biol.* 48: 75-97.
- Gachon C, Mingam A and Charrier B (2004). Real-time PCR: what relevance to plant studies? *J. Exp. Bot.* 55: 1445-1454.
- Guiltinan MJ, Verica J, Zhang D and Figueira A (2008). Genomics of *Theobroma Cacao*, "The Food of the Gods". In: *Genomics of Tropical Crop Plants* (Moore PH and Ming R, eds.). 1st edn. Springer, New York, 145-170.
- Gutierrez N, Giménez MJ, Palomino C and Avila CM (2010). Assessment of candidate reference genes for expression studies in *Vicia faba* L. by real-time quantitative PCR. *Mol. Breed.* DOI: 10.1007/s11032-010-9456-7.
- Hong SY, Seo PJ, Yang MS, Xiang F, et al. (2008). Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR. *BMC Plant Biol.* 8: 112.
- Huggett J, Dheda K, Bustin S and Zumla A (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* 6: 279-284.
- Huis R, Hawkins S and Neutelings G (2010). Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). *BMC Plant Biol.* 10: 71.
- Iskandar HM, Simpson RS, Casu RE and Bonnett GD (2004). Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane. *Plant Mol. Biol. Rep.* 22: 325-337.
- Jeong YM, Mun JH, Lee I, Woo JC, et al. (2006). Distinct roles of the first introns on the expression of *Arabidopsis* profilin gene family members. *Plant Physiol.* 140: 196-209.
- Leal GA Jr, Albuquerque PS and Figueira A (2007). Genes differentially expressed in *Theobroma cacao* associated with resistance to witches' broom disease caused by *Crinipellis pernicioso*. *Mol. Plant Pathol.* 8: 279-292.
- Lepelletier M, Cheminade G, Tremillon N and Simkin A (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.
- Liu JJ, Zamani A and Ekramoddoullah AK (2010). Expression profiling of a complex thaumatin-like protein family in western white pine. *Planta* 231: 637-651.
- Magneschi L, Kudahettige RL, Alpi A and Perata P (2009). Expansin gene expression and anoxic coleoptile elongation in rice cultivars. *J. Plant Physiol.* 166: 1576-1580.
- Martin RC, Hollenbeck VG and Dombrowski JE (2008). Evaluation of reference genes for quantitative RT-PCR in *Lolium perenne*. *Crop Sci.* 48: 1881-1887.
- Migocka M and Papierniak A (2010). Identification of suitable reference genes for studying gene expression in cucumber

- plants subjected to abiotic stress and growth regulators. *Mol. Breed.* DOI: 10.1007/s11032-010-9487-0.
- Mondego JM, Carazzolle MF, Costa GG, Formighieri EF, et al. (2008). A genome survey of *Moniliophthora perniciosa* gives new insights into Witches' Broom Disease of cacao. *BMC Genom.* 9: 548.
- Nicot N, Hausman JF, Hoffmann L and Evers D (2005). Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J. Exp. Bot.* 56: 2907-2914.
- Park SJ, Cotter PA and Gunsalus RP (1995). Regulation of malate dehydrogenase (MDH) gene expression in *Escherichia coli* in response to oxygen, carbon, and heme availability. *J. Bacteriol.* 177: 6652-6656.
- Pires JL, Cascardo JCM, Lambert SV and Figueira A (1998). Increasing cocoa butter yield through genetic improvement of *Theobroma cacao* L.: Seed fat content variability, inheritance, and association with seed yield. *Euphytica* 103: 115-121.
- Pre M, Caillet V, Sobilo J and McCarthy J (2008). Characterization and expression analysis of genes directing galactomannan synthesis in coffee. *Ann. Bot.* 102: 207-220.
- Reid KE, Olsson N, Schlosser J, Peng F, et al. (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.
- Rozen S and Skaletsky H (2000). Primer3 on the WWW for General Users and for Biologists Programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz S and Misener S, eds.). Humana Press, Totowa, 365-386.
- Salmona J, Dussert S, Descroix F, de Kochko A, et al. (2008). Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. *Plant Mol. Biol.* 66: 105-124.
- Schmidt GW and Delaney SK (2010). Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol. Genet. Genom.* 283: 233-241.
- Silva CRS (2005). Caracterização de Sequências Gênicas Expressas (EST) Durante o Desenvolvimento de Sementes de *Theobroma cacao* L. com Ênfase na Biossíntese de Ácidos Graxos e Triglicerídeos. Doctoral thesis, Centro de Energia Nuclear na Agricultura, USP, Piracicaba.
- Silva CRS and Figueira A (2005). Phylogenetic analysis of *Theobroma* (Sterculiaceae) based on Kunitz-like trypsin inhibitor sequences. *Plant Syst. Evol.* 250: 93-104.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3: RESEARCH0034.
- Verica JA, Maximova SN, Strem MD, Carlson JE, et al. (2004). Isolation of ESTs from cacao (*Theobroma cacao* L.) leaves treated with inducers of the defense response. *Plant Cell Rep.* 23: 404-413.
- World Cocoa Foundation (2010). Available at [<http://www.worldcocoafoundation.org>]. Accessed August 13, 2010.