

RESEARCH ARTICLE

Selection of Reference Genes for Expression Study in Pulp and Seeds of *Theobroma grandiflorum* (Willd. ex Spreng.) Schum

Lucas Ferraz dos Santos¹, Raner José Santana Silva¹, Daniel Oliveira Jordão do Amaral¹, Márcia Fabiana Barbosa de Paula^{1*}, Loeni Ludke Falcão², Thierry Legavre³, Rafael Moyses Alves⁴, Lucilia Helena Marcellino², Fabienne Micheli^{1,3}*

1 Universidade Estadual de Santa Cruz (UESC), Departamento de Ciências Biológicas (DCB), Centro de Biotecnologia e Genética (CBG), Rodovia Ilhéus-Itabuna, km 16, 45662-900 Ilhéus-BA, Brazil, **2** Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, 70770-917, Brazil, **3** CIRAD, UMR AGAP, F-34398 Montpellier, France, **4** Embrapa Amazônia Oriental, 66095-903 Belém-PA, Brazil

☯ These authors contributed equally to this work.

✉ Current address: Universidade Federal do Mato Grosso (UFMT), Instituto de Ciências da Saúde (ICS), Av. Alexandre Ferronato, 1200, Sinop, MT, Brazil

* fabienne.micheli@cirad.fr



OPEN ACCESS

Citation: Ferraz dos Santos L, Santana Silva RJ, Oliveira Jordão do Amaral D, Barbosa de Paula MF, Ludke Falcão L, Legavre T, et al. (2016) Selection of Reference Genes for Expression Study in Pulp and Seeds of *Theobroma grandiflorum* (Willd. ex Spreng.) Schum. PLoS ONE 11(8): e0160646. doi:10.1371/journal.pone.0160646

Editor: Hector Candela, Universidad Miguel Hernández de Elche, SPAIN

Received: March 23, 2016

Accepted: July 23, 2016

Published: August 8, 2016

Copyright: © 2016 Ferraz dos Santos et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: LFS was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; Project CAPES-EMBRAPA “Theobroma”); RJSS was supported by the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB, project n°DTE0038/2013); DOJA and MFBP were supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, projects 401924/2012-2 and 501715/2013-4, respectively). LHM was

Abstract

Cupuassu (*Theobroma grandiflorum* [Willd. ex Spreng.] Schum) is a species of high economic importance in Brazil with great potential at international level due to the multiple uses of both its seeds and pulp in the industry of sweets and cosmetics. For this reason, the cupuassu breeding program focused on the selection of genotypes with high pulp and seed quality—selection associated with the understanding of the mechanisms involved in fruit formation. Gene expression is one of the most used approaches related to such understanding. In this sense, quantitative real-time PCR (qPCR) is a powerful tool, since it rapidly and reliably quantifies gene expression levels across different experimental conditions. The analysis by qPCR and the correct interpretation of data depend on signal normalization using reference genes, i.e. genes presenting a uniform pattern of expression in the analyzed samples. Here, we selected and analyzed the expression of five genes from cupuassu (ACP, ACT, GAPDH, MDH, TUB) to be used as candidates for reference genes on pulp and seed of young, maturing and mature cupuassu fruits. The evaluation of the gene expression stability was obtained using the NormFinder, geNorm and BestKeeper programs. In general, our results indicated that the GAPDH and MDH genes constituted the best combination as reference genes to analyze the expression of cupuassu samples. To our knowledge, this is the first report of reference gene definition in cupuassu, and these results will support subsequent analysis related to gene expression studies in cupuassu plants subjected to different biotic or abiotic conditions as well as serve as a tool for diversity analysis based on pulp and seed quality.

the beneficiary of Cirad funds for three months-training at Cirad (Action Incitative). FM was supported by a CNPq Productivity Grant (PQ1). The studies were supported by CAPES-EMBRAPA "Theobroma" and FAPESB (n°DTE0038/2013) projects, both coordinated by FM.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: EST, expressed sequence tags; SSR, simple sequence repeat; qPCR, quantitative PCR; TUB, tubulin; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; ACT, actin; MDH, malate dehydrogenase; ACP, acyl carrier protein; NF, normalization factor; SD, standard deviation; CV, coefficient of variation.

Introduction

Cupuassu (*Theobroma grandiflorum* [Willd. ex Spreng.] Schum) is a fruit species from the Amazon region that belongs to the Malvaceae family, as the cacao tree (*Theobroma cacao* L., used for chocolate fabrication) [1]. Cupuassu is a species of high economic importance in Brazil with great potential at international level due to the multiple uses of both its seeds and pulp. The pulp is used in the industry for candy, ice cream, liquor, and juice production [2, 3], while the seeds are used for the manufacturing of a product similar to chocolate called *cupulate*, as well as in the cosmetic industry [4, 5]. From the beginning of its domestication in the middle '70s, cupuassu plantations suffered production losses due to its susceptibility to the main pathogen of this species, the fungus *Moniliophthora perniciosa*, responsible for witches' broom disease [6]. Because the Amazon region is the potential center of origin of both cupuassu and *M. perniciosa*, it is accepted that the fungus co-evolved with the plant and succeeded in surpassing the plant resistance through an adaptation process. For this reason, the cupuassu breeding program focused on the selection of genotypes with resistance to witches' broom disease allied to high pulp and seed quality, both important characteristics to producers. Agronomic and molecular markers have been previously developed [7–9] and ESTs and SSR-ESTs have been identified [9], constituting the first step for subsequent deep molecular analysis of expressed sequences related to cupuassu resistance or quality. The expression analysis of sequences in which polymorphic markers are identified is an advantage, not only for the breeding program, but also for the understanding of the mechanisms involved in any of the physiological processes studied.

Quantitative PCR (qPCR) is one of the most precise and sensitive techniques to analyze gene expression patterns in organisms subjected to biotic or abiotic stress, as well as throughout the developmental stages of such organisms [10]. The expression analysis of specific genes by qPCR depends on a signal normalization, which is achieved by using genes with a uniform pattern of expression in the considered samples, as referenced [11]. The normalization step is essential for a right and confident interpretation of the data obtained by qPCR [12, 13]. For this reason, it is necessary to previously test the stability of the expression of the candidate genes to be used as reference in the normalization, as carried out for other plant species [14–17] including cacao [18]. Such reference genes could be used in routine laboratory analysis by several research groups worldwide.

Here, we selected five genes from cupuassu EST databank [9] according to their putative functions, known in the literature as constitutive and which present good results as reference genes in other plant species. Such sequences were: i) tubulin (TUB) involved in cellular mechanism [19]; ii) glyceraldehyde-3-phosphate dehydrogenase (GAPDH), which participates in the Calvin cycle and glucose pathways [20, 21]; iii) actin (ACT), which is a cytoskeleton component and cell division regulator [22]; iv) malate dehydrogenase (MDH), an oxidoreductase from Krebs cycle involved in embryo metabolism and development and plant growth [23–25]; and v) acyl carrier protein (ACP) involved in fatty acid biosynthesis [26]. With the advance of the qPCR technique, some software packages supporting reference gene stability analysis also were developed. Among them, NormFinder, geNorm and BestKeeper programs allow the statistical identification of the best internal controls inside a group of candidate genes for normalization [27–29]. The use of such statistical algorithms simplifies the selection of adequate reference genes by calculation of the stability. Here, we analyzed by qPCR the expression pattern of the five candidate genes described above, potentially useful as reference genes for expression studies in cupuassu.

Material and Methods

Primer design for qPCR

Five gene sequences previously obtained from cupuassu [9] were used: tubulin (TUB), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), actin (ACT), malate dehydrogenase (MDH), and acyl carrier protein (ACP). The sequences are available in the [S1 Table](#). Sequences were aligned with the cacao genome using the Blast tool available in the database CocoaGenDB (<http://cocoagendb.cirad.fr/gbrowse/tools.html>). Specific primers were designed using the Primer3Plus program (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>), and the following parameters were analyzed: amplicon size 100–250 pb, GC content of 40–60%, primer size of 20–23 bp and annealing temperature of 60°C. To avoid cross-reaction between genes, the amplified regions were checked for the following characteristics: size difference, melting temperature, GC content, and GC/AT ratio [30] ([S2 Table](#))

Plant material

Samples were collected from the cupuassu genotype 174 (Coari) maintained in the field at Embrapa Amazônia Oriental (PA, Brazil) and selected in the frame of the breeding program of this institution for its characteristics of resistance to witches' broom disease and fruit quality. Seed and pulp were collected from young, maturing, and mature fruits. For each developmental stage and for each tissue, three biological samples were collected (1 biological unit = 1 fruit); a total of nine seed samples (3 units x 3 developmental stages) and nine pulp samples (3 units x 3 developmental stages) were obtained.

RNA extraction and cDNA synthesis

Total RNA was extracted from the samples described above using the CTAB method [31, 32] with modifications [33]. Cupuassu tissues were ground in liquid nitrogen and the extraction carried out with buffer containing 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), and 2% β -mercaptoethanol (v/v). Samples were then treated with chloroform, and the total RNA was precipitated with 2 M of LiCl overnight at 4°C. RNA was resuspended in SSE buffer (1 M NaCl, 0.5% SDS, 10 mM Tris-HCl pH8, 1 mM EDTA pH8), then extracted with 1 volume of phenol, followed by two extractions with chloroform/isoamyl alcohol (24:1), and then precipitated with 2.5 volumes of ethanol 100%. After resuspension in DEPC water, the RNA was quantified using the Nanodrop 2000 spectrophotometer (Thermo Scientific). Afterwards, 1 μ g of RNA was treated with DNase I according to the manufacturer's recommendations (Fermentas Lifes Sciences). The integrity of the RNA was checked on 1% agarose electrophoresis gel. The cDNA synthesis was performed in a final volume of 20 μ l using the cDNA RevertAid First Strand cDNA Synthesis Kit according to the manufacturer's recommendations (Thermo Scientific). The cDNA were quantified by Nanodrop 2000 (Thermo Scientific), and 100 ng/ μ l of each cDNA sample were used for the qPCR analysis.

Quantitative PCR

Expression analysis by qPCR was performed using standard settings of the Stratagene MX3005P system (Agilent Technologies). The qPCR reaction consisted of 100 ng/ μ l of cDNA, 0.3 μ M of each primer from candidate reference genes ([Table 1](#)), and 1X of Maxima™ SYBR Green/ROX qPCR Master Mix (Thermo Scientific) in a total volume of 12.5 μ l. Cycling conditions were: 50°C for 2 min., 95°C for 10 min followed by 40 cycles at 95°C for 15s, 58°C for 35s, and 72°C for 30s, with detection of the fluorescent signal at the end of each extension cycle. To

Table 1. Candidate reference genes, primer characteristics, and efficiency obtained by RT-qPCR. CV: coefficient of variation, F: forward; R: reverse.

Gene	Cell function	Primers (5'-3')	Amplicon size (bp)	Efficiency	R ²	CV
ACP	Fatty acid and polyketide biosynthesis	• F: AGGATGCAGCTGACCTGATT • R: AACTTGGGGTGGATTTCATCA	177	0.86	88	5.5
ACT	Cytoskeleton structure protein	• F: TGAGTTCACTTGACACAGGACA • R: CCTTCCAGCAGATGTGGATT	153	0.88	98	8.54
GAPDH	Glycolysis and gluconeogenesis	• F: TGACTTGATCCGACACATGG • R: TCCCATCCATCCAAAACAT	154	0.88	90	4.68
MDH	Citric acid cycle and gluconeogenesis	• F: ATTCAGAAGGGCGTTTCCTT • R: CCATAGCACAGGGTTTGGT	164	0.86	95	3.43
TUB	Cytoskeleton structure protein	• F: ATGTTGGTGAGGGAATGGAG • R: TTCAGCCACCAAATCTGTGA	215	0.86	91	6.85

doi:10.1371/journal.pone.0160646.t001

verify that each primer pair produced only a single PCR product, a dissociation analysis was carried out under the following cycling conditions: 95°C for 1 min, 55°C for 30 s, and 95°C for 30 s, and it was analyzed with MxPro QPCR software (Agilent Technologies). The gene expression level was analyzed on 3 experimental replicates for all samples. Experiments included a negative control (no template DNA), and amplification efficiency of each primer pair was analyzed using two dilutions [1:50 and 1:100 (v/v)] of each cDNA sample. Real-time data acquisition was performed by the Stratagene MX3005P system containing the MxPro QPCR software (Agilent Technologies), which provided the values of cycle threshold (Ct) and of fluorescence. Amplification efficiency (E) was accessed using Miner 2.2 software [34]

Reference gene stability analyses

To establish the ideal reference gene among the diverse biological conditions, gene expression data were analyzed using the statistical approach used by the geNorm, NormFinder and BestKeeper software. The geNorm VBA applet for Microsoft Excel determines the most stable reference genes from a set of tested genes in a given cDNA sample panel, and it calculates a gene expression normalization factor (NF) for each tissue sample based on the geometric mean of a user-defined number of reference genes [28]. The geNorm program calculates the gene expression stability measure *M* for a reference gene as the average pairwise variation *V* for that gene with all other tested reference genes. Stepwise exclusion of the gene with the highest *M* value allows for ranking of the tested genes according to their expression stability [28]. The geNorm program also determines the optimal number of reference genes needed for qPCR experiment using pairwise variation (V_n/V_{n+1}) analysis between the normalization factors (NF_n and NF_{n+1}). The NormFinder program uses Ct values transformed into relative quantities (Q), estimated by $Q = E^{(Ct_{\text{minimum}} - Ct_{\text{sample}})}$, where *E* is the amplification efficiency and the Ct minimum is the lowest Ct value (sample with the highest expression among all samples) [28]. The NormFinder program identifies the best reference gene by ranking all candidates to the 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. The BestKeeper program determines the best reference genes and combines them into an index, which can be used as standard in the same way such as any single reference gene [29]. The BestKeeper Excel template offers three main measures to detect the most stable genes: the standard deviation (SD) of the Ct of all samples for one gene, the correlation coefficient with the BestKeeper index (*r*), and the coefficient of variation (CV) of a potential reference gene [29, 35]. The BestKeeper guideline

does not indicate which of these three measures is more reliable nor has more weight to select the best reference gene. According to previous studies, only SD, only r , or a combination of SD + CV have been already used [36–38]. On the other hand, each potential reference gene was evaluated for the three measures offered by the BestKeeper (SD, r , and CV), and the mean of the rankings was calculated to determine the final rank of each gene, as described by Olias and collaborators [35].

Results and Discussion

Candidate reference gene selection, primer design and efficiency analysis

Based on previous studies, the ACP, ACT, GAPDH, MDH, and TUB genes were selected as possible candidate reference genes in cupuassu expression analysis. All these genes presented constitutive expression patterns in other plant species and some of them, such as ACP, presented a special interest for the *Theobroma* genus. The ACP gene is an important co-factor of several metabolic pathways [39], including those related to lipid storage occurring in seeds, and for this reason it was previously chosen in cacao [18] and also may be a good candidate for cupuassu. The MDH gene is also an interesting candidate due to its involvement in plant growth and embryo development [25], one subject addressed here (fruit under maturation process). Specific primers for each candidate's reference genes (Table 1) were tested in all samples (i.e. pulp and seed, each one in the 3 maturation fruit stage—see [Material and methods](#)) to confirm the fragment size and the amplification of a single PCR product. Dissociation curves showed that after 35 cycles of amplification, all the primers used for the analysis presented a unique peak (i.e., a unique PCR product) proving the primer specificity (S1 Fig). Both cDNA dilution 1:50 and 1:100 were tested, and no significant difference was observed (data not shown); thus, the 1:100 dilution was used.

Expression profile of candidate reference genes

The raw gene expression levels were estimated by Ct for all the samples—all together or according to the tissue (pulp vs seeds) or to the developmental stage (young, maturing or mature fruit) (Fig 1). Mean Ct values presented variations for each gene between samples: Ct varied from 22.2 to 24.0 for ACP, from 18.7 to 21.3 for ACT, from 19.8 to 21.3 for GAPDH, from 23.1 to 24.9 for MDH, 20.7 to 23.5 for TUB (S3 Table; Fig 1A to 1F). Generally, the ACT and GAPDH genes presented the lowest Ct using the different cupuassu samples (Fig 1A, 1B, 1C, 1D and 1F), while ACP presented the highest (Fig 1A, 1C, 1D and 1E). The global Ct variation was low—from 18.7 (for ACT) to 24.9 (for MDH)—and was consistent (similar or lower) with the one observed in other studies of reference genes previously published in other organisms [12, 16, 35, 40]. Such low Ct variation also indicated that the candidate genes presented similar behavior regardless of the biological condition analyzed (Fig 1; S3 Table) and for this reason may be used in combination (2 or 3 reference genes) for RT-qPCR normalization.

Expression stability analysis

The expression stability of the five reference candidate genes was analyzed using the geNorm, NormFinder, and BestKeeper programs. The stability was evaluated in cupuassu pulp and seeds during the fruit development stages: i) all stages and tissues put together; ii) according to the tissue; and iii) according to the fruit developmental stage. In our conditions, the geNorm program estimated that the optimal number of reference genes was at least 3 according to the samples analyzed (cut-off of 0.15; S4 Table). However, as the guideline, the authors of the

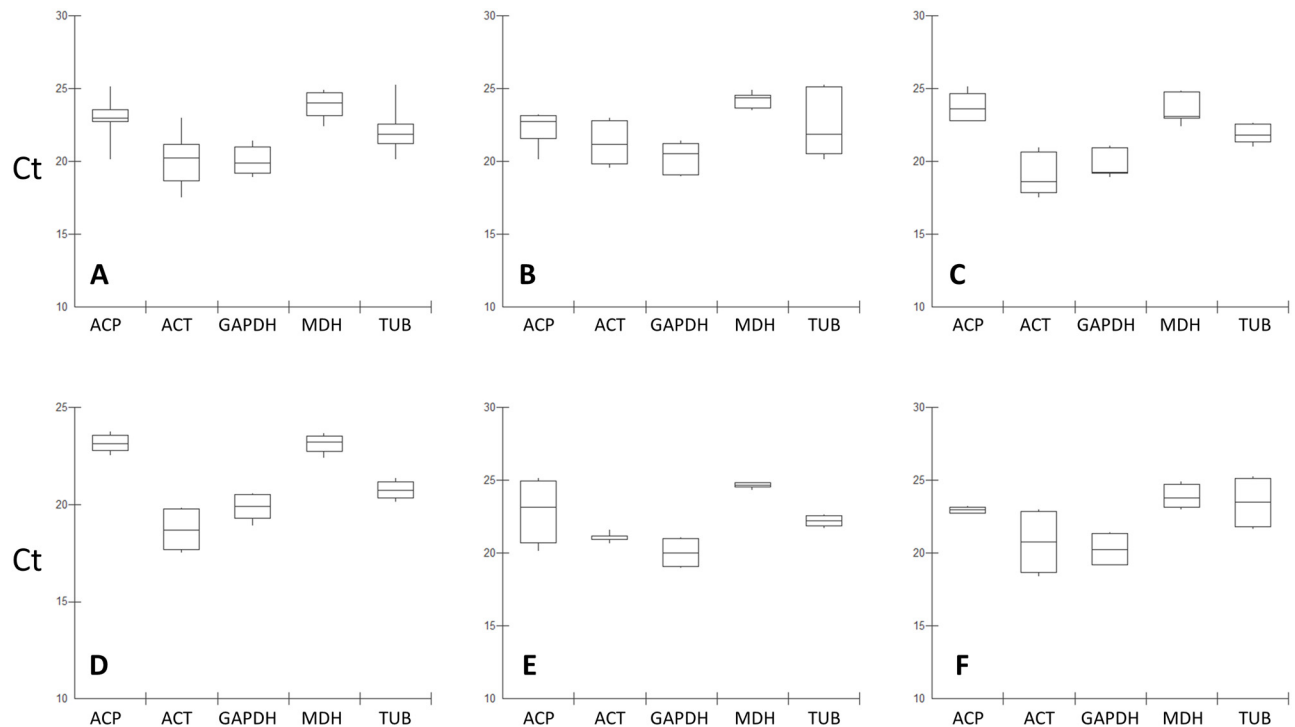


Fig 1. Box-plots showing Ct variation for each candidate reference gene estimated by RT-qPCR for the different samples. A. All samples (all tissues, all developmental stages). **B.** Pulp samples (all developmental stages). **C.** Seed samples (all developmental stages). **D.** Young fruit samples (all tissues). **E.** Maturing samples (all tissues). **F.** Mature samples (all stages). The line inside the box represents the median; the two parts of the box on each side of the line indicated the 25 and 75 percentiles. Whiskers represent the maximum and minimum values.

doi:10.1371/journal.pone.0160646.g001

geNorm program indicated that the proposed 0.15 value must not be taken as a too strict cut-off for determination of the optimal number of reference genes. It can be that a valid normalization strategy is obtained just using the two best reference genes, resulting in a much more accurate and reliable normalization than if one single reference gene is used [41]. The geNorm program organized the candidate genes according to their stability (Fig 2). In the all stages/tissues sample and in the pulp sample, the two more stable genes were GAPDH and MDH (Fig 2A and 2B). In seeds, ACP and GAPDH were the most stable genes (Fig 2C). In fruit samples, ACP and TUB (young fruit), TUB and MDH (maturing fruit), and TUB and GAPDH (mature fruits) were the most stable, respectively (Fig 2D to 2F).

The NormFinder program uses an algorithm for identification and optimization of reference genes among a candidate group. It ranked the candidate reference genes according to their stability of expression for a sample group (e.g., pulp) and according to the experimental design (e.g., throughout the fruit development). The algorithm is based on a mathematic model of gene expression and used a strong statistic network to estimate not only the global variation of expression of the candidate reference genes, but also the variation between the sample sub-groups from the initial sample set [27]. In the all stages/tissues samples, GAPDH and MDH were the most stable genes (0.0189 and 0.0182, respectively), as well as in the young fruit sample (0.011 for both; Table 2). In the pulp, MDH and TUB presented the best expression stability (0.0047 and 0.0091, respectively; Table 2), as well as for the maturing fruit sample (0.007 for both; Table 2). In seeds, ACT and TUB were the most stable genes (0.0197 for both), while in the mature fruit, GAPDH and TUB presented the best stability (0.013 and 0.014, respectively; Table 2). As a general analysis made by the NormFinder program, the best

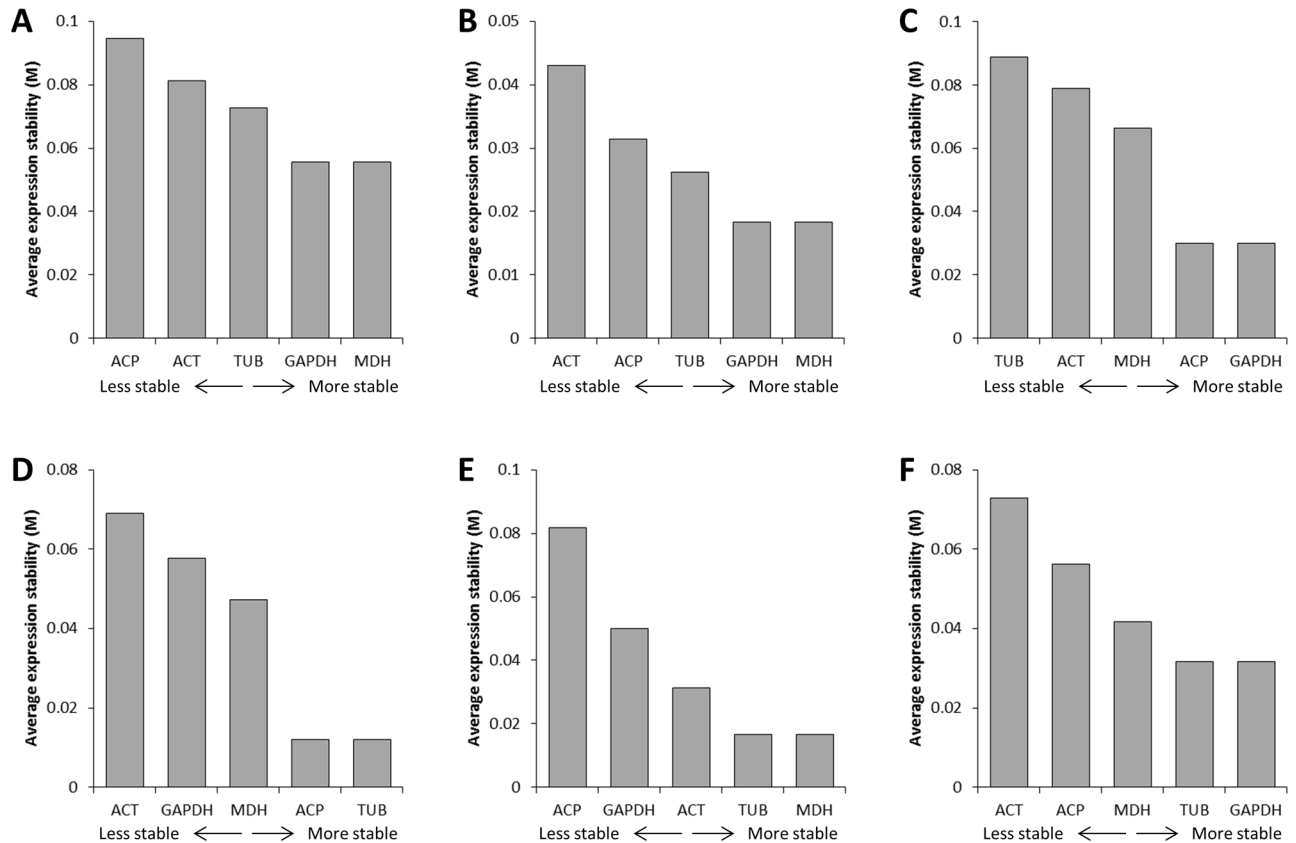


Fig 2. Gene expression stability obtained by geNorm. **A.** All samples (all tissues, all developmental stages). **B.** Pulp samples (all developmental stages). **C.** Seed samples (all developmental stages). **D.** Young fruit samples (all tissues). **E.** Maturing samples (all tissues). **F.** Mature samples (all stages). Mean expression stability (*M*) was calculated following stepwise exclusion of the least stable gene across all the treatment groups.

doi:10.1371/journal.pone.0160646.g002

combination of two genes for normalization in cupuassu samples was GAPDH and MDH (0.014; Table 2). Generally, the result we obtained here presented very high quality and stability compared with other works; our best combination presented a stability of 0.014, much lower than the one selected in other studies (0.06) [42]—value that corresponds with our worse stability (ACP; Table 2).

The BestKeeper program allowed the determination of three main measures (SD, *r*, and CV; see Material and methods) which when combined resulted in a gene ranking (Table 3). The

Table 2. Expression stability of the candidate reference genes obtained with NormFinder program. Data in bold indicates the best two genes for the sample group analyzed.

	ACP	ACT	GAPDH	MDH	TUB	Best combination of two genes*
All stages/all tissues	0.061	0.0595	0.0189	0.0182	0.0342	GAPDH and MDH (0.014)
Pulp (all stages)	0.0862	0.0185	0.0347	0.0047	0.0091	
Seeds (all stages)	0.1190	0.0197	0.1013	0.0772	0.0197	
Young fruit (all tissues)	0.118	0.015	0.011	0.011	0.122	
Maturing fruit (all tissues)	0.19	0.011	0.08	0.007	0.007	
Mature fruit (all tissues)	0.14	0.022	0.013	0.138	0.014	

* Given by the NormFinder program for the overall analysis. The best combination value is indicated inside brackets.

doi:10.1371/journal.pone.0160646.t002

Table 3. Best reference genes obtained by the BestKeeper program. SD: standard deviation; *r*: coefficient of correlation; CV: coefficient of variation. Best reference genes are indicated in bold.

		ACP	ACT	GAPDH	MDH	TUB
All stages/all tissues	SD*	0.92 (3)	1.49 (5)	0.91 (2)	0.77 (1)	1.16 (4)
	CV*	4.02 (2)	7.40 (5)	4.53 (3)	3.24 (1)	5.22 (4)
	<i>r</i> *	0.383 (5)	0.854 (3)	0.871 (1)	0.831 (4)	0.855 (2)
	Mean of rankings	3.33	4.33	2	2	3.33
Pulp (all stages)	SD*	0.79 (2)	1.21 (5)	0.82 (3)	0.85 (4)	0.49 (1)
	CV*	3.33 (2)	6.36 (5)	4.12 (4)	3.63 (3)	2.25 (1)
	<i>r</i> *	0.871 (5)	0.982 (3)	0.988 (1)	0.987 (2)	0.944 (4)
	Mean of rankings	3	4.33	2.67	3	2
Seeds (all stages)	SD*	0.98 (3)	1.11 (4)	0.83 (2)	0.45 (1)	1.82 (5)
	CV*	4.44 (3)	5.23 (4)	4.11 (2)	1.87 (1)	8.12 (5)
	<i>r</i> *	0.609 (5)	0.877 (2)	0.751 (3)	0.659 (4)	0.959 (1)
	Mean of rankings	3.66	3.33	2.33	2	3.66
Young fruit (all tissues)	SD*	0.45 (3)	1.04 (5)	0.68 (4)	0.44 (2)	0.44 (2)
	CV*	1.94 (2)	5.55 (5)	3.43 (4)	1.89 (1)	2.15 (3)
	<i>r</i> *	-0.941 (5)	0.969 (2)	0.991 (1)	0.936 (3)	-0.910 (4)
	Mean of rankings	3.33	4	3**	2	3**
Maturing fruit (all tissues)	SD*	2.13 (5)	0.22 (2)	0.98 (4)	0.17 (1)	0.38 (3)
	CV*	9.31 (5)	1.04 (2)	4.91 (4)	0.70 (1)	1.73 (3)
	<i>r</i> *	0.995 (1)	-0.738 (5)	0.992 (2)	0.915 (4)	0.984 (3)
	Mean of rankings	3.66	3	3.33	2	3
Mature fruit (all tissues)	SD*	0.19 (1)	2.13 (5)	1.06 (3)	0.81 (2)	1.69 (4)
	CV*	0.84 (1)	10.30 (5)	5.23 (3)	3.38 (2)	7.20 (4)
	<i>r</i> *	0.986 (4)	1.000 (1)	0.998 (2)	0.987 (3)	1.000 (1)
	Mean of rankings	2	3.66	2.66**	2.66**	3

* Rank is indicated inside brackets.
 ** Same ranking for a given condition.

doi:10.1371/journal.pone.0160646.t003

ranking indicated that GAPDH and MDH were the best genes in the all stages/all tissues sample, GAPDH and TUB in the pulp sample, GAPDH and MDH in seeds, MDH and GAPDH/TUB in the case of young fruit, MDH and TUB for maturing fruit, and ACP and GAPDH/MDH for mature fruits (Table 3). Even if the calculation and statistics used in the geNorm, NormFinder and BestKeeper programs were different, the results obtained for the normalization of the expression in cupuassu tissues were slightly the same. As presented in Table 4, the GAPDH and MDH genes were present both or individually in all the combinations of two or

Table 4. Main reference genes indicated for each physiological condition using the geNorm, NormFinder and BestKeeper programs. Genes are indicated in increasing order of stability.

	geNorm	NormFinder	BestKeeper
All stages/all tissues	MDH/GAPDH/TUB	MDH/GAPDH/TUB	GAPDH/MDH
Pulp (all stages)	MDH/GAPDH/TUB	MDH/TUB/ACT	GAPDH/TUB
Seeds (all stages)	GAPDH/ACP/MDH	ACT/TUB/MDH	GAPDH/MDH
Pulp and seeds (young fruit)	ACP/TUB/MDH	MDH/GAPDH/ACT	MDH/ GAPDH or TUB
Maturing pulp and seeds	MDH/TUB/ACT	MDH/TUB/ACT	MDH/TUB
Mature pulp and seeds	GAPDH/TUB/MDH	GAPDH/TUB/ACT	ACP/GAPDH or MDH

doi:10.1371/journal.pone.0160646.t004

three best reference genes. Generally, the third-best gene was TUB, followed by ACT or ACP (Table 4).

A reference gene needs to present constant expression level, but the stability of this gene may suffer variations according to the tissue or biological condition analyzed such as developmental stage or response to stress [43, 44]. All genes evaluated presented stability values adequate for the normalization of the target gene in expression studies using RT-qPCR, with a highlight for the GAPDH and MDH (Table 4). According to three different gene-expression stability analysis programs, these two genes presented high stability in all the tested conditions and also presented compatibility to be used together in RT-qPCR expression analyses (Table 2). These results are consistent with studies carried out for other plant species such as in coffee seeds, *Vitis vinifera* fruit pericarp, developing organs/tissues, or plants under stress, where GAPDH has been identified as one of the most stable among the tested genes [18, 44–46]. MDH was also considered as a good reference gene in similar analysis in Pearl millet [47] and in *Theobroma cacao* [18].

Conclusion

In general, our results indicate that the GAPDH and MDH genes constituted the best combination as reference genes to analyze gene expression in cupuassu samples. In the third position, the TUB gene may be envisaged. To our knowledge, this is the first report of reference gene identification in cupuassu. This analysis will support subsequent expression studies in cupuassu plants subjected to different biotic or abiotic conditions as well as serve as a tool for diversity analysis based on pulp and seed quality.

Supporting Information

S1 Fig. Dissociation curves obtained for the five genes used in this study. Temperature is presented in the x-axis while the fluorescence is presented in the y-axis.
(DOCX)

S1 Table. Five sequences of cupuassu used in this study.
(DOCX)

S2 Table. Amplicon characteristics.
(DOCX)

S3 Table. Ct mean for each reference candidate gene in the different tissues and stages analyzed. Maximum and minimum are indicated in bold. (*) indicates two maximum values for the considered gene.
(DOCX)

S4 Table. Estimation of the optimal number of reference genes required for accurate normalization based on pairwise variation ($V_n/n+1$) analysis. The asterisk indicates the lowest V value in each condition. Bold indicate the variation ≤ 0.15 (suggested cut-off).
(DOCX)

Acknowledgments

We thank Dr. Claudia Fortes Ferreira (Embrapa CNPMPF, Brazil) for the English-language revision.

Author Contributions

Conceptualization: TL LHM FM.

Formal analysis: LFS RJSS FM.

Funding acquisition: FM.

Investigation: LFS RJSS DOJA MFBP LLF.

Methodology: LFS RJSS TL LHM FM.

Project administration: FM.

Resources: RMA FM.

Software: RJSS.

Supervision: LHM FM.

Visualization: LFS FM.

Writing - original draft: LFS FM.

Writing - review & editing: LFS FM.

References

1. Clement C, De Cristo-Araújo M, Coppens D'Eeckenbrugge G, Alves Pereira A, Picanço-Rodrigues D. Origin and domestication of native amazonian crops. *Diversity*. 2010; 2(1):72–106. doi: [10.3390/d2010072](https://doi.org/10.3390/d2010072)
2. Calzavara BBG, Muller CH, Kahwage ONN. *Fruticultura Tropical: o cupuaçuzeiro—cultivo, beneficiamento e utilização do fruto*. Belém: EMBRAPA-CPATU; 1984. 101 p.
3. Cohen KO, Jackix MNH. Estudo do liquor de cupuaçu. *Food Science and Technology (Campinas)*. 2005; 25:182–90.
4. Carvalho AV, García NHP, Farfán JA. Proteínas da semente de cupuaçu e alterações devidas à fermentação e à torração. *Food Science and Technology (Campinas)*. 2008; 28:986–93.
5. Alves RM, Silva CRS, Silva MSC, Silva DCS, Sebbenn AM. Diversidade genética em coleções amazônicas de germoplasma de cupuaçuzeiro [*Theobroma grandiflorum* (Willd. ex Spreng.) Schum.]. *Revista Brasileira de Fruticultura*. 2013; 35:818–28.
6. Alves RM, Resende MDV. Avaliação genética de indivíduos e progênes de cupuaçuzeiro no estado do Para e estimativas de parâmetros genéticos. *Revista Brasileira de Fruticultura*. 2008; 30(3):696–701.
7. Alves RM, Sebbenn AM, Artero AS, Figueira A. Microsatellite loci transferability from *Theobroma cacao* to *Theobroma grandiflorum*. *Molecular Ecology Notes*. 2006; 6(4):1219–21. doi: [10.1111/j.1471-8286.2006.01496.x](https://doi.org/10.1111/j.1471-8286.2006.01496.x)
8. Maia MCC, Resende MDV, Oliveira LC, Alves RM, Silva JL, Rocha MM, et al. Análise genética de famílias de meios-irmãos de cupuaçuzeiro. *Pesquisa Florestal Brasileira*. 2011; 31:123–30. <http://dx.doi.org/10.4336/2011.pfb.31.66.123>.
9. Ferraz dos Santos L, Moreira Fregapani R, Falcão LL, Togawa RC, Costa MMdC, Lopes UV, et al. First Microsatellite Markers Developed from Cupuassu ESTs: Application in Diversity Analysis and Cross-Species Transferability to Cacao. *PLoS ONE*. 2016; 11(3):e0151074. doi: [10.1371/journal.pone.0151074](https://doi.org/10.1371/journal.pone.0151074) PMID: [26949967](https://pubmed.ncbi.nlm.nih.gov/26949967/)
10. Gachon C, Mingam A, Charrier B. Real-time PCR: what relevance to plant studies? *J Exp Bot*. 2004; 55. doi: [10.1093/jxb/erh181](https://doi.org/10.1093/jxb/erh181)
11. Bastolla F. Seleção e avaliação de genes de referência para estudos de expressão gênica em *Eucalyptus*: Universidade Federal do Rio Grande do Sul; 2007.
12. Demidenko NV, Logacheva MD, Penin AA. Selection and Validation of Reference Genes for Quantitative Real-Time PCR in Buckwheat (*Fagopyrum esculentum*) Based on Transcriptome Sequence Data. *PLoS ONE*. 2011; 6(5):e19434. doi: [10.1371/journal.pone.0019434](https://doi.org/10.1371/journal.pone.0019434). PMC3093374. PMID: [21589908](https://pubmed.ncbi.nlm.nih.gov/21589908/)

13. Expósito-Rodríguez M, Borges AA, Borges-Pérez A, Pérez JA. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biology*. 2008; 8(1):1–12. doi: [10.1186/1471-2229-8-131](https://doi.org/10.1186/1471-2229-8-131)
14. Jain M, Nijhawan A, Tyagi AK, Khurana JP. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications*. 2006; 345(2):646–51. <http://dx.doi.org/10.1016/j.bbrc.2006.04.140>. PMID: [16690022](https://pubmed.ncbi.nlm.nih.gov/16690022/)
15. Mariot RF, de Oliveira LA, Voorhijzen MM, Staats M, Hutten RCB, Van Dijk JP, et al. Selection of Reference Genes for Transcriptional Analysis of Edible Tubers of Potato (*Solanum tuberosum* L.). *PLoS ONE*. 2015; 10(4):e0120854. doi: [10.1371/journal.pone.0120854](https://doi.org/10.1371/journal.pone.0120854) PMID: [25830330](https://pubmed.ncbi.nlm.nih.gov/25830330/)
16. Amil-Ruiz F, Garrido-Gala J, Blanco-Portales R, Folta KM, Muñoz-Blanco J, Caballero JL. Identification and Validation of Reference Genes for Transcript Normalization in Strawberry (*Fragaria × ananassa*) Defense Responses. *PLoS ONE*. 2013; 8(8):e70603. doi: [10.1371/journal.pone.0070603](https://doi.org/10.1371/journal.pone.0070603) PMID: [23940602](https://pubmed.ncbi.nlm.nih.gov/23940602/)
17. Løvdal T, Lillo C. Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. *Analytical Biochemistry*. 2009; 387(2):238–42. <http://dx.doi.org/10.1016/j.ab.2009.01.024>. doi: [10.1016/j.ab.2009.01.024](https://doi.org/10.1016/j.ab.2009.01.024) PMID: [19454243](https://pubmed.ncbi.nlm.nih.gov/19454243/)
18. Pinheiro TT, Litholdo CG Jr., Sereno ML, Leal GA Jr., Albuquerque PSB, Figueira A. Establishing references for gene expression analyses by RT-qPCR in *Theobroma cacao* tissues. *Genetics and Molecular Research*. 2012; 10(4):3291–305. doi: [10.4238/2011.November.17.4](https://doi.org/10.4238/2011.November.17.4)
19. Das L, Bhattacharya B, Basu G. Rationalization of paclitaxel insensitivity of yeast β -tubulin and human β III-tubulin isotype using principal component analysis. *BMC Research Notes*. 2012; 5(1):1–10. doi: [10.1186/1756-0500-5-395](https://doi.org/10.1186/1756-0500-5-395)
20. Petersen J, Brinkmann H, Cerff R. Origin, Evolution, and Metabolic Role of a Novel Glycolytic GAPDH Enzyme Recruited by Land Plant Plastids. *Journal of Molecular Evolution*. 2003; 57(1):16–26. doi: [10.1007/s00239-002-2441-y](https://doi.org/10.1007/s00239-002-2441-y) PMID: [12962302](https://pubmed.ncbi.nlm.nih.gov/12962302/)
21. Zaffagnini M, Fermani S, Costa A, Lemaire SD, Trost P. Plant cytoplasmic GAPDH: redox post-translational modifications and moonlighting properties. *Frontiers in Plant Science*. 2013; 4. doi: [10.3389/fpls.2013.00450](https://doi.org/10.3389/fpls.2013.00450)
22. Paavilainen VO, Bertling E, Falck S, Lappalainen P. Regulation of cytoskeletal dynamics by actin-monomer-binding proteins. *Trends in Cell Biology*. 2004; 14(7):386–94. doi: [10.1016/j.tcb.2004.05.002](https://doi.org/10.1016/j.tcb.2004.05.002) PMID: [15246432](https://pubmed.ncbi.nlm.nih.gov/15246432/)
23. Centeno DC, Osorio S, Nunes-Nesi A, Bertolo ALF, Carneiro RT, Araújo WL, et al. Malate Plays a Crucial Role in Starch Metabolism, Ripening, and Soluble Solid Content of Tomato Fruit and Affects Post-harvest Softening. *The Plant Cell*. 2011; 23(1):162–84. doi: [10.1105/tpc.109.072231](https://doi.org/10.1105/tpc.109.072231) PMID: [21239646](https://pubmed.ncbi.nlm.nih.gov/21239646/)
24. Beeler S, Liu H-C, Stadler M, Schreier T, Eicke S, Lue W-L, et al. Plastidial NAD-Dependent Malate Dehydrogenase Is Critical for Embryo Development and Heterotrophic Metabolism in Arabidopsis. *Plant Physiology*. 2014; 164(3):1175–90. doi: [10.1104/pp.113.233866](https://doi.org/10.1104/pp.113.233866). PMC3938612. PMID: [24453164](https://pubmed.ncbi.nlm.nih.gov/24453164/)
25. Wang Z-A, Li Q, Ge X-Y, Yang C-L, Luo X-L, Zhang A-H, et al. The mitochondrial malate dehydrogenase 1 gene GhmMDH1 is involved in plant and root growth under phosphorus deficiency conditions in cotton. *Scientific Reports*. 2015; 5:10343. doi: [10.1038/srep10343](https://doi.org/10.1038/srep10343). PMC4503954. PMID: [26179843](https://pubmed.ncbi.nlm.nih.gov/26179843/)
26. Jing F, Cantu DC, Tvaruzkova J, Chipman JP, Nikolau BJ, Yandea-Nelson MD, et al. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. *BMC Biochemistry*. 2011; 12(1):1–16. doi: [10.1186/1471-2091-12-44](https://doi.org/10.1186/1471-2091-12-44)
27. Andersen CL, Jensen JL, Ørntoft TF. 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 Research*. 2004; 64(15):5245–50. doi: [10.1158/0008-5472.can-04-0496](https://doi.org/10.1158/0008-5472.can-04-0496) PMID: [15289330](https://pubmed.ncbi.nlm.nih.gov/15289330/)
28. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002; 3. doi: [10.1186/gb-2002-3-7-research0034](https://doi.org/10.1186/gb-2002-3-7-research0034)
29. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. *Biotechnol Lett*. 2004; 26. doi: [10.1023/b:BILE.0000019559.84305.47](https://doi.org/10.1023/b:BILE.0000019559.84305.47)
30. Ririe KM, Rasmussen RP, Wittwer CT. Product Differentiation by Analysis of DNA Melting Curves during the Polymerase Chain Reaction. *Analytical Biochemistry*. 1997; 245(2):154–60. <http://dx.doi.org/10.1006/abio.1996.9916>. PMID: [9056205](https://pubmed.ncbi.nlm.nih.gov/9056205/)
31. Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter / Isymb*. 1993; 11(2):113–6.

32. Logemann J, Schell J, Willmitzer L. Improved method for the isolation of RNA from plant tissues. *Analytical Biochemistry*. 1987; 163(1):16–20. [http://dx.doi.org/10.1016/0003-2697\(87\)90086-8](http://dx.doi.org/10.1016/0003-2697(87)90086-8). PMID: [2441623](https://pubmed.ncbi.nlm.nih.gov/2441623/)
33. Santana RJS, Silva-Werneck JO, Falcao LL, Marcellino LH. Extração de RNA de cupuaçuzeiro (*Theobroma grandiflorum*) para estudos de expressão gênica. Congresso Nacional de Ciências Biológicas; Recife, Brazil 2011. p. 387–91.
34. Zhao S, Fernald RD. Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. *Journal of computational biology: a journal of computational molecular cell biology*. 2005; 12(8):1047–64. doi: [10.1089/cmb.2005.12.1047](https://doi.org/10.1089/cmb.2005.12.1047). PMC2716216.
35. Olias P, Adam I, Meyer A, Scharff C, Gruber AD. Reference Genes for Quantitative Gene Expression Studies in Multiple Avian Species. *PLoS ONE*. 2014; 9(6):e99678. doi: [10.1371/journal.pone.0099678](https://doi.org/10.1371/journal.pone.0099678). PMC4057121. PMID: [24926893](https://pubmed.ncbi.nlm.nih.gov/24926893/)
36. Hibbeler S, Scharsack JP, Becker S. Housekeeping genes for quantitative expression studies in the three-spined stickleback *Gasterosteus aculeatus*. *BMC Molecular Biology*. 2008; 9(1):1–10. doi: [10.1186/1471-2199-9-18](https://doi.org/10.1186/1471-2199-9-18)
37. Maroufi A, Van Bockstaele E, De Loose M. Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. *BMC Molecular Biology*. 2010; 11(1):1–12. doi: [10.1186/1471-2199-11-15](https://doi.org/10.1186/1471-2199-11-15)
38. Chang E, Shi S, Liu J, Cheng T, Xue L, Yang X, et al. Selection of Reference Genes for Quantitative Gene Expression Studies in *Platyclusus orientalis* (Cupressaceae) Using Real-Time PCR. *PLoS ONE*. 2012; 7(3):e33278. doi: [10.1371/journal.pone.0033278](https://doi.org/10.1371/journal.pone.0033278). PMC3316566. PMID: [22479379](https://pubmed.ncbi.nlm.nih.gov/22479379/)
39. Byers DMBM, Gong H. Acyl carrier protein: structure–function relationships in a conserved multifunctional protein family. *Biochemistry and Cell Biology*. 2007; 85(6):649–62. doi: [10.1139/O07-109](https://doi.org/10.1139/O07-109) PMID: [18059524](https://pubmed.ncbi.nlm.nih.gov/18059524/).
40. Cheng D, Zhang Z, He X, Liang G. Validation of Reference Genes in *Solenopsis invicta* in Different Developmental Stages, Castes and Tissues. *PLoS ONE*. 2013; 8(2):e57718. doi: [10.1371/journal.pone.0057718](https://doi.org/10.1371/journal.pone.0057718). PMC3585193. PMID: [23469057](https://pubmed.ncbi.nlm.nih.gov/23469057/)
41. geNorm software manual update. Available: <http://medgen.ugent.be/~jvdesomp/genorm>.
42. Wan H, Zhao Z, Qian C, Sui Y, Malik AA, Chen J. Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber. *Analytical Biochemistry*. 2010; 399(2):257–61. <http://dx.doi.org/10.1016/j.ab.2009.12.008>. doi: [10.1016/j.ab.2009.12.008](https://doi.org/10.1016/j.ab.2009.12.008) PMID: [20005862](https://pubmed.ncbi.nlm.nih.gov/20005862/)
43. Suzuki T, Higgins PJ, Crawford DR. Control selection for RNA quantitation. *Biotechniques*. 2000; 29.
44. Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG. Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. *BMC Molecular Biology*. 2009; 10(1):1–11. doi: [10.1186/1471-2199-10-1](https://doi.org/10.1186/1471-2199-10-1)
45. Reid K, Olsson N, Schlosser J, Peng F, Lund S. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biol*. 2006; 6:27. doi: [10.1186/1471-2229-6-27](https://doi.org/10.1186/1471-2229-6-27) PMID: [17105665](https://pubmed.ncbi.nlm.nih.gov/17105665/)
46. Carvalho K, de Campos MKF, Pereira LFP, Vieira LG. Reference gene selection for real-time quantitative polymerase chain reaction normalization in “Swingle” citrumelo under drought stress. *Analytical Biochemistry*. 2010; 402(2):197–9. doi: [10.1016/j.ab.2010.03.038](https://doi.org/10.1016/j.ab.2010.03.038) PMID: [20363209](https://pubmed.ncbi.nlm.nih.gov/20363209/)
47. Reddy PS, Reddy DS, Sharma KK, Bhatnagar-Mathur P, Vadez V. Cloning and validation of reference genes for normalization of gene expression studies in pearl millet [*Pennisetum glaucum* (L.) R. Br.] by quantitative real-time PCR. *Plant Gene*. 2015; 1:35–42. <http://dx.doi.org/10.1016/j.plgene.2015.02.001>.