Selection and validation of appropriate reference genes for quantitative real-time PCR analysis in *Salvia hispanica*

Rahul Gopalam[1,2], Sunny D. Rupwate[1¤], Ajay W. Tumaney[1,2]*

1 Department of Lipid Science, CSIR-Central Food Technological Research Institute (CSIR-CFTRI), Mysore, Karnataka, India, **2** Academy of Scientific and Innovative Research (AcSIR), CSIR-Central Food Technological Research Institute (CSIR-CFTRI), Mysore, Karnataka, India

¤ Current address: Dr. John Barnabas Post Graduate School for Biological Studies, Department of Chemistry, Ahmednagar College, Ahmednagar, India

* ajay@cftri.res.in

Abstract

Quantitative real-time polymerase chain reaction (qRT-PCR) has become the most popular choice for gene expression studies. For accurate expression analysis, it is pertinent to select a stable reference gene to normalize the data. It is now known that the expression of internal reference genes varies considerably during developmental stages and under different experimental conditions. For *Salvia hispanica*, an economically important oilseed crop, there are no reports of stable reference genes till date. In this study, we chose 13 candidate reference genes viz. Actin11 (ACT), Elongation factor 1-alpha (EF1-α), Eukaryotic translation initiation factor 3E (ETIF3E), alpha tubulin (α-TUB), beta tubulin (β-TUB), Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), Cyclophilin (CYP), Clathrin adaptor complex (CAC), Serine/threonine-protein phosphatase 2A (PP2A), FtsH protease (FtsH), 18S ribosomal RNA (18S rRNA), S-adenosyl methionine decarboxylase (SAMDC) and Rubisco activase (RCA) and the expression levels of these genes were assessed in a diverse set of tissue samples representing vegetative stages, reproductive stages and various abiotic stress treatments. Two of the widely used softwares, geNorm and Normfinder were used to evaluate the expression stabilities of these 13 candidate reference genes under different conditions. Results showed that GAPDH and CYP expression remain stable throughout in the different abiotic stress treatments, CAC and PP2A expression were relatively stable under reproductive stages and α-TUB, PP2A and ETIF3E were found to be stably expressed in vegetative stages. Further, the expression levels of Diacylglycerol acyltransferase (DGAT1), a key enzyme in triacylglycerol synthesis was analyzed to confirm the validity of reference genes identified in the study. This is the first systematic study of selection of reference genes in *S. hispanica*, and will benefit future expression studies in this crop.

Introduction

Chia (*Salvia hispanica* L.) is a flowering plant that belongs to the mint family (Lamiaceae) and is native to parts of South America and Mexico. It is an important oilseed crop, having about

60% α-linolenic acid (ALA) in its seed triacylglycerol and is the highest reported terrestrial plant source of this omega-3 fatty acid [1]. Besides ALA, the seeds of this plant are also a wealthy source of dietary fiber, protein, essential minerals and vitamins [2]. The transcriptome analysis has been performed recently on the developing seeds of *S. hispanica* and different enzymes have been identified that are proposed to play a pivotal role in the flux of ALA into triacylglycerol (TAG) [3]. To further our understanding of biosynthetic pathway leading to the accumulation of ALA in TAG, gene expression analysis is an important tool. Currently, there are no reports on internal reference genes in *S. hispanica*. Therefore, the exploration of reference genes for precise transcript normalization in seed development as well as under different experimental conditions is important for the future research to be carried out on *S. hispanica*.

Gene expression analysis has become an important tool in multiple fields of biological research. Real time quantitative reverse transcription PCR (also referred to as qRT-PCR) is now the most popular approach to measure the levels of gene expression; its main advantages being high sensitivity, efficiency, specificity and a broad quantification range [4,5]. For accurate analysis of gene expression, normalization of gene expression data with suitable internal reference gene(s) is needed [6]. Normalization is extremely important to correct the variations that might arise due to the quality and amount of template used, both of which can have an effect on the efficiency of reactions [7]. Some researchers have shown that two or more reference genes may be needed for data normalization [8,9]. This method requires the use of internal reference genes expressed constitutively to account for the variability seen at different steps of experimental procedure.

The most commonly used reference genes (also referred to as housekeeping genes) are those that are involved in basic and ubiquitous cellular processes such as the glycolytic pathway, ribosomal subunit synthesis, protein folding and components of cytoskeleton. Most of the traditionally used reference genes validated in plants include, glyceraldehydes-3-phosphate dehydrogenase *(GAPDH)*, ubiquitin *(UBQ)*, β-actin *(ACT)*, α-tubulin *(TUB)*, elongation factors *(EF)*,*and 18S* ribosomal RNA [10–14]. These genes are assumed to have constant expression levels between different samples and in response to different environmental stimuli. However, recent studies have shown that the expression levels vary under different experimental conditions in various tissues and organs [15–18]. Therefore, it is important to choose multiple stably expressed reference genes other than the traditional ones for the precise normalization of gene expression studies.

In the recent years, several new reference genes with stable expression have been identified as ideal genes such as, *F-box* protein in soybean [19], protein phosphatase 2A (*PP2A)* in buffalo grass [20] and Sorghum [21], clathrin adaptor complex subunit family protein (*CAC)* and TIP41-family protein in *Brassica juncea* [22]. Hence, it is important to choose the suitable reference genes across various experimental conditions in different species.

Many studies of stable reference genes have been reported in various plant species such as *Arabidopsis thaliana* [6, 23], *B. juncea* [22], *Oryza sativa* [16, 24], tea plant [25] tomato [26], *Glycine max* [27, 28], potato [11], Sorghum [21], flax [12], *Arachis hypogea* [29], *Triticum aestivum* [30], *Cajanus cajan* [31], *Zea mays* [32, 33] and so on. To the best of our knowledge, no studies have been carried out regarding the selection of stable reference genes in *S. hispanica*.

In the current study, we have analyzed the expression of 13 candidate reference genes in diverse tissue samples of *S. hispanica*, broadly categorized into three distinct tissue sets (abiotic stress treatments, vegetative stages and reproductive stages). Two statistical algorithms, geNorm [8], Normfinder [7], were performed to evaluate the expression stability of reference genes and to determine the optimal number of genes required for normalization in different tissue sets. Furthermore, the expression level of Diacylglycerol acyltransferase 1 (DGAT1) was checked using the stable reference genes identified in the present study to validate its effectiveness.

Materials and methods

Plant materials and stress treatments

Chia (*S. hispanica*) variety CHIAmpion B-1 was used for all the experiments. The samples have been categorized in three different experimental sets. Pre-germinated seeds on a filter paper saturated with water were grown in plastic pots in green house. Seedlings were assessed at the second true leaf stage and the healthy ones were used for i) abiotic stress group and ii) vegetative groups.

In the vegetative group, a total of 5 tissues namely, cotyledons, root, shoot, young leaf (before flowering) and mature leaf (after flowering) were collected from the plants. For stress treatments, seedlings at the second true leaf stage were transferred to separate beakers each containing 200 mM NaCl, 15% PEG 6000, 200 μM Salicylic acid (SA) and 100 μM abscisic acid (ABA) and were incubated for 5 h. For cold and heat treatments, seedlings in pots were placed in chambers at $4 \pm 1°C$ and $42 \pm 1°C$ respectively, for 5 h. Control plants were kept in room temperature at $25 \pm 1°C$. Leaves were collected from three independent biological replicates for each treatment.

In the reproductive group, a total of 6 tissues namely, bud, seed 3 days after flowering (DAF), seed 7 DAF, seed 14 DAF, seed 21 DAF and seed 28 DAF were collected from the plants grown in the green house. All tissues were immediately frozen in liquid nitrogen and stored at -80°C till further use.

Total RNA isolation and cDNA synthesis

Total RNA was extracted using Spectrum Plant Total RNA Kit (Sigma, USA) following the manufacturer's instructions. The integrity of RNA samples was verified by performing 1.5% (w/v) agarose gel electrophoresis and the quality was determined using Nanodrop 2000 Spectrophotometer (NanoDrop Technologies, Thermo Scientific, USA). The quantity was measured using Qubit 3.0 Fluorometer (Thermo Scientific, USA). Only the RNA samples with absorption ratios of $A_{260/280}$ = 2.0–2.1 and $A_{260/230}$ higher than 2.0 were used for cDNA synthesis. The integrity of all RNA samples was further measured using Agilent 2100 Bioanalyzer (Agilent technologies, Palo Alto, CA) and RIN (RNA integrity number) of > 6.5 were used for analysis (S1 File).

An aliquot of 1 μg of total RNA was used for cDNA synthesis with a final volume of 20 μL using Thermo Scientific Verso cDNA synthesis kit with RT Enhancer, to remove any contaminating DNA, by following the manufacturer's instructions. The cDNA was diluted 1:10 with nuclease-free water for use in qRT-PCR.

Selection of candidate reference genes and primer design

Based on previous literature reports of stable expression in different plant species, 13 candidate reference genes were shortlisted. The amino acid sequences of their corresponding homologs taken from TAIR (https://www.arabidopsis.org/) database served as a query for a TBLASTN search against assembled transcripts of *S. hispanica* [3]. Local blast (TBLASTN) was done using the software given in the following link (https://www.blaststation.com/intl/en/blaststation2.php). The best hit for each gene was chosen based on the bit score and E-value. The 13 candidate reference genes include *ACT11*, *EF1-α*, *ETIF3E*, *α-TUB*, *β-TUB*, *GAPDH*, *CAC*, *RCA*, *FtsH*, *PP2A*, *CYP*, *18S rRNA and SAMDC*. The candidate reference genes, their homologs in Arabidopsis and GenBank accessions are provided in Table 1. Primer pairs were designed using Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) [34] based on the following criteria: melting temperatures of 58–62°C, primer lengths of 18–20 bp, GC content of 40–60%, and amplicon lengths of 100–150 bp. All the primer sequences in the current study are listed in Table 2. The specificity of each primer pair for its target was further confirmed by Primer-BLAST.

Table 1. Candidate reference genes and their GenBank accessions in *S. hispanica*.

Gene name (abbreviation)	GenBank accession number	Arabidopsis homolog locus	Arabidopsis TBLASTN E-value	Identity (%) to *A. thaliana*
Actin11 (*ACT*)	MF621045	AT3G12110	0.0	96%
Elongation factor-1alpha (EF1-α)	MF621042	AT5G60390	0.0	95%
Eukaryotic translation Initiation factor 3E (ETIF3E)	MF621041	AT3G57290	0.0	82%
Alpha tubulin α-(TUB)	MF621046	AT5G19780	0.0	87%
Beta tubulin β-(TUB)	MF579140	AT4G20890	0.0	90%
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	MF574098	AT1G13440	1e-160	86%
Clathrin adaptor complex (CAC)	MF621044	AT5G46630	0.0	90%
Serine/threonine-protein phosphatase (PP2A)	MF579139	AT1G10430	1e-177	94%
Rubisco activase (RCA)	MF621043	AT2G39730	1e-165	85%
FtsH protease (FtsH)	MF621047	AT2G26140	0.0	74%
Cyclophilin (CYP)	MF579138	AT5G35100	1e-100	77%
Sadenosylmethionine decarboxylase (SAMDC)	MF579137	AT3G25570	1e-123	65%
18S ribosomal RNA (18SrRNA)	MF623893	AT3G41768	0.0	96%

https://doi.org/10.1371/journal.pone.0186978.t001

qRT-PCR analysis

qRT-PCR amplification was carried out on the Bio-Rad CFX96 Real-Time PCR system C1000 Thermal Cycler (Bio-Rad, Hercules, CA) using iQ SYBR Green Supermix (Bio-Rad, Hercules, CA). Each 10 µL reaction mixture included 3 µL of diluted cDNA, 5 µL of iQ SYBR Green Supermix, 0.5 µL each primer (0.5 µM) and 1 µL of Milli-Q H$_2$O. Reaction conditions include an initial denaturation step at 95˚C for 5 mins followed by 39 cycles of 95˚C for 15 s, respective annealing temperatures for 30 s and 72˚C for 30 s, after which a melt curve was generated at 55–95˚C to confirm the specificity of each primer pair. Each qRT-PCR reaction was performed with three biological and technical replicates and a no-template (NTC) control was used.

To determine the specificity of primer pairs used in the current study, agarose gel electrophoresis and a melt curve analysis were performed. Firstly, a gradient was performed using mixed cDNA to determine the optimal annealing temperature of each primer pair (S1 Table). PCR amplifications were then carried out under the conditions; 94˚C for 30 s, 39 cycles of 94˚C for 30 s, respective annealing temperatures for 30s, 68˚C for 30 s followed by a final extension step at 68˚C for 5 min. Each sample with a total reaction volume of 100 µL contained 68.5 µL of Milli-Q H$_2$O, 10 µL of Taq Buffer, 10 µL of dNTP mix, 5 µL of each primer (diluted to 10 µM), 0.5 µL of Taq Polymerase (New England Biolabs) and 1 µL of template. Amplified PCR products were run on a 2% agarose gel for confirmation. Products of correct size were gel extracted using Macharey-Nagel Gel Extraction Kit.

The gel purified DNA product concentration was measured by checking the absorbance at 260nm. The copy number was calculated by using the formula given in the following link (http://cels.uri.edu/gsc/cndna.html).

The purified DNA product was diluted with double distilled water to obtain a standard series differing by 10-fold. A series of 10 fold of five dilutions of cDNA (100–10,00,000) was

Table 2. Comprehensive details of genes and primer sets used for qRT-PCR.

S. No	Gene Name	Primers (F/R) 5'-3'	Amplicon length (bp)	Amplicon Tm (˚C)	PCR efficiency (E%)	(R²)
1	Actin11	ACTGGAATGGTCAAGGCTGG TCTTTCTGGCCCATCCCAAC	113	84.5	97.3	0.999
2	Elongation factor-1alpha	CTGTCCAGGAGCCAAAGAGG TCAACTCTTCCGACTGGCAC	103	82.5	104	0.999
3	Eukaryotic translation Initiation factor 3E	TGTGGAAACTACTCCGGTGC CTGCTGCCAATTTTCCCCAC	103	81	105.6	0.996
4	Alpha tubulin	CTCGCGCATTGACCACAAAT GAGCTTCGCTGAACTCACCT	101	82.5	104.3	0.999
5	Beta tubulin	GTACACCGGGGAAGGAATGG TTCGTCAGCAGTCGCATCTT	106	82.5	97.5	0.997
6	Glyceraldehyde-3-phosphate dehydrogenase	TCAAGGAGGAGTCTGAGGGA CAATTCCAGCCTTGGCATCG	120	81.5	109.7	0.995
7	Clathrin adaptor complex	CGGCTTCCGCGATTTACTTC GCACTCGAAAAGCATCCACC	102	83	99.9	0.999
8	Serine/threonine-protein phosphatase	GGCGATATCCACGGACAGTT CCGAATAGTACCCACGGTCG	115	82.5	94	0.999
9	Rubisco activase	CCGAAGATGACGCTGGAGAA AAGAGCTGCGTCCTTCAAGT	111	84	105.1	0.997
10	FtsH protease	GCGCCAAAGCAGTGAGTATG CTCGTGGTAGGCTGTCATCC	119	82	99.6	0.999
11	Cyclophilin	ACCAACCATTTCTTCCGGGT CTCAGCTTCCAACCTCTGCA	105	82.5	89.3	0.999
12	S-adenosyl methionine decarboxylase	GCTTCATATTCCCGGGTGCT TCCCATCACATAGGCTTCGC	116	83	89.3	0.998
13	18S ribosomal RNA	CGGGTGACGGAGAATTAGGG TACCTCCCCGTGTCAGGATT	111	84.5	109.5	0.993
14	# Diacylglycerol acyltransferase1 (DGAT1)	GGAAGGGTTGGGTGGTAAGG CCCTCTCCACGGCATACAAA	144	78.5	99.2	0.999

The PCR efficiency (E%) and regression coefficient (R²) were calculated by the CFX96 system.

Used for normalization during reproductive stages.

used to determine the PCR efficiency of each primer pair by qRT-PCR. A standard curve was generated using linear regression and the slope based on the Cq values for all dilutions in a series. The qBaseplus software calculated the PCR efficiency using following equation:

$$\text{Efficiency}\% = 10^{\left(\frac{-1}{\text{slope}}\right)} \text{X } 100\%$$

Data analysis

The quantification cycle (Cq) values of reference genes under different experimental conditions were recorded by CFX96 system. Two statistical programs, geNorm and Normfinder analysis were performed to rank the stability of suitable reference genes in different studied experimental conditions. The raw Cq values for all genes were corrected according to their PCR efficiencies and then changed to relative quantities. The relative expression values were then used in analysis to determine expression stability. geNorm calculates the gene expression stability value (*M*) based on pairwise variation (V) of a particular gene in comparison with all the other reference genes. The default set value is 1.5; gene with the lowest *M*-value is the most stably expressed one. Pair-wise variation analysis (Vn/n+1) was performed to decide the number of optimal reference genes needed for normalization in each sample set using qBaseplus

software (v:3.0; Biogazelle, Belgium) [35]. A (Vn/n+1) value <0.15 indicates that no additional reference genes are needed for normalization. Normfinder measures the gene expression stability by estimating the intra- and inter-group variations. As with geNorm, the lower the stability value of the gene, the more stable it is.

Reference gene validation

To validate the reliability of qRT-PCR data, the expression of a key target gene in triacylglycerol synthesis, DGAT1 was analyzed using the two most stable and a least stable genes identified in the study. The sequence of DGAT1 was extracted from *S. hispanica* transcriptome data and the amplification efficiency was calculated with qBaseplus software. The relative expression levels were calculated using qBaseplus software and represented as fold change [35].

Results

Verification of primer specificity and amplification efficiency analysis

A total of 13 candidate reference genes representing different classes have been selected for this study. The gene names, amplicon length and Tm, amplification efficiency (E) and correlation coefficients (R^2) are listed in the Table 2. All the primer pairs of 13 reference genes yielded a single PCR product of expected size upon amplification of pooled cDNA with no formation of primer dimers or non-specific amplification (Fig 1). Moreover, the melt curve analysis confirmed the specificity of amplicons as observed by the presence of a single peak and it also generated the melting temperatures (S2 File). In addition, the complete ORFs of all the genes were sequence verified and the amplicon sequences of the respective genes used in the current study are given in S2 Table. Amplification efficiencies of qPCR reactions ranged from 89.3% for CYP and SAMDC to 109.1% for 18S rRNA and correlation coefficients (R^2) of the standard curve ranged from 0.993 for 18S rRNA to 0.999 for CAC (S3 File).

Cq value analysis

The RNA isolated from different tissues was reverse transcribed into cDNA and used as template in qRT-PCR. The expression levels of 13 candidate reference genes in different tissue sets were examined using a SYBR Green-based qPCR assay, and the expression was determined using the quantification cycle (Cq) values. A low Cq value represents high expression level and vice-versa. The mean Cq values of candidate genes ranged from 19.65 to 28.62, with most falling in the range of 23 to 26. Across all the samples, 18S rRNA exhibited the lowest mean Cq value of 19.65 ± 1.13 (mean ± SD) indicating high expression, followed by GAPDH (20.6 ± 1.51) and EF1-α (22.87 ± 2.3), while PP2A (28.18 ± 2.2) and FtsH (28.62 ± 0.9) were the least expressed genes. The Cq value distribution of all genes in different samples are represented as a box-plot (Fig 2). The genes having high SD of Cq values exhibited more variable expression compared to those with lower SD. FtsH showed the smallest variation in gene expression (28.62 ± 0.9), while RCA showed most variable levels of expression (23.17 ± 4.5).

Expression stability analysis of candidate reference genes

The stability values of reference gene(s) may vary from one experimental condition to the other. To obtain a reliable dataset of the reference genes under different experimental conditions, two most popular algorithms, geNorm and Normfinder were performed.

geNorm analysis. geNorm analysis was performed to rank the 13 candidate reference genes by calculating expression stability value (*M*). The raw Cq values obtained for all genes were corrected according to their PCR efficiencies and then changed to relative quantities.
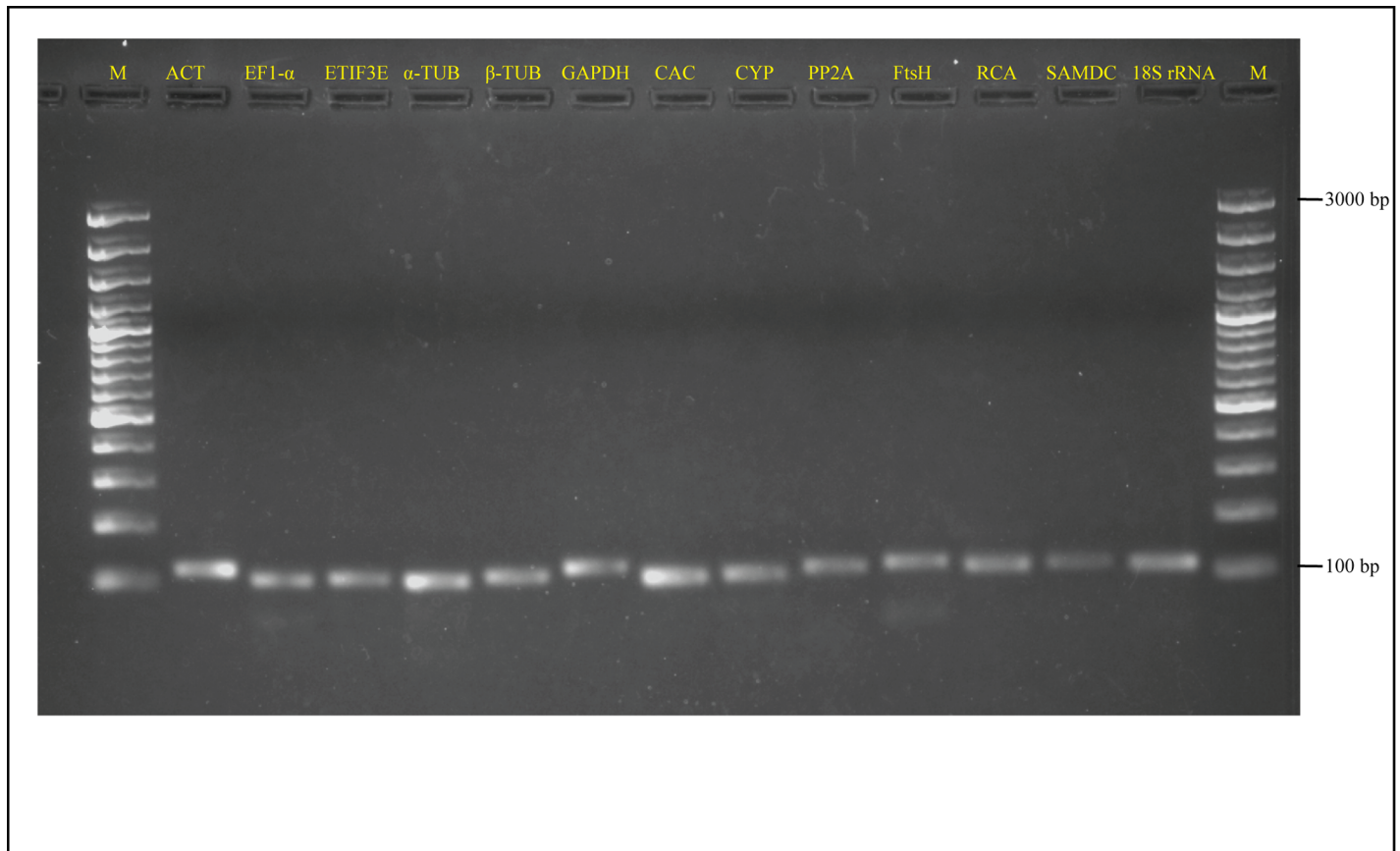
Fig 1. Agarose gel (2%) electrophoresis showing PCR amplified products of the expected size for 13 candidate reference genes. M: 100 bp marker (Thermo Scientific).

These relative expression values were used in analysis to determine expression stability. *M*-value is negatively correlated to stability and the genes with lower *M*-value are more stable. geNorm recommends a cutoff *M*-value of 1.5 and the genes with *M*-value < 1.5 are considered as stably expressed. The expression stability ranking of tested genes in different tissue sets along with their *M*-values are listed in Fig 3. In abiotic stress, the lowest *M*-value was obtained for CYP and GAPDH (0.762 and 0.777 respectively) and they represent the genes with most stable expression, followed by FtsH (0.827) and α-TUB (0.965). EF1-α (1.614), 18S rRNA (1.701) and SAMDC (1.853) were the least stable genes based on the *M*-value. In reproductive stages, CAC and PP2A (*M*-value of 0.571 and 0.624) were the most stable genes while 18S rRNA and RCA with *M*-values of 1.772 and 1.95 were least stable. In vegetative stages, GAPDH and α-TUB (0.491 and 0.5) were the most stable genes while RCA (1.699) was least stable. 12 genes had a *M*-value of less than 1.5 and thus performed well in vegetative stages.

The Vn/n+1 between normalization factors calculated by geNorm will help to determine the optimal number of reference genes needed for accurate normalization. A Vn/n+1 value <0.15 indicates that introducing an additional gene will make no significant contribution to normalization [8]. In the vegetative group, the V3/4 value was less than 0.15 (0.135), which suggested that three genes are sufficient for normalization. In the abiotic stress and reproductive group, the lowest pairwise variation value was above 0.15 (Fig 4).
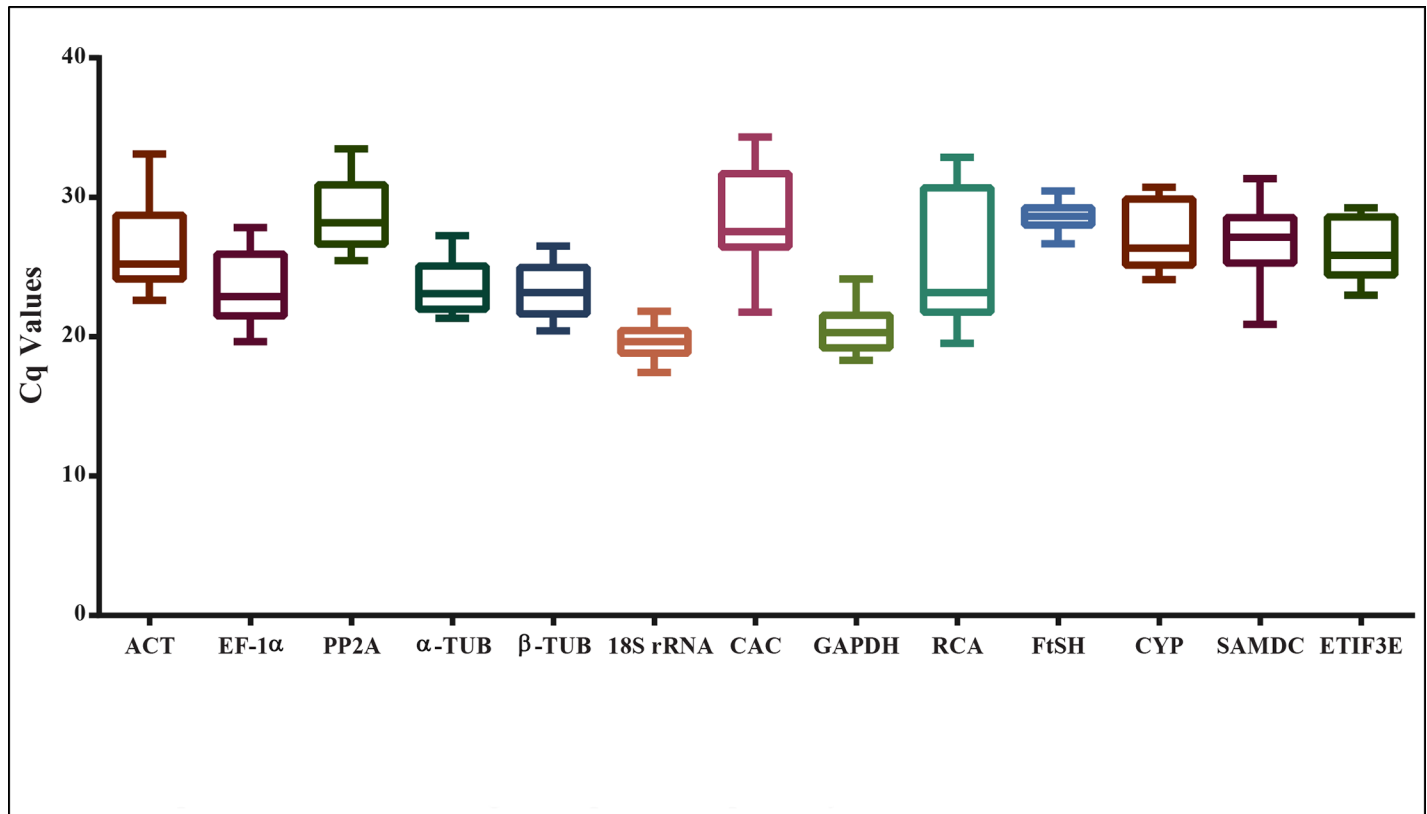
Fig 2. Distribution of quantification cycle (Cq) values across all experimental sets in *S. hispanica*. The line across the box-plot depicts the median. Lower and upper boxes represent 25th and 75th percentiles. Whiskers represent the lowest and highest Cq values respectively.
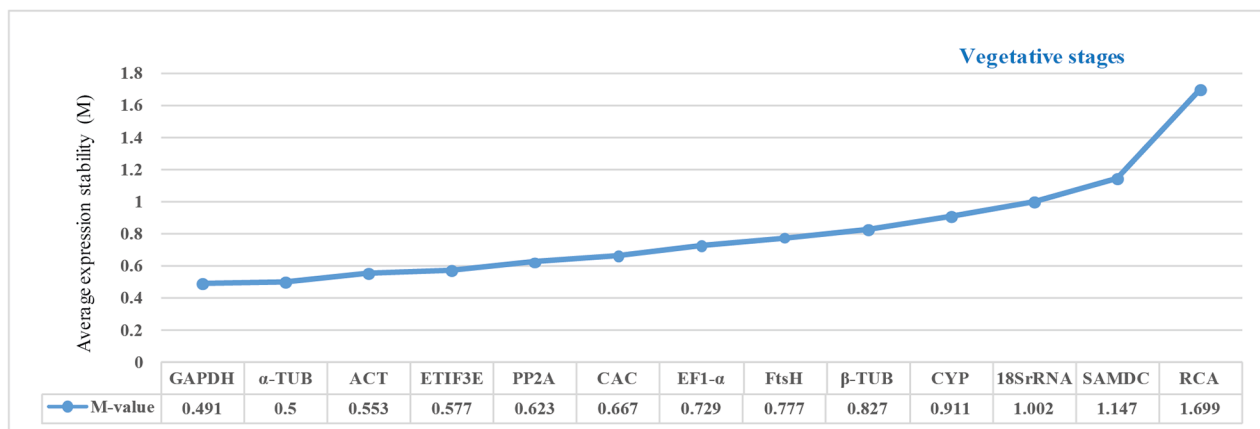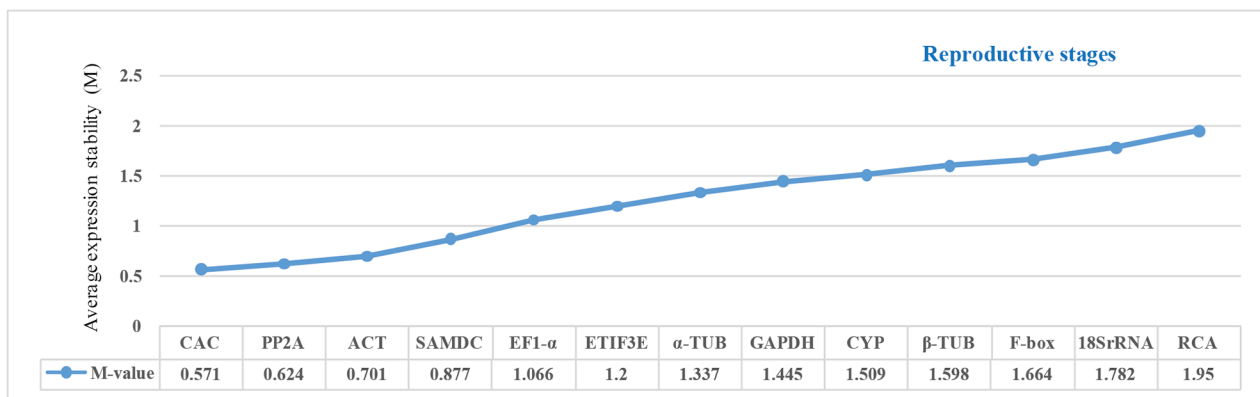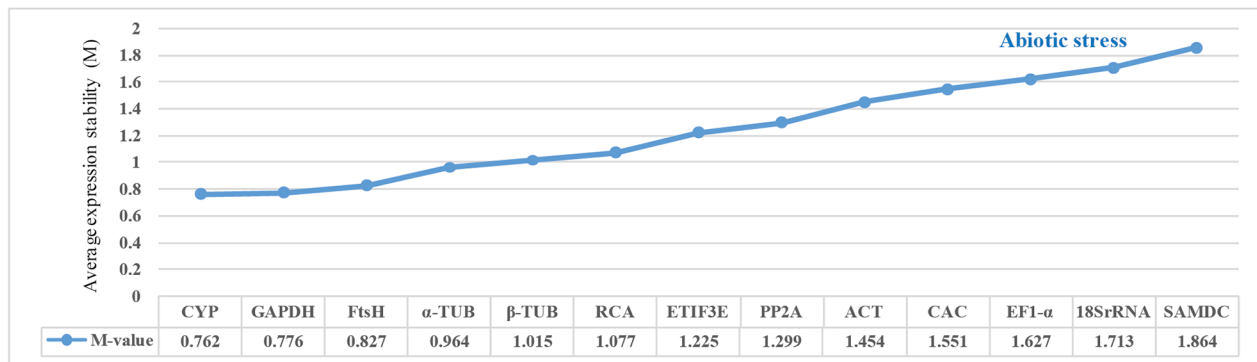
Normfinder analysis. Normfinder is a Microsoft Excel™ based program that requires the input data to be given on a linear scale. Therefore, the raw Ct values obtained for each gene should be transformed into relative expression levels according to the given formula $2^{-\Delta Ct}$ (ΔCt = Ct value of each sample—minimum Ct value).

These values are then imported into Normfinder software (https://moma.dk/normfinder-software) to analyze gene expression stability. It measures the stability based on intra and inter-group expression variations, and a low stability value indicates stable gene expression. The gene stability rankings generated by Normfinder are shown for each group and the stability values are also included in Fig 5. In the abiotic stress group, CYP (stability value 0.264) and GAPDH (0.278) were ranked the most stable genes (also ranked similarly by geNorm), while SAMDC (0.929) was the least stable gene. In the reproductive group, α-TUB (0.389) and PP2A (0.451) emerged as the most stable genes, followed by CAC (0.730) while geNorm ranked α-TUB at seventh place. 18S rRNA (1.464) and RCA (1.829) were the two least stable genes, as with the geNorm analysis. In the vegetative group, FtsH (0.253) and α-TUB (0.255) occupied the top positions, while geNorm ranked FtsH at a medium position. Finally, Normfinder also reported SAMDC and RCA as the least stable genes in accordance with geNorm.

Validation of the best and least ranked reference genes

To validate the genes identified in the current study, we checked the expression levels of a key enzyme, DGAT1 in different stages of seed development (reproductive stages).

	CYP	GAPDH	FtsH	α-TUB	β-TUB	RCA	ETIF3E	PP2A	ACT	CAC	EF1-α	18SrRNA	SAMDC
M-value	0.762	0.776	0.827	0.964	1.015	1.077	1.225	1.299	1.454	1.551	1.627	1.713	1.864

	CAC	PP2A	ACT	SAMDC	EF1-α	ETIF3E	α-TUB	GAPDH	CYP	β-TUB	F-box	18SrRNA	RCA
M-value	0.571	0.624	0.701	0.877	1.066	1.2	1.337	1.445	1.509	1.598	1.664	1.782	1.95

	GAPDH	α-TUB	ACT	ETIF3E	PP2A	CAC	EF1-α	FtsH	β-TUB	CYP	18SrRNA	SAMDC	RCA
M-value	0.491	0.5	0.553	0.577	0.623	0.667	0.729	0.777	0.827	0.911	1.002	1.147	1.699

Most stable genes   **Least stable genes**

Fig 3. Gene expression stability (M) of 13 candidate reference genes across different experimental groups as assayed by geNorm. M-value is inversely related to gene stability. The cutoff M-value was set at 1.5; the direction of arrows indicate the most stable and least stable genes in the graph. The most stable genes are listed on the left and least stable genes are listed on the right.

Triacylglycerols are the major storage forms of energy and have crucial roles in seed oil accumulation, germination and seedling development. DGAT1 plays a crucial role in the final step of TAG synthesis by catalyzing final acylation in pathway of diacylglycerol (DAG) to form TAG [36]. Overexpression of DGAT1 has been performed in different plant species and this has enhanced the TAG content [37, 38]. It is considered the rate-limiting enzyme in TAG accumulation, and is being exploited to increase the oil content through genetic engineering in oil plants and microalga for economic benefit [36, 39]. The expression of DGAT1 was normalized using the two best genes, PP2A and CAC in combination and individually. For both genes, the transcript abundance increased at second stage (7DAF) compared to first stage
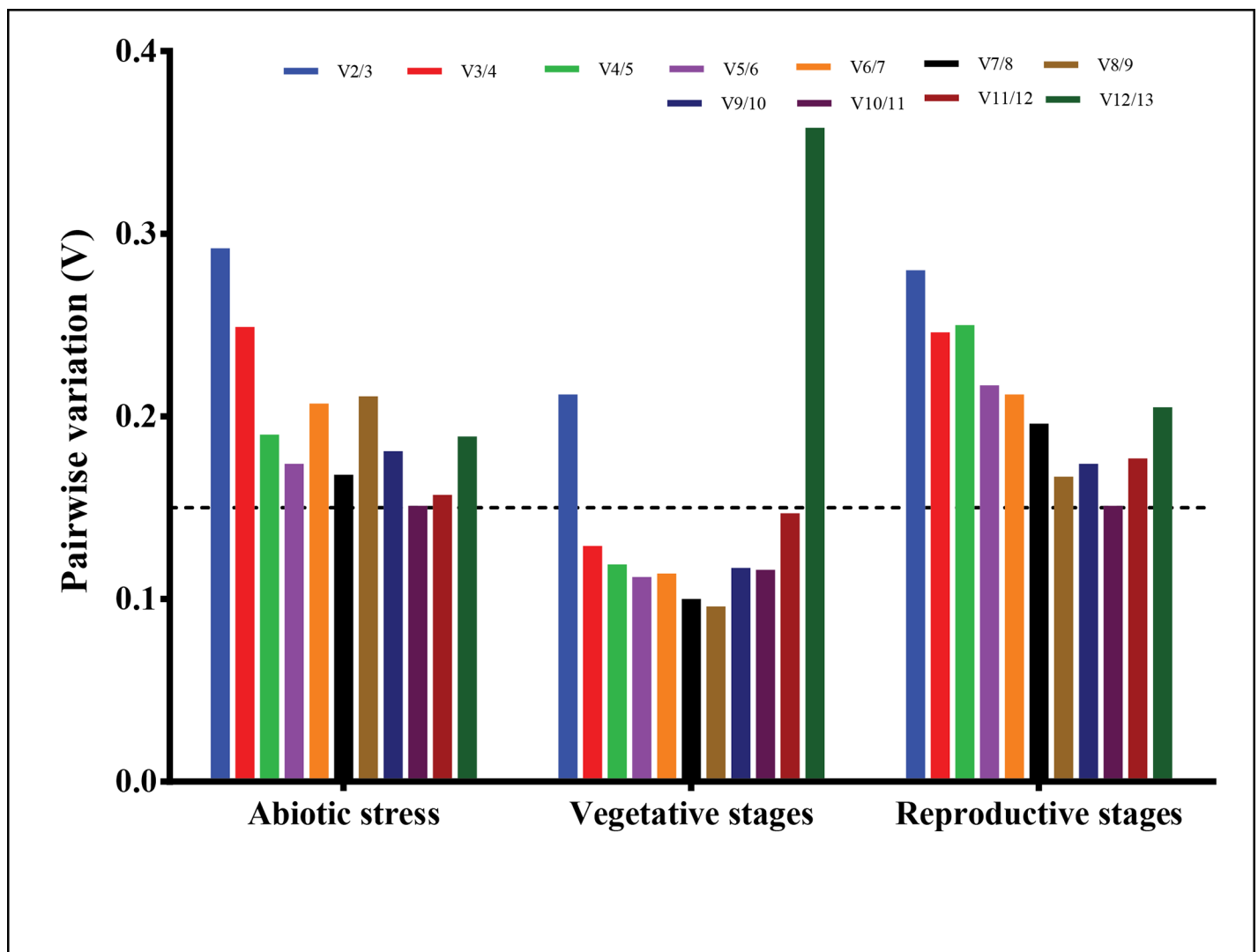

Fig 4. Optimal number of references genes required for normalization across different experimental sets as calculated by geNorm. Dotted line indicates cut-off value of 0.15, below which additional reference genes are not necessary for normalization.
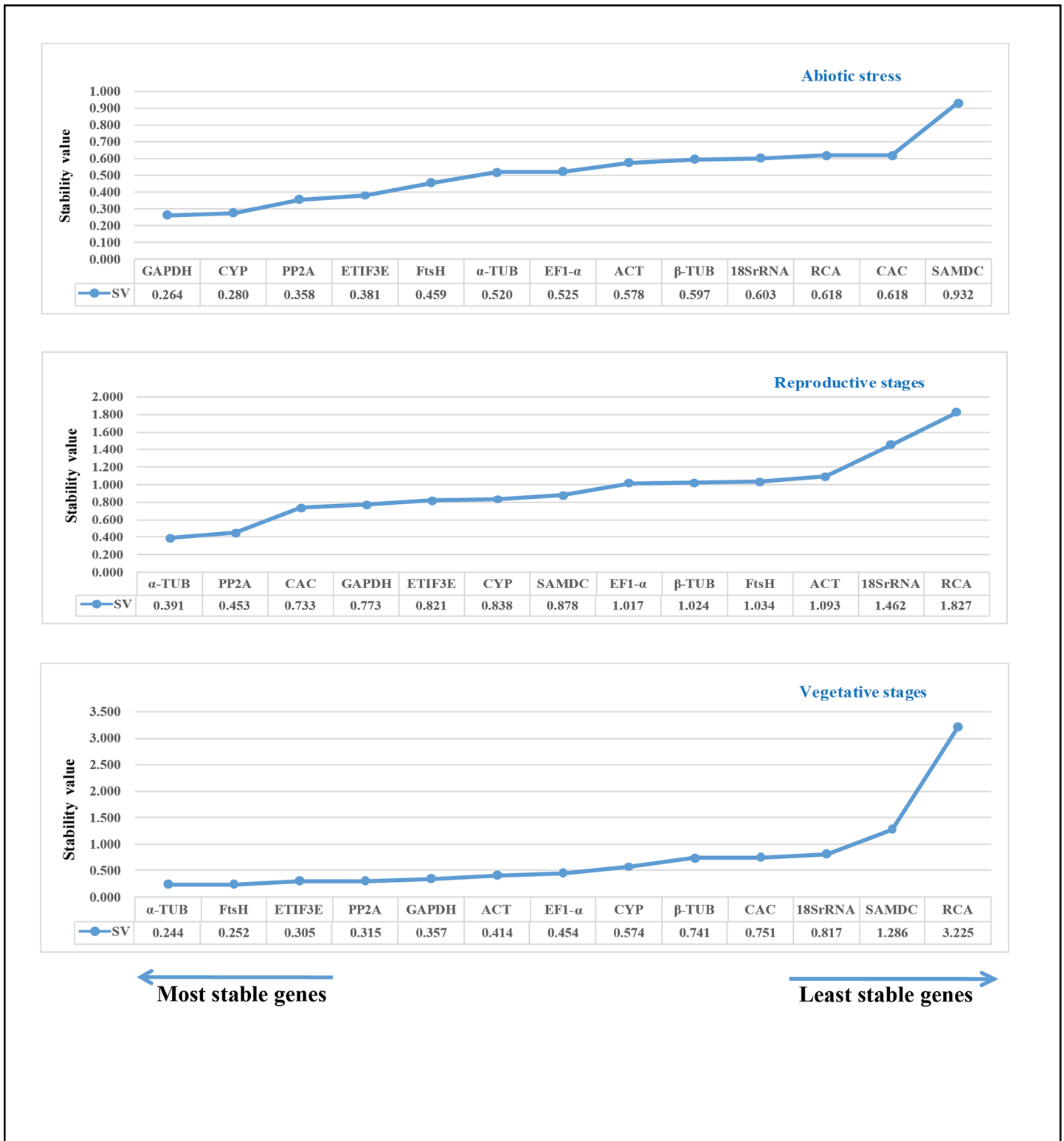
Fig 5. Stability values of 13 candidate reference genes across different experimental groups as assayed by Normfinder. Lower stability values indicate highly stable genes and higher stability values indicate less stable genes. The direction of the arrows indicate the most stable and least stable genes in the graph. The most stable genes are listed on the left and least stable genes are listed on the right.
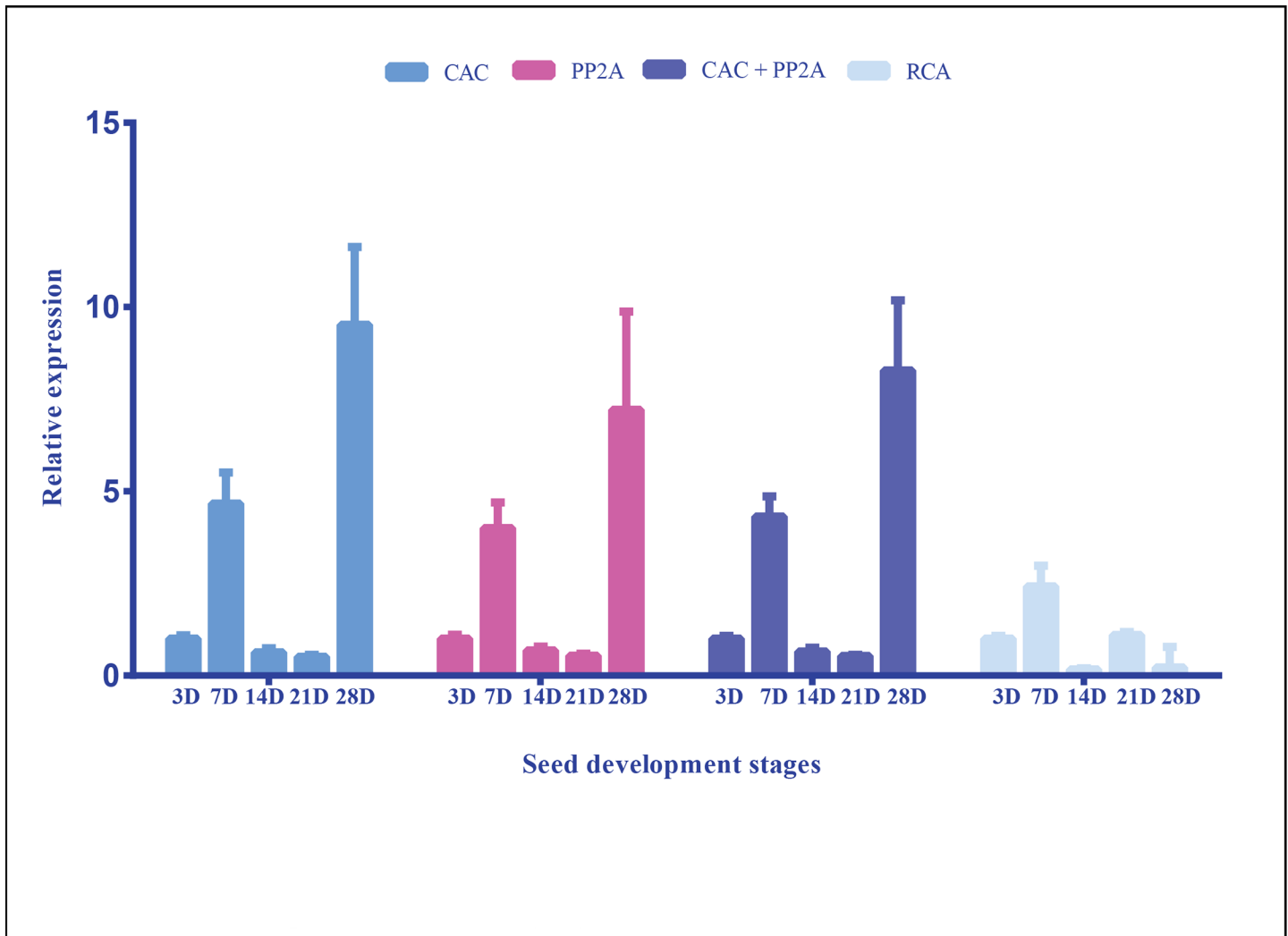
https://doi.org/10.1371/journal.pone.0186978.g005

Fig 6. Validation of the reference genes in reproductive stage group. The relative expression levels of DGAT1 was normalized by using the two most stable genes, PP2A and CAC in combination and individually, and also with the least stable gene, RCA. D represents DAF–Days after flowering. Values shown are relative expression levels ± SD (n = 3).

(3DAF), and finally increased to a maximum at the final stage of seed maturation. When PP2A and CAC were used in combination, the same pattern was seen. However, normalization with the least reported stable gene, RCA, the expression pattern was completely changed (Fig 6). The results highlight the importance of validated reference genes for target normalization.

Discussion

Currently, qRT-PCR is considered as the most appropriate method for gene expression profiling due to its high sensitivity, accuracy and reproducibility [40]. Precise measurement of target gene expression requires the use of suitable reference genes that are stably expressed. However, random or inappropriate selection of reference genes for normalization may lead to misinterpretation of results [41]. Recent reports have shown that the expression of commonly used reference genes varies with different experimental conditions. Therefore, it is mandatory to validate reference genes for each experimental condition in different species [15]. Hence,

the aim of the study was to identify ideal reference genes with stable expression levels under different experimental conditions in *S. hispanica*.

We checked the expression levels and stability of 13 candidate reference genes across different tissue samples, i.e. vegetative stages, reproductive stages and six abiotic stress treatments in *S. hispanica*. Two softwares namely, geNorm and Normfinder were used to identify the stable reference genes in the studied experimental conditions. As the two algorithms operate differently, the results were slightly different in some conditions. For example, in reproductive stages, geNorm ranked CAC and PP2A as the two most stable genes while Normfinder ranked α-TUB and PP2A as the best genes. α-TUB was ranked at a medium position in geNorm, however CAC was ranked third in Normfinder. In vegetative stages, FtsH was ranked at a medium position in geNorm but was ranked second in Normfinder. The gene α-TUB was ranked top in both algorithms. However, both geNorm and Normfinder recommended CYP and GAPDH as the suitable reference genes in abiotic stress treatments. In addition, the two least stable genes for each set were the same in geNorm and Normfinder. The apparent divergence in ranking is due to the discrepancies in the two algorithms to measure stability [42]. The geNorm works upon stepwise exclusion of the least stable genes, which is based on the expression stability value (*M*) [8]. Normfinder, an excel based tool takes inter- and intra-group variations into account and combines them into a stability value [7]. The difference in the ranking by the two algorithms was also observed in many other studies [14, 43–45].

Our results are in accordance with previous studies, wherein CYP was reported to be the stably expressed gene under abiotic stress conditions in peanut [29], sorghum [21] and salt-stressed *S. tuberosum* [11]. The CYP gene also performed well in switchgrass [46]. Similarly, GAPDH was reported to be a stably expressed gene under stress conditions in chickpea [13], tall fescue [44], heat and salinity stress in pigeon pea [31], and hormone stimuli in parsley [14]. In contrast, GAPDH was shown to be a least stable gene in *B. juncea*. SAMDC, the least stable gene in abiotic stress treatment was reported to be unstable in switchgrass [46] and 18S rRNA was also reported as unstable in Jute [42] and oxytropis [47]. Further, PP2A was found to be stably expressed in maturing *B. napus* embryos [48], developmental stages in *A. thaliana* [6] and life stages in Striga [49]. CAC was also reported to be the stable gene during development in *B. juncea* [22] and *B. arvensis* [50]. Similar to the present study, RCA was also reported to be the least stable gene in reproductive development in papaya [51].

Data normalization with a single reference gene may give rise to erroneous results [52] and thus usage of two or more reference genes has become the common way. geNorm has suggested to use three reference genes for vegetative stages as the V3/4 value was less than 0.15. For reproductive stages and abiotic stress, the value was above 0.15. However, the V (pairwise variation) value of 0.15 should not be considered as an absolute one and high V values have been reported in few species. [53–55]. Vandesompele et al., 2002 have proposed that minimum three reference genes are necessary for precise normalization [8]. However, the usage of multiple reference genes may lead to misinterpretation of results.

DGAT1, a key enzyme involved in triacylglycerol synthesis was used to validate the reference genes identified in the study. DGAT catalyzes the synthesis of TAG from diacylglycerol and is pivotal to oil accumulation. The two best reference genes, CAC and PP2A were used alone and in combination for normalization. With both genes, there was an increase in transcript expression at 7DAF (second stage) compared to 3DAF (first stage) and then finally increased in the final stage (28DAF). When normalized using the least stable gene RCA, the expression pattern was not conclusive.

In conclusion, we evaluated the expression stabilities of 13 candidate reference genes in different abiotic stresses (NaCl, PEG, SA, ABA, heat, and cold), vegetative and reproductive stages of *S. hispanica*. Two statistical algorithms, geNorm and Normfinder were used to assess

the stability of reference genes, and some slight variation was observed because of the way both algorithms operate to calculate stability. Overall, we propose that in *S. hispanica*, CYP and GAPDH can be used for abiotic stresses; PP2A and CAC for reproductive development, α-TUB, ETIF3E and PP2A for vegetative stages. In addition, the expression pattern of DGAT1 was analyzed using the best reference genes identified in the study. The reference genes identified will benefit future studies of gene expression in various tissues of S. *hispanica*.

Supporting information

S1 File. RNA integrity values determined by Bioanalyzer.
(PPTX)

S2 File. Melt curves of 13 genes under different experimental conditions.
(PPTX)

S3 File. Amplification efficiencies of primers used in the study.
(PPTX)

S1 Table. Annealing temperatures of primers used in the study.
(DOCX)

S2 Table. Sequences of the PCR amplicons.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Sunny D. Rupwate, Ajay W. Tumaney.

Data curation: Rahul Gopalam, Sunny D. Rupwate, Ajay W. Tumaney.

Formal analysis: Rahul Gopalam, Sunny D. Rupwate, Ajay W. Tumaney.

Funding acquisition: Ajay W. Tumaney.

Investigation: Rahul Gopalam, Sunny D. Rupwate, Ajay W. Tumaney.

Methodology: Ajay W. Tumaney.

Project administration: Sunny D. Rupwate, Ajay W. Tumaney.

Resources: Ajay W. Tumaney.

Supervision: Sunny D. Rupwate, Ajay W. Tumaney.

Validation: Rahul Gopalam.

Visualization: Rahul Gopalam, Ajay W. Tumaney.

Writing – original draft: Rahul Gopalam.

Writing – review & editing: Rahul Gopalam, Sunny D. Rupwate, Ajay W. Tumaney.

References

1. Coates W. Protein content, oil content and fatty acid profiles as potential criteria to determine the origin of commercially grown chia (Salvia hispanica L.). Industrial Crops and Products. 2011 Sep 30; 34 (2):1366–71.

2. Muñoz LA, Cobos A, Diaz O, Aguilera JM. Chia seed (Salvia hispanica): an ancient grain and a new functional food. Food reviews international. 2013 Oct 2; 29(4):394–408.

3. Sreedhar RV, Kumari P, Rupwate SD, Rajasekharan R, Srinivasan M. Exploring triacylglycerol biosynthetic pathway in developing seeds of Chia (Salvia hispanica L.): a transcriptomic approach. PloS one. 2015 Apr 13; 10(4):e0123580. https://doi.org/10.1371/journal.pone.0123580 PMID: 25875809

4. Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. Journal of molecular endocrinology. 2002 Aug 1; 29(1):23–39. https://doi.org/10.1677/jme.0.0290023 PMID: 12200227

5. Bustin SA, Benes V, Nolan T, Pfaffl MW. Quantitative real-time RT-PCR–a perspective. Journal of molecular endocrinology. 2005 Jun 1; 34(3):597–601. https://doi.org/10.1677/jme.1.01755 PMID: 15956331

6. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant physiology. 2005 Sep 1; 139 (1):5–17. https://doi.org/10.1104/pp.105.063743 PMID: 16166256

7. 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 Aug 1; 64(15):5245–50. https://doi.org/10.1158/0008-5472.CAN-04-0496 PMID: 15289330

8. 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 biology. 2002 Jun 18; 3(7):research0034–1. https://doi.org/10.1186/gb-2002-3-7-research0034 PMID: 12184808

9. Schmid H, Cohen CD, Henger A, Irrgang S, Schlöndorff D, Kretzler M. Validation of endogenous controls for gene expression analysis in microdissected human renal biopsies. Kidney international. 2003 Jul 31; 64(1):356–60. https://doi.org/10.1046/j.1523-1755.2003.00074.x PMID: 12787429

10. 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 Feb 15; 11(1):15. https://doi.org/10.1186/1471-2199-11-15 PMID: 20156357

11. Nicot N, Hausman JF, Hoffmann L, Evers D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. Journal of experimental botany. 2005 Nov 1; 56 (421):2907–14. https://doi.org/10.1093/jxb/eri285 PMID: 16188960

12. Huis R, Hawkins S, Neutelings G. Selection of reference genes for quantitative gene expression normalization in flax (Linum usitatissimum L.). BMC plant biology. 2010 Apr 19; 10(1):71. https://doi.org/10.1186/1471-2229-10-71 PMID: 20403198

13. Garg R, Sahoo A, Tyagi AK, Jain M. Validation of internal control genes for quantitative gene expression studies in chickpea (Cicer arietinum L.). Biochemical and biophysical research communications. 2010 May 28; 396(2):283–8. https://doi.org/10.1016/j.bbrc.2010.04.079 PMID: 20399753

14. Li J, Han J, Hu Y, Yang J. Selection of reference genes for quantitative real-time PCR during flower development in tree Peony (Paeonia suffruticosa Andr.). Frontiers in plant science. 2016; 7. https://doi.org/10.3389/fpls.2016.00516 PMID: 27148337

15. Schmittgen TD, Zakrajsek BA. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. Journal of biochemical and biophysical methods. 2000 Nov 20; 46(1):69–81. https://doi.org/10.1016/S0165-022X(00)00129-9 PMID: 11086195

16. 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 Jun 30; 345(2):646–51. https://doi.org/10.1016/j.bbrc.2006.04.140 PMID: 16690022

17. Hong SY, Seo PJ, Yang MS, Xiang F, Park CM. Exploring valid reference genes for gene expression studies in Brachypodium distachyon by real-time PCR. BMC plant biology. 2008 Nov 7; 8(1):112. https://doi.org/10.1186/1471-2229-8-112 PMID: 18992143

18. Wang HL, Chen J, Tian Q, Wang S, Xia X, Yin W. Identification and validation of reference genes for Populus euphratica gene expression analysis during abiotic stresses by quantitative real-time PCR. Physiologia plantarum. 2014 Nov 1; 152(3):529–45. https://doi.org/10.1111/ppl.12206 PMID: 24720378

19. Yim AK, Wong JW, Ku YS, Qin H, Chan TF, Lam HM. Using RNA-Seq data to evaluate reference genes suitable for gene expression studies in soybean. PloS one. 2015 Sep 8; 10(9):e0136343. https://doi.org/10.1371/journal.pone.0136343 PMID: 26348924

20. Li W, Qian YQ, Han L, Liu JX, Sun ZY. Identification of suitable reference genes in buffalo grass for accurate transcript normalization under various abiotic stress conditions. Gene. 2014 Aug 15; 547 (1):55–62. https://doi.org/10.1016/j.gene.2014.06.015 PMID: 24914494

21. Reddy PS, Reddy DS, Sivasakthi K, Bhatnagar-Mathur P, Vadez V, Sharma KK. Evaluation of sorghum [Sorghum bicolor (L.)] reference genes in various tissues and under abiotic stress conditions for quantitative real-time PCR data normalization. Frontiers in plant science. 2016; 7. https://doi.org/10.3389%2Ffpls.2016.00529 PMID: 27200008

22. Chandna R, Augustine R, Bisht NC. Evaluation of candidate reference genes for gene expression normalization in Brassica juncea using real time quantitative RT-PCR. PloS one. 2012 May 11; 7(5): e36918. https://doi.org/10.1371/journal.pone.0036918 PMID: 22606308

23. Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A. Normalisation of real-time RT-PCR gene expression measurements in Arabidopsis thaliana exposed to increased metal concentrations. Planta. 2008 May 1; 227(6):1343–9. https://doi.org/10.1007/s00425-008-0706-4 PMID: 18273637

24. Ji Y, Tu P, Wang K, Gao F, Yang W, Zhu Y, et al. Defining reference genes for quantitative real-time PCR analysis of anther development in rice. Acta Biochim. Biophys. Sin.(Shanghai). 2014 Apr 1; 46:305–12. https://doi.org/10.1093/abbs/gmu002 PMID: 24492537

25. Wu ZJ, Tian C, Jiang Q, Li XH, Zhuang J. Selection of suitable reference genes for qRT-PCR normalization during leaf development and hormonal stimuli in tea plant (Camellia sinensis). Scientific reports. 2016; 6. https://doi.org/10.1038%2Fsrep19748 PMID: 26813576

26. 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 Dec 22; 8(1):131. https://doi.org/10.1186/1471-2229-8-131 PMID: 19102748

27. Jian B, Liu B, Bi Y, Hou W, Wu C, Han T. Validation of internal control for gene expression study in soybean by quantitative real-time PCR. BMC molecular biology. 2008 Jun 23; 9(1):59. https://doi.org/10.1186/1471-2199-9-59 PMID: 18573215

28. Nakayama TJ, Rodrigues FA, Neumaier N, Marcelino-Guimarães FC, Farias JR, De Oliveira MC, et al. Reference genes for quantitative real-time polymerase chain reaction studies in soybean plants under hypoxic conditions. Genet Mol Res. 2014 Jan 1; 13:860–71. https://doi.org/10.4238/2014.February.13.4 PMID: 24615050

29. Reddy DS, Bhatnagar-Mathur P, Cindhuri KS, Sharma KK. Evaluation and validation of reference genes for normalization of quantitative real-time PCR based gene expression studies in peanut. PloS one. 2013 Oct 22; 8(10):e78555. https://doi.org/10.1371/journal.pone.0078555 PMID: 24167633

30. Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. BMC molecular biology. 2009 Feb 20; 10(1):11. https://doi.org/10.1186/1471-2199-10-11 PMID: 19232096

31. Sinha P, Singh VK, Suryanarayana V, Krishnamurthy L, Saxena RK, Varshney RK. Evaluation and validation of housekeeping genes as reference for gene expression studies in pigeonpea (Cajanus cajan) under drought stress conditions. PloS one. 2015 Apr 7; 10(4):e0122847. https://doi.org/10.1371/journal.pone.0122847 PMID: 25849964

32. Manoli A, Sturaro A, Trevisan S, Quaggiotti S, Nonis A. Evaluation of candidate reference genes for qPCR in maize. Journal of plant physiology. 2012 May 15; 169(8):807–15. https://doi.org/10.1016/j.jplph.2012.01.019 PMID: 22459324

33. Lin F, Jiang L, Liu Y, Lv Y, Dai H, Zhao H. Genome-wide identification of housekeeping genes in maize. Plant molecular biology. 2014 Nov 1; 86(4–5):543–54. https://doi.org/10.1007/s11103-014-0246-1 PMID: 25209110

34. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic acids research. 2012 Aug 1; 40(15):e115-. https://doi.org/10.1093/nar/gks596 PMID: 22730293

35. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome biology. 2007 Feb 9; 8(2):R19. https://doi.org/10.1186/gb-2007-8-2-r19 PMID: 17291332

36. Lung SC, Weselake RJ. Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. Lipids. 2006 Dec 1; 41(12):1073–88. https://doi.org/10.1007/s11745-006-5057-y PMID: 17269553

37. Andrianov V, Borisjuk N, Pogrebnyak N, Brinker A, Dixon J, Spitsin S, et al. Tobacco as a production platform for biofuel: overexpression of Arabidopsis DGAT and LEC2 genes increases accumulation and

shifts the composition of lipids in green biomass. Plant biotechnology journal. 2010 Apr 1; 8(3):277–87. https://doi.org/10.1111/j.1467-7652.2009.00458.x PMID: 20051035

38. Taylor DC, Zhang Y, Kumar A, Francis T, Giblin EM, Barton DL, et al. Molecular modification of triacyl-glycerol accumulation by over-expression of DGAT1 to produce canola with increased seed oil content under field conditions This paper is one of a selection of papers published in a Special Issue from the National Research Council of Canada–Plant Biotechnology Institute. Botany. 2009 Jun 9; 87(6):533–43.

39. Weselake RJ. Storage lipids. Plant Lipids: Biology, Utilisation and Manipulation. 2005:162–225.

40. Gutierrez L, Mauriat M, Pelloux J, Bellini C, Van Wuytswinkel O. Towards a systematic validation of references in real-time RT-PCR. The Plant Cell. 2008 Jul 1; 20(7):1734–5. https://doi.org/10.1105/tpc.108.059774 PMID: 18664615

41. Dheda K, Huggett JF, Chang JS, Kim LU, Bustin SA, Johnson MA, et al. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Analytical biochemistry. 2005 Sep 1; 344(1):141–3. https://doi.org/10.1016/j.ab.2005.05.022 PMID: 16054107

42. Niu X, Qi J, Zhang G, Xu J, Tao A, Fang P, et al. Selection of reliable reference genes for quantitative real-time PCR gene expression analysis in Jute (Corchorus capsularis) under stress treatments. Frontiers in plant science. 2015 Oct 14; 6:848. https://doi.org/10.3389/fpls.2015.00848 PMID: 26528312

43. Kong Q, Yuan J, Gao L, Zhao S, Jiang W, Huang Y, et al. Identification of suitable reference genes for gene expression normalization in qRT-PCR analysis in watermelon. PLoS One. 2014 Feb 28; 9(2): e90612. https://doi.org/10.1371/journal.pone.0090612 PMID: 24587403

44. Yang Z, Chen Y, Hu B, Tan Z, Huang B. Identification and validation of reference genes for quantification of target gene expression with quantitative real-time PCR for tall fescue under four abiotic stresses. PloS one. 2015 Mar 18; 10(3):e0119569. https://doi.org/10.1371/journal.pone.0119569 PMID: 25786207

45. Ma R, Xu S, Zhao Y, Xia B, Wang R. Selection and validation of appropriate reference genes for quantitative real-time PCR analysis of gene expression in Lycoris aurea. Frontiers in plant science. 2016; 7. https://doi.org/10.3389%2Ffpls.2016.00536 PMID: 27200013

46. Gimeno J, Eattock N, Van Deynze A, Blumwald E. Selection and validation of reference genes for gene expression analysis in switchgrass (Panicum virgatum) using quantitative real-time RT-PCR. PloS one. 2014 Mar 12; 9(3):e91474. https://doi.org/10.1371/journal.pone.0091474 PMID: 24621568

47. Zhuang H, Fu Y, He W, Wang L, Wei Y. Selection of appropriate reference genes for quantitative real-time PCR in Oxytropis ochrocephala Bunge using transcriptome datasets under abiotic stress treatments. Frontiers in plant science. 2015; 6. https://doi.org/10.3389/fpls.2015.00475 PMID: 26175743

48. Chen X, Truksa M, Shah S, Weselake RJ. A survey of quantitative real-time polymerase chain reaction internal reference genes for expression studies in Brassica napus. Analytical biochemistry. 2010 Oct 1; 405(1):138–40. https://doi.org/10.1016/j.ab.2010.05.032 PMID: 20522329

49. Fernández-Aparicio M, Huang K, Wafula EK, Honaas LA, Wickett NJ, Timko MP, et al. Application of qRT-PCR and RNA-Seq analysis for the identification of housekeeping genes useful for normalization of gene expression values during Striga hermonthica development. Molecular biology reports. 2013 Apr 1; 40(4):3395–407. https://doi.org/10.1007/s11033-012-2417-y PMID: 23271128

50. Gadkar VJ, Filion M. Validation of endogenous reference genes in Buglossoides arvensis for normalizing RT-qPCR-based gene expression data. SpringerPlus. 2015 Apr 15; 4(1):178. https://doi.org/10.1186/s40064-015-0952-4 PMID: 25918683

51. Zhu X, Li X, Chen W, Chen J, Lu W, Chen L, et al. Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. PloS one. 2012 Aug 31; 7(8): e44405. https://doi.org/10.1371/journal.pone.0044405 PMID: 22952972

52. Veazey KJ, Golding MC. Selection of stable reference genes for quantitative rt-PCR comparisons of mouse embryonic and extra-embryonic stem cells. PloS one. 2011 Nov 10; 6(11):e27592. https://doi.org/10.1371/journal.pone.0027592 PMID: 22102912

53. 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 Apr 15; 399(2):257–61. https://doi.org/10.1016/j.ab.2009.12.008 PMID: 20005862

54. Tian C, Jiang Q, Wang F, Wang GL, Xu ZS, Xiong AS. Selection of suitable reference genes for qPCR normalization under abiotic stresses and hormone stimuli in carrot leaves. PLoS One. 2015 Feb 6; 10 (2):e0117569. https://doi.org/10.1371/journal.pone.0117569 PMID: 25658122

55. Marum L, Miguel A, Ricardo CP, Miguel C. Reference gene selection for quantitative real-time PCR normalization in Quercus suber. PloS one. 2012 Apr 18; 7(4):e35113. https://doi.org/10.1371/journal.pone.0035113 PMID: 22529976.