Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean

Franceli Rodrigues Kulcheski [a], Francismar Correa Marcelino-Guimaraes [b], Alexandre Lima Nepomuceno [b], Ricardo Vilela Abdelnoor [b], Rogério Margis [a,*]

[a] Centre of Biotechnology, Laboratory of Genomes and Plant Population, Building 43431, Federal University of Rio Grande do Sul–UFRGS, P.O. Box 15005, CEP 91501-970, Porto Alegre, RS, Brazil
[b] EMBRAPA Soja, Rodovia Carlos João Strass, Distrito de Warta, CEP 86001-970, Londrina, PR, Brazil

ARTICLE INFO

ABSTRACT

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is a robust and widely applied technique used to investigate gene expression. However, for correct analysis and interpretation of results, the choice of a suitable gene to use as an internal control is a crucial factor. These genes, such as housekeeping genes, should have a constant expression level in different tissues and across different conditions. The advances in genome sequencing have provided high-throughput gene expression analysis and have contributed to the identification of new genes, including microRNAs (miRNAs). The miRNAs are fundamental regulatory genes of eukaryotic genomes, acting on several biological functions. In this study, miRNA expression stability was investigated in different soybean tissues and genotypes as well as after abiotic or biotic stress treatments. The present study represents the first investigation into the suitability of miRNAs as housekeeping genes in plants. The transcript stability of 10 miRNAs was compared to those of six previously reported housekeeping genes for the soybean. In this study, we provide evidence that the expression stabilities of miR156b and miR1520d were the highest across the soybean experiments. Furthermore, these miRNAs genes were more stable than the most commonly protein-coding genes used in soybean gene expression studies involving RT-qPCR.

© 2010 Elsevier Inc. All rights reserved.

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR)[1] is a powerful technique that is commonly used to study gene expression due to its high sensitivity, good reproducibility, and wide dynamic quantification range. Many experiments have used RT-qPCR to measure the gene expression variation between two different biological groups, for example, a treated versus a control pool of samples. However, in many cases, the variation is caused by a discrepancy in the initial sample amount, efficiency of nucleic acid recovery, RNA degradation, differences in sample quality, pipetting errors, or variation in cDNA synthesis efficiency [1]. In such cases, the normalization process is fundamental for correcting nonspecific variation that can affect quantification results [2].

Several normalization strategies have been proposed, but the most commonly used method is relative quantification. Using this method, the expression level of a gene of interest is normalized to the expression level of a reference gene [2]. An ideal reference gene should be expressed at a constant level across various conditions, such as developmental stages or tissue types, and its expression

should be unaffected by experimental parameters [3–5]. Currently, the reference genes frequently used are protein-coding genes, especially "housekeeping genes," which are involved in basic cellular processes [6], such as cytoskeleton construction (actins), glycolysis (glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*)), protein folding (cyclophilin), synthesis of ribosomal subunits (rRNA), electron transport (succinate dehydrogenase complex, *SDH*), and protein degradation (ubiquitin) [7]. These genes are thought to have constant expression levels among different samples and are frequently used to normalize gene expression levels without proper validation [7]. However, the expression of a number of housekeeping genes, although constant under some experimental conditions, varies considerably under other conditions [6,8–11]. In fact, when these genes are used as reference genes under experimental conditions that differ from those conditions in which their stability of expression was originally tested, the results can be misinterpreted [11].

The advances in genome sequencing have provided high-throughput gene expression analysis and have contributed to the identification of a wide range of new gene products, such as the small noncoding RNAs, especially microRNAs (miRNAs). Discovered some decades ago, miRNAs are fundamental regulatory genes of eukaryotic genomes that regulate several biological functions

including hormonal control, immune responses, and adaptation to a variety of biotic and abiotic stresses [12–15]. Although the small size of miRNA (19–24 nucleotides) creates challenges for their detection, recent innovative adaptations to RT-qPCR have resulted in improvements in gene expression profiling. The development of stem–loop primers provided a tool for detecting and characterizing mature miRNAs by PCR [16,17]. In addition, stem–loop primers can be used to specifically convert the mature functional miRNA into its DNA complement and can potentially be used for multiplex reverse transcription (RT) reactions [16]. Due to the advances in the ability to characterize miRNA expression, research involving these small RNAs has increased in recent years. However, experiments involving miRNA expression are still a challenge in many research areas due to the lack of proper control genes for normalizing these transcripts. Davoren and co-workers, while searching for suitable reference genes for the normalization of microRNA expression, identified two miRNAs (*Let-7a* and *miR-16*) constantly expressed in human breast cancer tissues [17]. Peltier and Latham [18] also searched for appropriated reference genes for their miRNA RT-qPCR studies and found miRNA expression levels were the most constant RNA levels in their analysis. They identified three miRNAs that were highly consistent in expression across 13 healthy tissues and 5 tumor tissues in humans. These miRNAs were statistically superior to the most commonly used reference RNAs used in miRNA RT-qPCR cancer experiments [18]. Galiveti et al. [19] also reported the detection of five non-protein-coding RNAs as appropriate housekeeping genes in human tissues. Currently, there are no data reporting the expression stability pattern of microRNAs in plant tissues.

Recognizing the potential of miRNAs as a reference gene in RT-qPCR analysis, we evaluated these non-protein-coding genes for use in normalizing gene expression in the soybean. This study is the first in which miRNAs were evaluated for stability alongside other mRNA genes and also tested as potential reference genes for both miRNA and mRNA gene expression in plant tissues. For this study, we selected the soybean (*Glycine max* Merryll L.), which is the major legume crop worldwide and already has established mRNA housekeeping genes [7,20,21]. We compared the expression level of 10 soybean miRNAs (*miR156a*, *miR156b*, *miR167ab*, *miR167c*, *miR171a*, *miR171b*, *miR172ab*, *miR396a*, *miR1520c*, and *miR1520d*) with six common mRNA housekeeping genes (*ACT*, *CDPK*, *CYP2*, *ELF1B*, *F-BOX*, and *TUA*) and found that many of the miRNAs showed better expression stability than the protein-coding housekeeping genes, indicating that these genes can be used as optimal reference genes for both miRNAs and protein-coding genes in RT-qPCR analysis.

Material and methods

Plant materials and treatments

Drought assay

For drought treatment, we used the soybean (*G. max* Merryll L.) cultivars 'Embrapa 48' as a drought-tolerant standard [22] and 'BR 16' as a sensitive standard. The plants were grown in a greenhouse at Embrapa-soybean in Londrina, Brazil, using a hydroponic system compound for plastic containers (30 liters) and an aerated 6.6 pH-balanced nutrient solution [22,23]. Briefly, seeds were pregerminated on moist filter paper in the dark at $25 \pm 1\,^\circ$C and $65 \pm 5\%$ relative humidity. Then, the plantlets were placed in polystyrene supports in such a way that the roots of the seedlings were completely immersed in the solution. Each tray containing the seedlings was maintained in a greenhouse at $25 \pm 2\,^\circ$C and $60 \pm 5\%$ relative humidity under natural daylight (photosynthetic photon flux density (PPFD) = $1.5 \times 10^3\ \mu$mol m^{-2} s^{-1}, equivalent to

8.93×10^4 lx) and a 12 h day. After 15 days, seedlings with the first trifoliate leaf fully developed (V2 developmental stage) were submitted to different treatments. They were removed from the hydroponic solution and kept in a tray in the dark without nutrient solution or water for 0 min (T0 or control) and 125 min of stress (T125). The roots and leaves of both genotypes were analyzed as biological duplicate (T0) or triplicates (T125). They were immediately frozen in liquid nitrogen and stored at $-80\,^\circ$C until RNA extraction.

Asian soybean rust assay

Asian soybean rust (ASR) reaction was assessed after inoculation with a field population of *Phakopsora pachyrhizi* and performed in the greenhouse at Embrapa-soybean Londrina, Brazil. The soybean plants were grown in a pot-based system. The 'Embrapa 48' genotype was used as a susceptible host plant, which develops a susceptible lesion (TAN) after ASR infection [24], and 'PI561356' was used as the resistant host, which carries the ASR resistance mapped on linkage group G (personal communication). Urediniospores were collected by tapping infected leaves over a plastic tray and were then diluted in distilled water with 0.05% Tween 20 to a final concentration of 3×10^5 spores/mL. This suspension of spores was sprayed onto the plantlets at the V2 developmental stage. The same solution minus the spores was used for the mock inoculations. Following the ASR or mock inoculations, water-misted bags were placed over all plantlets for 1 day to aid the infection process and to prevent the cross-contamination of mock-infected plants. One trifoliolate leaf for each plant was collected at 192 h after inoculation for RNA extraction. Three biological replicates of each genotype were analyzed for both treatments.

Genotype and tissue assay

For the genotype assay, 'Embrapa 48', 'BR 16', and 'PI561356' were analyzed. Leaf samples under nonstress conditions from both the drought and the ARS assays at the same V2 developmental stage were used. The tissue assay was performed by analyzing the roots and leaves from the 'Embrapa 48' genotype under nonstress conditions and during the V2 developmental stage.

RNA extraction and cDNA synthesis

For each treatment, total RNA was isolated by extraction with Trizol, (Invitrogen) and the quality was evaluated by electrophoresis on a 1.0% agarose gel. The cDNA synthesis was carried out by multiplex technique [25,26] from approximately 2 μg of total RNA. Each reaction was primed with a pool of 0.5 μM 10 gene-specific stem–loop primers [16] (IDT) and 1 μM oligonucleotide dT24V (Invitrogen). Before transcription, RNA and primers were mixed with RNase-free water to a total volume of 10 μL and incubated at 70 °C for 5 min followed by ice-cooling. Then, 6 μL 5× RT-Buffer (Promega), 1 μL of 5 mM dNTP (Ludwig), and 1 μL MML-V RT Enzyme 200 U (Promega) were added for a final volume of 30 μL. The synthesis was performed at 42 °C for 30 min on a Veriti Thermal Cycler (Applied Biosystem), and inactivation of the enzyme was completed at 85 °C for 5 min. All cDNA samples were 50-fold diluted with RNase-free water before being used as a template in RT-qPCR analysis.

Selection of candidate housekeeping genes and primer design

The six protein-coding genes, *ELF1B*, *CYP2*, *ACT*, *TUA*, *F-BOX*, and *CDPK* (Table 1), were selected based on previous reports of housekeeping genes for the soybean [7,20,21]. The 10 miRNAs were chosen based on our previous gene expression studies in the soybean (data not shown). During the analyses of 43 miRNAs available on miRBase database at http://www.mirbase.org (release 12.0) for

Table 1
Primer sequences and amplicon characteristics for each of the 16 reference gene candidates.

Acronym	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	Amplicon size (bp)	Efficiency	R^2	G.max loci[a]	Function	Locus accession number[b]
ACT	CGGTGGTTCTATCTTGGCATC	GTCTTTCGCTTCAATAACCCTA	142	1.89	0.998	Gm18:61274000-61274248	Cytoskeletal structural protein	BW652479
CDPK	TAAAGAGCACCATGCCTATCC	TGGTTATGTGAGCAGATGCAA	97	1.97	0.986	Gm10:46251505-46251601	CDPK-related protein kinase	AW396185
CYP2	CGGGACCAGTGTGCTTCTTCA	CCCCTCCACTACAAAGGCTCG	154	1.80	0.996	Gm12:1802441-1802594	Protein folding	CF806591
ELF1B	GTTGAAAAGCCAGGGGACA	TCTTACCCCTTGAGCGTGG	118	1.86	0.996	Gm02:49033741-49033961	Translational elongation	EV279336
F-BOX	AGATAGGGAAATGGTGCAGGT	CTAATGGCAATTGCAGCTCTC	93	1.88	0.983	Gm12:3676720-3677719	F-Box protein family	CD397253
TUA	AGGTCGGAAACTCCTGCTGG	AAGGTGTTGAAGGCGTCGTG	159	1.95	0.982	Gm05:34705808-34705919	Cytoskeletal structural protein	CA801144
156a	TGACAGAAGAGAGTGAGCAC	GTGCAGGGTCCGAGGT	70	1.94	0.992	Gm17: 6149963-6150084	miRNA	MIMAT0001686
156b	TGACAGAAGAGAGAGAGCACA	″	71	2.02	0.992	Gm14: 990334-990453	miRNA	MIMAT0001692
167ab	TGAAGCTGCCAGCATGATCTA	″	71	2.00	0.992	Gm19: 41871231-41871349 Gm02: 14635614-14635734	miRNA	MIMAT0001679 MIMAT0001680
167c	TGAAGCTGCCAGCATGATCTG	″	71	1.92	0.998	Gm07: 39778512-39778886	miRNA	MIMAT0007355
171a	TGAGCCGTGCCAATATCACGA	″	71	2.00	0.997	Gm12: 8443106-8443207	miRNA	MIMAT0007358
171b	CGAGCCGAATCAATATCACTC	″	71	1.91	0.999	Gm04: 46988579-46988670	miRNA	MIMAT0007363
172ab	AGAATCTTGATGATGCTGCAT	″	71	2.00	0.979	Gm12: 6110704-6110862 Gm13: 40401673-40401825	miRNA	MIMAT0001682 MIMAT0001683
396a	TTCCACAGCTTTCTTGAACTG	″	71	1.90	0.999	Gm13: 26338134-26338273	miRNA	MIMAT0001687
1520c	TTCAATAAGAACGTGACACGTGA	″	73	1.97	0.992	Gm17: 37893185-37893324	miRNA	MIMAT0007395
1520d	ATCAGAACATGACACGTGACAA	″	72	2.00	0.996	Gm07: 119451-119561	miRNA	MIMAT0007379

[a] Position of CDS or pre-miRNA sequences.
[b] NCBI or miRBase database accession number.

RT-qPCR analysis in a drought-stress assay, we observed that miR156a, miR156b, miR167ab, miR167c, miR171a, miR171b, miR172-ab, miR396a, miR1520c, and miR1520d had uniform expression along a range of different conditions (unpublished data).

All primers of the 16 candidate reference genes are listed in Table 1. Primer sequences for the six mRNA housekeeping genes were chosen based on current literature [7,20]. The miRNA primers were designed based on the fully tested miRNA sequence (forward), and the reverse primer was the universal reverse primer for miRNA [16]. The stem–loop primers, used for miRNAs cDNA synthesis, were designed according to Chen et al. [16]. The stem–loop sequence consists of 44 conserved and 6 variable nucleotides that are specific to the 3′ end of the miRNA sequence (5′ GTCGTATCCAGTGCAGGGT CCGAGGTATTCGCACTGGATACGAC NNNNNN 3′).

Reverse transcription quantitative PCR

All quantitative PCR were performed in an ABI 7500 Real-Time PCR System (Applied Biosystem) using SYBR Green I (Invitrogen) to detect double-strand cDNA synthesis. Reactions were done in a volume of 24 μL containing 12 μL of diluted cDNA (1:50), 1× SYBR Green I (Invitrogen), 0.025 mM dNTP, 1× PCR Buffer, 3 mM MgCl₂, 0.25 U Platinum Taq DNA Polymerase (Invitrogen), and 200 nM of each reverse and forward primers. In all miRNA RT-qPCR the Universal Reverse primer (5′ GTGCAGGGTCCGAGGT 3′) was used. All samples were analyzed as technical triplicates with a no-template control also included. The conditions were set as the following: an initial polymerase activation step for 5 min at 94 °C, 40 cycles of 15 s at 94 °C for denaturation, 10 s at 60 °C for annealing, and 25 s at 72 °C for elongation. At the end of cycling protocol, a melting-curve analysis was included and fluorescence measured from 65 to 99 °C. Threshold and baselines were manually determined using the ABI 7500 Real-Time PCR System SDS Software v2.0.

Data analysis

Primer efficiency (E) and correlation coefficient (R^2) were calculated using SDS software (ABI 7500 Real-Time PCR System v. 2.0)

based on a standard curve generated using a twofold dilution series of one sample over five dilution points that were measured in duplicate. PCR amplification efficiencies were calculated for each candidate endogenous control with the formula $E = 10^{-1/\text{slope}}$, using the slope of the plot, C_t versus log input of cDNA. The stability of each candidate gene expression was analyzed using geNorm [27] and NormFinder [28] software. For these analyses, the threshold cycle (C_t) values were converted to quantities via the comparative C_t method. Both geNorm and NormFinder tools were used to determine the most stable reference genes. GeNorm software was also used to determine the optimal number of genes required for normalization. It starts calculating the pairwise variation (V) $V2/3$ between the NF2 (including the two most stable reference genes) and the normalization factor NF3 (including the three most stable reference genes). Then, it performs a stepwise calculation of the $Vn/n + 1$ between the NFn and the NF$n + 1$.

Reference gene validation

To determine how the use of different reference genes can affect the normalization of the expression data for a gene of interest, we calculated the significant difference in the mean expression of two target genes between drought-stressed and control subgroups. We selected the miR1513 (5′ CGCCTGAGAGAAAGCCATGACTTAC 3′) as a miRNA target gene and the CDPK as a protein-coding or mRNA target gene. The expressions of the two target genes were normalized using three different strategies: (1) with all candidate reference genes individually; (2) with the two most stable reference genes and also with the two most stable mRNA genes selected by NormFinder; (3) with the two and three most stable reference genes, and also with the two and three most stable mRNA reference genes according to geNorm software. To calculate the relative expression of the two target genes, we used the $2^{-\Delta\Delta Ct}$ method [29]. Student's t test was performed to compare pairwise differences in expression. The parameters of two-tailed distribution and two samples assuming unequal variances were established. The means were considered significantly different when $P < 0.05$.

Results and discussion

PCR efficiency and amplification specificity

The amplification efficiency for each primer pair was determined in an RT-qPCR assay using duplicates of a twofold dilution series from a cDNA template. Primer efficiency indicates the amplicon doubling rate of a specific primer pair during a PCR. When the efficiency is 100% or 2 (1 + efficiency = 2), it indicates that the cDNA target is duplicated at every PCR cycle during the exponential phase. The efficiency values of the 16 candidate reference genes ranged from 1.80 to 2.02, as listed in Table 1. About 75% of all primer pairs were in the range of 1.9–2.02, reflecting the high quality of the PCR. Amplification of the specific transcript was confirmed by the appearance of a single peak in the melting-curve analysis following completion of the amplification reaction (in Supplementary Data). Furthermore, the correlation coefficient (R^2) was calculated. The R^2 value indicates the quality of the fit of the standard curve to the plotted data points. All primer pairs showed an $R^2 > 0.98$ (Table 1), indicating a strong relationship between the detected C_t values and the corresponding relative amount of template in all the amplification reactions.

Expression stability of candidate reference genes

To investigate the suitability of 10 miRNAs as reference genes, we analyzed their expression stability along with 6 housekeeping genes commonly employed in soybean studies involving RT-qPCR analysis. The expression stability of these 16 genes was analyzed based on five different data sets: (1) all 31 samples including different tissues, cultivars, abiotic treatments, and biotic treatments; (2) different tissues; (3) different genotypes; (4) abiotic stress consisting of drought-stress; and (5) biotic stress relative to samples from plants infected with ARS.

GeNorm analysis

The program geNorm uses an algorithm to calculate an M value, which is a gene expression stability measurement defined as the mean pairwise variation for a given gene compared to the remaining tested genes. Genes with the lowest M value indicate the most stable expression, whereas the highest M values indicate the least stable expression [27]. When we analyzed all 31 samples (Fig. 1A), the average expression stability values (M) of *miR156b* and *miR1520d* were lower than those of the other miRNAs, indicating that they are the most stable candidate genes. Interestingly, we observed that, in this set of analysis, all miRNA genes were more stable than mRNA housekeeping genes. Tubulin-A (*TUA)* was the most unstable gene with an M value over 1.2. Similar results were found when the sample sets of different tissues (Fig. 1B), different genotypes (Fig. 1C), and different abiotic conditions (Fig. 1D) were analyzed. The mean expressions of *miR156b* and *miR1520d* were the most stable out of all tested reference genes. In most cases, the miRNAs were the most stable genes, except for *miR396a*, which was less stable than mRNA housekeeping genes in the different tissues and genotypes data sets. In the biotic stress data set, the lowest M value was the *miR156a* and *miR156b* pair (Fig. 1E). The *miR1520d* was the third most stable candidate. Although in the first four analyses, *TUA* (Fig. 1A–C) and *CDPK* (Fig. 1D) mRNA genes were the most variable, the *miR167c* and *miR171b* were the most unstable genes for the biotic stress panel (Fig. 1E). The results also indicate that, generally, the miRNAs are more stable than the mRNAs tested (Fig. 1), but not always. The *miR171b*, which was the fourth most stable gene in tissue, genotypic, and abiotic approaches, showed low stability in the biotic assay. Considering the performance of the mRNA housekeeping genes, we compared our results with previous studies. Jian et al. [7] reported that *ELF1B* and *CYP2* were the most stable genes among different soybean tissues. Although *CYP2* had the lowest stability value of all mRNA candidates, *ELF1B* was the second most variable gene in our tissue assay. Furthermore, nine of the miRNA genes were more stable than the *CYP2* gene after geNorm analysis (Fig. 1A). In another study, *F-BOX* was described as the most stable gene among different tissues. *F-BOX* and *CDPK* were the most stable genes in samples infected with ASR and control [20]. However, in our results, for tissue and ASR data sets, both mRNA genes exhibited higher M values compared with several miRNA candidates (Fig. 1A and F). However, in all five analyses, the miRNAs were among the most stable genes indicated by the geNorm.

Although the stability analysis based on M-value points determines the most stable genes, Vandesompele et al. [27] recommended that in order to measure expression levels accurately, a normalization method using multiple housekeeping genes instead of one is required. To account for this recommendation, geNorm also computes a normalization factor (NF) and estimates the optimal number of reference genes according to the pairwise variation (V). A pairwise variation of 0.15 is accepted as a cutoff, below which the inclusion of an additional control gene is not required for reliable normalization [27]. To determine the optimal number of internal control genes for normalization, we calculated the pairwise variation for all five data sets (Fig. 1F). The results indicate that, in all approaches, the combination of the two most stable genes would be sufficient for normalization purposes because the $V2/3$ value was lower than 0.15 (Fig. 1F). Based on this graph, we can conclude that the addition of a third, fourth, or fifth reference gene does not cause a remarkable decrease in the V value. Additionally, using a combination of the seven most stable genes, the V value actually increases slightly for the tissue, genotypic, abiotic, and biotic assays but still remains below the cutoff value (0.15). In our analyses, the combination of *miR156b* and *miR1520d* was appropriate for comparisons of all sample sets ($V < 0.11$), different tissues ($V < 0.08$), contrasting genotypes ($V < 0.09$), and also abiotic stress treatment ($V < 0.06$), whereas the *miR156a* and *miR156b* pair was suitable for biotic treatment ($V < 0.04$).

NormFinder analysis

Stability of expression was also analyzed using the program NormFinder. Its strategy is based on a mathematical model of gene expression that enables an estimation of the intra- and intergroup variation, which are then combined into a stability value [28]. Candidate control genes with the minimal intra- and intergroup variation will have the lowest stability value and will be ranked at the top. Using this program, we analyzed five sets in a similar manner as in the geNorm analysis: (1) all 31 sample sets with two different analyses; all samples with no subgroups and another where groups were divided into stressed and nonstressed subgroups; (2) tissue sets were subgrouped into root and leaf; (3) genotype sets were divided into three different cultivar subgroups; (4) abiotic sets had drought and control subgroups, and (5) biotic stress sets were grouped into ARS infection and noninfection subgroups (Table 2; Supplementary File 1). Interestingly, the *miR156a* was the most stable gene in four out of the five data sets, including both subgroups investigated in each sample data set, and it was the second most stable gene in the biotic stress group. *TUA* was among the three genes with the worst stability in all the data sets. *CDPK* was the least stable in the abiotic set, and *miR167c* and *miR171b*, such as *miR1520c*, were the most variable genes in the biotic and abiotic stress sets, respectively. Interestingly, these results are similar to those obtained by geNorm. Despite a visible variation in the rankings of the other genes generated by geNorm and NormFinder algorithms, in general, all analyzed sets showed a marked separation
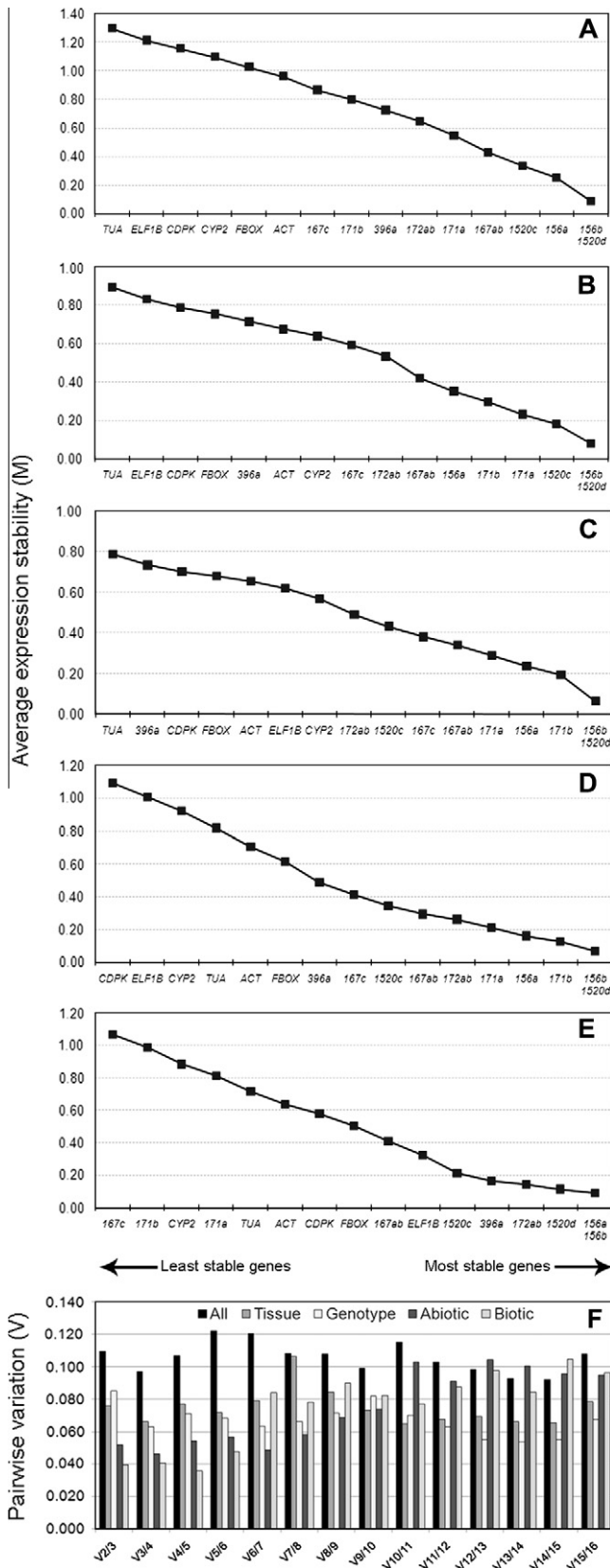
Fig. 1. Average expression stability values (*M*) of candidate reference genes after geNorm analysis: (A) all 31 samples including different tissues, cultivars, abiotic and biotic treatments; (B) different tissues; (C) different genotypes; (D) abiotic stress, consisting of drought stress and a control; and (E) biotic stress, relative to samples infected with ARS and noninfected (control); (F) determination of the number of housekeeping genes for normalization in each of the situations by geNorm analysis.

between the miRNA and the mRNA genes. As observed in the geNorm output, the miRNA genes were more stable compared to the mRNA candidates. Again, the best combination for the normalization of gene expression was two miRNA genes (Table 2).

Validation of putative reference genes

In order to demonstrate the suitability of the putative reference genes under investigation, we monitored the difference in the mean expression of two target genes from the drought-stress subgroup on normalization with different control genes. We decided to analyze two distinct classes of genes: one miRNA and one mRNA. Because there is currently no published data about miRNA expression in soybean, we decided to select the *miR1513* based on our previous RT-qPCR screening of the 43 soybean miRNAs from miRBase release 12.0. We observed that *miR1513* expression was clearly downregulated in soybean plants under drought conditions (unpublished data). The other target gene was *CDPK* (calcium-dependent protein kinases). This gene was one of the candidate reference genes; however, it was ranked as the least stable gene by geNorm and NormFinder analysis during drought-stress. Also, *CDPK* has been associated with drought-stress in plants. For example, studies with Arabidopsis [30,31] and rice [32] have demonstrated that this gene is overexpressed in drought situations.

Case 1: miRNA as target gene

In the first round of analyses, the target *miR1513* was individually normalized by each of the 16 candidate reference genes. A significant difference in gene expression under drought conditions compared to control samples was only determined with normalization based on *miR172ab*, *miR1520d*, *miR156b*, *miR171a*, and *miR156a* (Table 3). In the six cases where the target gene was normalized with an mRNA candidate reference gene, no significant difference was found between the compared situations (Table 3). In the next analysis, we normalized the target gene with *miR156a* and *miR171a*, the best pair for normalization comparison as indicated by NormFinder, and the differences between the subgroups, using miRNAs as reference genes, remained significant. We also selected the two most stable mRNA genes, *F-BOX* and *ACT*, following the NormFinder ranking and combined both with *miR1513* for normalization, and even so, no significant difference was observed between drought and control samples (Table 3).

Similar results were achieved when *miR1513* was normalized using either the two (*miR156b* and *miR1520d*) or three (*miR156b*, *miR1520d*, *miR171b*) best reference genes determined by geNorm. Although the top two reference genes elected by geNorm and NormFinder were not the same, both combinations were suitable for the normalization of the target gene. The mean-fold expression of *miR1513* under drought-stress conditions was significantly reduced to similar levels (0.4) with either *miR156a* and *miR171a* (NormFinder) or *miR156b* and *miR1520d* (geNorm) standardization (Table 3) compared to control samples.

We also selected the two most stable mRNA genes according to the ranking of NormFinder and geNorm for the drought-stress group. The *F-BOX* and *ACT* were determined to be the two best mRNA reference genes by both programs; however, when we normalized the target *miR1513* with these two reference genes, the difference in *miR1513* expression between the drought-stressed samples and the control samples was not significant. The same result was obtained when we added the *TUA* reference gene to these analyses. *TUA* was the third best mRNA reference gene by geNorm analysis. Again, no significant alteration in gene expression was determined when we normalized the miRNA target gene with three mRNA reference genes (Table 3).

In our investigations, we identified that the top ranked housekeeping genes by NormFinder (*miR156a* and *miR171a*) and geN-

Table 2
Ranking of candidate reference genes in order of their expression stability calculated by NormFinder.

Ranking order	All samples set		Tissues set (leaf::root)	Genotypes set (EM::BR::PI)	Abiotic stress set (drought::control)	Biotic stress set (ASR::control)
	Non-subgroups	(stress::control)				
1	156a	156a	156a	156a	156b	1520c
2	167ab	172ab	167ab	167c	1520d	156a
3	396a	1520d	171b	1520d	156a	167ab
4	172ab	156b	CYP2	156b	171b	172ab
5	156b	167ab	396a	171b	171a	156b
6	1520d	396a	156b	ELF1B	172ab	1520d
7	1520c	1520c	ACT	167ab	167c	396a
8	ACT	ACT	1520d	171a	167ab	ELF1B
9	F-BOX	F-BOX	F-BOX	F-BOX	396a	CYP2
10	171a	167c	1520c	CYP2	1520c	F-BOX
11	167c	171b	CDPK	CDPK	F-BOX	CDPK
12	171b	CYP2	172ab	172ab	ACT	ACT
13	CYP2	171a	171a	ACT	CYP2	171a
14	ELF1B	ELF1B	167c	1520c	ELF1B	TUA
15	CDPK	TUA	ELF1B	TUA	TUA	167c
16	TUA	CDPK	TUA	396a	CDPK	171b
Best combination of 2 genes		156a, 1520d	167ab, 171b	156a, 1520d	156a, 171a	172ab, 1520c

Note: The subgroups are indicated inside the parentheses; mRNA genes are in bold.

Table 3
Differential relative expression of the *miR1513* target gene when normalized individually with 16 different candidate housekeeping genes and when normalized by combining the most stable genes according to the NormFinder and geNorm analyses in abiotic stress treatment.

Normalizer	Mean fold change in gene expression				*P* value[b]
	Ctrl	SE[a]	Stress	SE[a]	
172ab	1.00	0.01	0.38	0.01	1.45E−04
1520d	1.00	0.00	0.39	0.02	0.002
156b	1.00	0.03	0.38	0.02	0.013
171a	1.00	0.05	0.45	0.09	0.026
156a	1.00	0.04	0.35	0.01	0.037
167c	1.00	0.07	0.33	0.03	0.068
171b	1.01	0.08	0.37	0.03	0.076
396a	1.01	0.10	0.37	0.18	0.096
1520c	1.01	0.08	0.37	0.01	0.102
CYP2	1.02	0.14	0.12	0.04	0.115
167ab	1.02	0.13	0.25	0.02	0.137
CDPK	1.24	0.52	0.04	0.01	0.349
TUA	1.03	0.18	2.82	1.94	0.530
ACT	1.05	0.23	0.84	0.63	0.817
ELF1B	1.03	0.17	1.14	0.86	0.926
F-BOX	1.04	0.21	0.99	0.74	0.958
NormFinder					
156a, 171a	1.00	0.03	0.40	0.05	1.8E−05
F-BOX, ACT	1.05	0.16	0.91	0.49	0.819
GeNorm					
156b, 1520d	1.0	0.02	0.4	0.02	5.9E−08
156b,1520d, 171b	1.0	0.03	0.4	0.01	1.3E−07
F-BOX, ACT	1.0	0.16	0.9	0.49	0.819
F-BOX, ACT, TUA	1.0	0.12	1.6	0.78	0.562

[a] *SE* (Standard Error).
[b] *P* values were calculated using the *t*-test, significance level $P < 0.05$.

orm (*miR156b*, *miR1520d*, *miR171b*) provided an accurate standardization for *miR1513* expression. When the target was normalized with either the two or three best reference genes based on our analysis, a difference in expression could be detected between stressed and control groups, in accordance with previously published data that has shown that *miR1513* is down-regulated under drought-stress conditions. However, when the miRNA target was normalized using *F-BOX* and *ACT* as well as *F-BOX*, *ACT* and *TUA*, which were the most stable among the

mRNA reference genes, no difference in expression was detected for *miR1513* between the drought-stressed and the control groups. If these mRNA housekeeping genes were used as reference genes for *miR1513*, without proper investigation, the results could be misinterpreted.

Case 2: mRNA as target gene

In the first set of analyses, the target gene *CDPK* was normalized individually with the remaining 15 housekeeping genes. We observed that with all individual normalizations, except with *ELF1B*, *TUA*, and *CYP2*, *CDPK* expression was significantly different between the drought and the control situation. Although the *CDPK* expression was very distinct between drought and control samples when normalized with *ELF1*, *TUA*, and *CYP2* genes, the values of the standard error (SE) were high, which explains why these marked differences are not statistically supported. In the second set, the *CDPK* expression was normalized with the two most stable genes (*miR156a* and *miR171a*) following the Norm-Finder ranking (Table 4). The expression of *CDPK* was also investigated using the two most stable mRNA genes identified in our study (*F-BOX* and *ACT*). In analyses, using miRNAs or mRNA as reference genes, the differences in target expression were significant between the stressed and the control groups. The third set of analyses was done based on the most stable genes selected by geNorm. *CDPK* expression was significantly different between drought and control when the target was normalized with the two (*miR156b* and *miR1520d*) and the three (*miR156b*, *miR1520d*, and *miR171b*) most stable genes.

Similar result was obtained when the two (*F-BOX* and *ACT*) and the three (*F-BOX*, *ACT*, and *TUA*) top-ranked mRNA genes were used for normalization (Table 4). These results show that normalizing with one protein-coding gene along with miRNAs or mRNA genes produces consistent results. It has been reported that the *CDPK* gene has increased expression during drought-stress [30–32]. In our research, we observed that this gene showed a fivefold increase in expression in plants under drought conditions even when expression was normalized using the two best miRNA or the two best mRNA housekeeping genes. From these findings, we conclude that miRNAs can be used as optimal reference genes not only for other miRNAs but also for protein-coding genes.

Table 4
Differential relative expression of the *CDPK* target gene when normalized individually with 15 different candidates housekeeping genes and when normalized by combining the most stable genes according to the NormFinder and geNorm analyses in abiotic stress treatment.

Normalizer	Mean fold change in gene expression				P value[b]
	Ctrl	SE[a]	Stress	SE[a]	
1520d	1.03	0.23	5.62	0.27	0.001
F-BOX	1.00	0.10	4.57	0.22	0.001
156b	1.03	0.26	6.14	0.46	0.003
171b	1.03	0.25	6.00	0.45	0.003
171a	1.03	0.23	5.95	0.48	0.004
156a	1.04	0.28	5.10	0.42	0.004
ACT	1.00	0.06	5.62	0.29	0.004
167c	1.03	0.23	3.84	0.23	0.005
1520c	1.00	0.06	5.41	0.43	0.008
396a	1.03	0.26	4.39	0.16	0.012
167ab	1.08	0.41	4.32	0.34	0.019
172ab	1.03	0.24	5.04	0.70	0.021
ELF1B	1.00	0.01	18.79	5.31	0.079
TUA	1.00	0.04	16.74	6.43	0.134
CYP2	1.00	0.01	10.08	6.22	0.282
NormFinder					
156a, 171a	1.03	0.13	5.52	0.31	9.4E−06
F-BOX, act	1.00	0.05	5.09	0.28	2.8E−05
GeNorm					
156b, 1520d	1.03	0.12	5.88	0.24	5.4E−07
156b,1520d, 171b	1.03	0.10	5.92	0.20	2.5E−10
F-BOX, ACT	1.00	0.05	5.09	0.28	2.8E−05
F-BOX, ACT, TUA	1.00	0.03	8.97	2.54	0.018

[a] SE (Standard Error).
[b] *P* value were calculated using the *f*-test, significance level $P < 0.05$).

Final considerations

In this study, we observed a marked difference between the expression stability of miRNA and mRNA candidate housekeeping genes. In general, miRNAs were the most stable genes across the five different sets analyzed by geNorm (Fig. 1A–E) and NormFinder (Table 2). Although the optimal combinations of genes selected by geNorm and NormFinder were not the same, both analyses did select the same class of genes as the most stably expressed genes. These differences in the ranking were previously reported [28] and thus affirm that the discrepancies observed in the NormFinder versus the geNorm results were caused by the differences between these two approaches. In conclusion, we provide evidence that miRNAs can have better expression stability than protein-coding genes. In addition, we demonstrate that microRNAs are optimal reference genes not only for other miRNAs but also for protein-coding genes in RT-qPCR analysis.

Ideally, a reference gene for quantitative gene expression studies should not be influenced by the experimental conditions, type of tissues, or developmental stages. In our work, we have found that miRNAs genes are, in general, more stable than the protein-coding genes. Analyses of the best reference gene among different classes of RNAs (mRNA, snRNA, and miRNA) were previously done using different human tissues [19] and comparing normal and malignant breast tissues [17]. As in our results, miRNAs have been shown to have a more stable expression than the other classes of RNAs. The reason why miRNAs expression is more stable than protein-coding genes remains unanswered.

Despite the use of miRNAs as reference genes in our and other studies, it must be considered that not all miRNAs have a constitutive expression among tissues and under stressing conditions. Indeed, several of them are modulated by environmental and hormonal conditions and are implicated in cellular and tissue differentiation. At the same time, if we consider a miRNA showing a biological function during a very narrow developmental stage

[33], its expression may remain stable in different organs and under several stressing conditions, presenting considerable variation just in a specific situation associated to its biological role.

It is important to stress that the use of miRNAs as reference genes in RT-qPCR-derived expression analysis of other miRNAs or even protein-coding genes does not impose technical difficulties or costs, since miRNA-specific primers can be mixed with the standard poli-T primer in the reverse transcription reaction.

Analyses in other plant and animal systems need to be undertaken to demonstrate the universality of the present results. Nonetheless, the evolutionary conservation of diverse miRNA families among distinct plant taxa suggests that the genes that encode miRNAs can be used as reference genes in place of the traditionally used protein-coding genes.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2010.07.020.

References

[1] S.A. Bustin, Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems, J. Mol. Endocrinol. 29 (2002) 23–39.
[2] A.M. Brunner, I.A. Yakovlev, S.H. Strauss, Validating internal controls for quantitative plant gene expression studies, BMC Plant Biol. 4 (2004) 14.
[3] O. Thellin, W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, E. Heinen, Housekeeping genes as internal standards: use and limits, J. Biotechnol. 75 (1999) 291–295.
[4] T.D. Schmittgen, B.A. Zakrajsek, Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR, J. Biochem. Biophys. Methods 46 (2000) 69–81.
[5] Z. Tong, Z. Gao, F. Wang, J. Zhou, Z. Zhang, Selection of reliable reference genes for gene expression studies in peach using real-time PCR, BMC Mol. Biol. 10 (2009) 71.
[6] T. Remans, K. Smeets, K. Opdenakker, D. Mathijsen, J. Vangronsveld, A. Cuypers, Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations, Planta 227 (2008) 1343–1349.
[7] B. Jian, B. Liu, Y. Bi, W. Hou, C. Wu, T. Han, Validation of internal control for gene expression study in soybean by quantitative real-time PCR, BMC Mol. Biol. 9 (2008) 59.
[8] N. Nicot, J.F. Hausman, L. Hoffmann, D. Evers, Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress, J. Exp. Bot. 56 (2005) 2907–2914.
[9] R.A. Volkov, I.I. Panchuk, F. Schoffl, Heat-stress-dependency and developmental modulation of gene expression: the potential of housekeeping genes as internal standards in mRNA expression profiling using real-time RT-PCR, J. Exp. Bot. 54 (2003) 2343–2349.
[10] T. Czechowski, M. Stitt, T. Altmann, M.K. Udvardi, W.R. Scheible, Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis, Plant Physiol. 139 (2005) 5–17.
[11] L. Gutierrez, M. Mauriat, J. Pelloux, C. Bellini, O. Van Wuytswinkel, Towards a systematic validation of references in real-time RT-PCR, Plant Cell 20 (2008) 1734–1735.
[12] I. Pedersen, M. David, MicroRNAs in the immune response, Cytokine 43 (2008) 391–394.
[13] R. Sunkar, V. Chinnusamy, J. Zhu, J.K. Zhu, Small RNAs as big players in plant abiotic stress responses and nutrient deprivation, Trends Plant Sci. 12 (2007) 301–309.
[14] O. Voinnet, Origin, biogenesis, and activity of plant microRNAs, Cell 136 (2009) 669–687.
[15] O. Voinnet, Post-transcriptional RNA silencing in plant-microbe interactions: a touch of robustness and versatility, Curr. Opin. Plant Biol. 11 (2008) 464–470.
[16] C. Chen, D.A. Ridzon, A.J. Broomer, Z. Zhou, D.H. Lee, J.T. Nguyen, M. Barbisin, N.L. Xu, V.R. Mahuvakar, M.R. Andersen, K.Q. Lao, K.J. Livak, K.J. Guegler, Real-time quantification of microRNAs by stem-loop RT-PCR, Nucleic Acids Res. 33 (2005) e179.

[17] P.A. Davoren, R.E. McNeill, A.J. Lowery, M.J. Kerin, N. Miller, Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer, BMC Mol. Biol. 9 (2008) 76.

[18] H.J. Peltier, G.J. Latham, Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues, RNA 14 (2008) 844–852.

[19] C.R. Galiveti, T.S. Rozhdestvensky, J. Brosius, H. Lehrach, Z. Konthur, Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR, RNA 16 (2010) 450–461.

[20] M. Libault, S. Thibivilliers, D.D. Bilgin, O. Radwan, M. Benitez, S.J. Clough, G. Stacey, Identification of four soybean reference genes for gene expression normalization, Plant Genome 1 (2008) 11.

[21] R. Hu, C. Fan, H. Li, Q. Zhang, Y.F. Fu, Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR, BMC Mol. Biol. 10 (2009) 93.

[22] P.K. Martins, B.Q. Jordão, N. Yamanaka, J.R.B. Farias, M.A. Beneventi, E. Binneck, R. Fuganti, R. Stolf, A.L. Nepomuceno, Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems, Genet. Mol. Biol. 31 (2008) 512–521.

[23] E.J. Hewitt, Nutrition of plants in culture media, in: F.C. Steward (Ed.), Inorganic Nutrition of Plants, Academic Press, London, 1963, pp. 97–360.

[24] M. van de Mortel, J.C. Recknor, M.A. Graham, D. Nettleton, J.D. Dittman, R.T. Nelson, C.V. Godoy, R.V. Abdelnoor, A.M. Almeida, T.J. Baum, S.A. Whitham, Distinct biphasic mRNA changes in response to Asian soybean rust infection, Mol. Plant Microbe Interact. 20 (2007) 887–899.

[25] K. Lao, N.L. Xu, V. Yeung, C. Chen, K.J. Livak, N.A. Straus, Multiplexing RT-PCR for the detection of multiple miRNA species in small samples, Biochem. Biophys. Res. Commun. 343 (2006) 85–89.

[26] F. Tang, P. Hajkova, S.C. Barton, K. Lao, M.A. Surani, MicroRNA expression profiling of single whole embryonic stem cells, Nucleic Acids Res. 34 (2006) e9.

[27] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002) RESEARCH0034.

[28] C.L. Andersen, J.L. Jensen, T.F. Orntoft, Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets, Cancer Res. 64 (2004) 5245–5250.

[29] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method, Methods 25 (2001) 402–408.

[30] T. Urao, T. Katagiri, T. Mizoguchi, K. Yamaguchi-Shinozaki, N. Hayashida, K. Shinozaki, Two genes that encode Ca(2+)-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis thaliana*, Mol. Gen. Genet. 244 (1994) 331–340.

[31] A.A. Ludwig, T. Romeis, J.D. Jones, CDPK-mediated signalling pathways: specificity and cross-talk, J. Exp. Bot. 55 (2004) 181–188.

[32] Y. Saijo, S. Hata, J. Kyozuka, K. Shimamoto, K. Izui, Over-expression of a single Ca^{2+}-dependent protein kinase confers both cold and salt/drought tolerance on rice plants, Plant J. 23 (2000) 319–327.

[33] J. Wang, B. Czech, D. Weigel, miR156-regulated SPL transcription factors define and endogenous flowering pathway in *Arabidopsis thaliana*, Cell 138 (2009) 738–749.