Electronic Journal of Biotechnology 75 (2025) 49–56

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

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

Research article

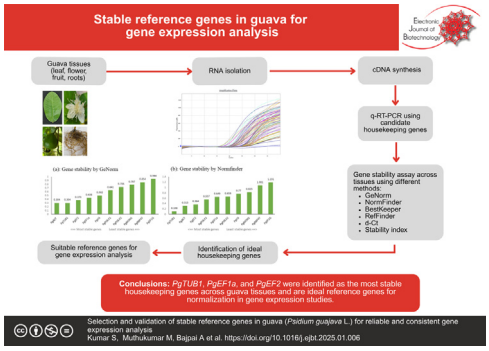
Selection and validation of stable reference genes in guava (*Psidium guajava* L.) for reliable and consistent gene expression analysis ☆

Sandeep Kumar [a,b], Manoharan Muthukumar [a,*], Anju Bajpai [a], Amar Kant Kushwaha [a], Israr Ahmad [a], Yashi Bajpai [a,b], Anshuman Singh [a], Thukkaram Damodaran [a], Mala Trivedi [b,*]

[a] *ICAR-Central Institute for Subtropical Horticulture, Rehmankhera, P.O, Kakori, Uttar Pradesh, India*
[b] *Amity University Uttar Pradesh, Lucknow Campus, Malhour, Lucknow, Nijampur Malhaur, Uttar Pradesh, India*

GRAPHICAL ABSTRACT

Selection and validation of stable reference genes in guava (*Psidium guajava* L.) for reliable and consistent gene expression analysis



ARTICLE INFO

Article history:
Received 18 September 2024
Accepted 31 January 2025
Available online 4 April 2025

Keywords:
Gene expression analysis
Guava
Housekeeping genes
Normalization
Psidium guajava L.
qRT-PCR
Reference genes
Selection
Validation

ABSTRACT

Background: Ideal stably expressing housekeeping gene to be used as internal control needs to be optimized for efficient gene expression analysis for accurate mRNA quantitation. In pursuit of identifying suitable reference genes for gene expression analysis in guava, a systematic study was conducted using different tissues of guava cv. Allahabad Safeda examining 10 housekeeping genes such as Actin (*PgACT*), Elongation factor 1G (*PgEF1G*), Elongation factor 2 (*PgEF2*), Tubulin (*PgTUB1*), Elongation factor 1 α (*PgEF1a*), Monensin sensitivity1 (*PgMON1*), Histone H3 (*PgH3*), RNA-binding protein 47 (*PgRBP47*), Polyubiquitin_X1 (*PgPOLX1*), and Polyubiquitin_X2 (*PgPOLX2*).
Results: qRT-PCR analysis showed amplification efficiencies ranging from 74.4% to 124.9%, with correlation coefficients exceeding 0.98. The stability of these genes' expression was evaluated using six methods: GeNorm, NormFinder, BestKeeper, RefFinder, comparative delta-Ct, and Stability index in which different methods identified *PgRBP47* as least stable and indicated its unsuitability as a reference gene but showed variations in the ranking of the genes for gene stability.
Conclusions: Comparison of all the methods and accounting for the top 3 ranks of gene stability, three genes i.e., *PgTUB1*, *PgEF1a*, and *PgEF2* were identified as more stable housekeeping genes across different

☆ Audio abstract available in Supplementary material.
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso
* Corresponding authors.
E-mail addresses: sandeep17apr@gmail.com (S. Kumar), muthukumarbt@gmail.com (M. Muthukumar), anju.bajpai@gmail.com (A. Bajpai), amar.kant20422@gmail.com (A.K. Kushwaha), israr15ahmad@gmail.com (I. Ahmad), yashi.bvm@gmail.com (Y. Bajpai), anshumaniari@gmail.com (A. Singh), damhort73@gmail.com (T. Damodaran), mtrivedi@lko.amity.edu (M. Trivedi).

https://doi.org/10.1016/j.ejbt.2025.01.006

tissues of guava and could be considered as ideal reference genes for normalization in gene expression studies of guava.

How to cite: Kumar S, Muthukumar M, Bajpai A, et al. Selection and validation of stable reference genes in guava (*Psidium guajava* L.) for reliable and consistent gene expression analysis. Electron J Biotechnol 2025;75. https://doi.org/10.1016/j.ejbt.2025.01.006.

1. Introduction

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) is one of the extensively used molecular techniques for quantifying gene expression levels because of its high sensitivity, specificity, accuracy, and capability for high-throughput analysis. Reliable and reproducible gene expression profiling results depend on the quantity and quality of the starting material (RNA), sampling procedures, enzyme efficiencies, presence of inhibitors, and basal transcription rate of the genes [1]. An important factor that controls the variations in the transcript activity is using one or two reference genes for data normalization to minimize technical variability among samples and accurately measure target gene expression levels [2,3]. Absolute and accurate quantitation of the reference gene among the constitutively expressing housekeeping genes is essential. Furthermore, the accuracy of qRT-PCR results depends on multiple factors, such as the quality of the complementary DNA (cDNA), the efficiency of the polymerase enzyme, the effectiveness of PCR amplification, and the stability of reference genes across different experimental conditions (treatments or developmental conditions/ stages), time (temporal) and tissues (spatial) [4,5,6].

The normalization process accounts for differences in the magnitude of gene expression in each sample, ensuring that results are comparable across different samples. In different crop plants, different scientific groups have identified reference genes for their data normalization and relative quantification of gene expression profiling. Zhang et al. [7] have demonstrated Ribosomal protein L2 (L2) and Tubulin (*TUB*) genes with the highest stability for gene expression studies in banana Guangfen No. 1 (ABB) while a combination of *ACT1* and *TUB* optimal for Cavendish cv. Brazilian (AAA). Zhu et al. [8] introduced *MaSPS1* as a novel banana reference gene. Nocum et al. [9] identified the *L2* gene as an ideal reference gene for characterizing gene expression in Banana bunchy top virus-resistant and susceptible varieties. In apples, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), ribosomal protein L2 (*RPL2*), and protein phosphatase 2A (*PP2A*) were reported as most suitable reference genes across various tissue types and under different biotic stress conditions [10]. Zhu et al. [11] reported that *WD40*, *ACT*, and *GAPDH* as the most stable reference genes in apples for expression analysis of peel samples while *EF-1α*, *CKL*, and *WD40* were for flesh samples. However, Song et al. [12] identified seven novel reference genes (*CYSP*, *NDUFS8*, *YLS8*, *EIF5A2*, *Gluc*, *GDT1*, and EF-Hand) in apples with stable expression across different tissue types, developmental stages, and biotic stress conditions. In grapevine, different combinations of *VvEF1-γ, PPR2, RRM1, EF1-α*, and Actin were reported as stable reference genes for expression analysis in different tissue types [13,14]. Similarly, in mango, the most commonly used reference genes are Actin, Tubulin, and Elongation Factor 1 [15,16,17,18]. However, there are reports on *GAPDH*, 25S ribosomal RNA, and *TUB-B* as stable reference genes under various experimental conditions [19,20,21].

From the above facts, it is evident that the most commonly reported reference genes for expression analysis include genes encoding Actin (*ACT*), α and β-tubulin (*TUA* and *TUB*), 18S ribosomal RNA (18S rRNA) and 28S rRNA, Histone B (*H2B*), elongation factor-1-alpha (*EF1-α*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), suppressor of K+ transport growth defect (*SKD1*), Thioredoxin-like protein (*YLS8*), Actin-depolymerizing factor 2 (*ADF2*), and Cytochrome B5 (*CYB5*) [4,22,23,24]. Notably, no universally stable reference gene consistently exhibits stable expression across all time points, tissues, and experimental conditions that can be invariably used in all crop plants. Moreover, the gene expression stability can vary due to species-specific or even tissue-specific traits in different plant species [25]. For instance, in tomatoes, the expression of some housekeeping genes was not stable, even at different leaf stages and bud sizes [26]. This poses issues of overestimating or underestimating the relative quantitation of gene expression levels if such genes are used as reference genes. The stability of reference genes is crucial in gene expression studies, as inconsistent expression can lead to inaccurate normalization and skewed qRT-PCR results [5,6]. Therefore, it is highly essential to identify the ideal reference gene specific to a crop with good stability that produces consistent results in the expression analysis. With the view of accomplishing this objective, the present study was undertaken to utilize a few housekeeping genes of guava for the selection of the best stable reference gene.

2. Material and methods

2.1. Plant materials

Guava (*Psidium guajava* L.) cultivar (cv.) Allahabad Safeda being popular cv. grown in many parts of the country was used in this study. Different tissues were collected from cv. Allahabad Safeda at the ICAR-Central Institute for Subtropical Horticulture (CISH), Lucknow, (latitude 26.90°N, longitude 80.76°E), Uttar Pradesh, India. Tissues representing different developmental stages such as juvenile, young, and mature leaves, flower buds, whole flowers, fruit pulp, fruit peel, whole fruits, and roots were harvested, immediately frozen in liquid nitrogen, and stored at −80°C until further processing.

2.2. RNA Extraction, quality control, quantitation, and cDNA synthesis

Total RNA was extracted from all the collected tissues using the TRIzol reagent (Invitrogen Bioservices India Pvt Ltd, Gurgaon, India) following the manufacturer's protocol, with minor modifications to enhance RNA yield and purity. RNA quality and quantity were assessed using a nanodrop spectrophotometer (Thermo Fisher Scientific, USA). RNA samples with an A260/A280 ratio between 1.8 and 2.0, and sharp ribosomal RNA bands (18S and 28S) on 1% agarose gel, were used for complementary DNA (cDNA) synthesis. The cDNA synthesis was carried out using the maxima first strand cDNA kit (ThermoFisher Scientific Pvt Ltd, India), with 2 μg of total RNA as the template, 5X Reaction Mix (4 μL), Maxima Enzyme Mix (2 μL), template RNA 2 μg, and rest of the volume made up with nuclease-free water to the reaction volume set up at 20 μL; as per instruction manual. All the components were

gently mixed, centrifuged, and reaction programmed at 1 cycle each of 10 min at 25°C, 30 min at 50°C, termination by heating at 85°C for 5 min and holding at 4°C for 20 min in the Prima96 PCR machine (HiMedia Laboratories India Pvt. Ltd., Mumbai). The cDNA concentration was checked using the QIAexpert nano-spectrophotometer (Qiagen India Pvt Ltd., New Delhi). The cDNA was diluted five-fold with nuclease-free water to achieve a concentration of 100 ng/µl before use in quantitative real–time polymerase chain reaction (qRT-PCR) assays.

2.3. Mining of housekeeping genes from the guava genome and designing qRT-PCR primers

Ten housekeeping genes were selected randomly based on literature, and the genomic sequences were mined from publicly available guava genome data (GCA_016432845.1). These candidate genes include actin (*PgACT*), elongation factor 1G (*PgEF1G*), elongation factor 2 (*PgEF2*), tubulin (*PgTUB1*), elongation factor 1α (*PgEF1α*), monensin sensitivity 1 (*PgMON1*), histone H3 (PgH3), RNA- binding protein 47 (*PgRBP47*), and two polyubiquitin genes (*PgPOLX1* and *PgPOLX2*). Primers for qRT-PCR were designed using Primer3Plus software [27], with parameters set with product size (90–250 bp), melting temperature (60–65°C), higher GC content, and minimal secondary structure formation. The designed primers (Table 1) were synthesized and obtained from Xcelris Laboratories India Pvt Ltd., Gujarat, India which was used for further experiments.

2.4. Gene expression analysis using qRT-PCR

The qRT-PCR assay was conducted using the Applied Biosystems™ 7500 system with the reaction set up involving 10 ng of cDNA, 5 picomoles each of forward and reverse primers, and then amplified using a SYBR® Green master mix. The reaction was performed by following a two-step program: Pre-PCR Read: 60°C for 30 s followed by holding stage at 95°C for 10 min for complete denaturation and enzyme activation and cycling stage at 95°C for 15 s (denaturation) followed by 60°C for 1 min (annealing and extension), repeated for data acquisition. To minimize biological and technical errors, each sample was evaluated in triplicate technical replicates and three biological replicates. Serially diluted cDNA samples (5^{-1}, 5^{-2}, 5^{-3}, 5^{-4}, and 5^{-5}) were used to construct standard curves, which were then used to calculate the efficiency of gene amplification (E) using the following formula;

$$E = \left(10^{-1/slope} - 1\right) * 100$$

2.5. Statistical analysis and selection of ideal reference gene for expression analysis in guava

The expression stability of ten candidate housekeeping genes in guava cv. Allahabad Safeda was assessed using four methods viz., GeNorm [28], NormFinder [29], BestKeeper [30], and Comparative delta-Ct [31]. Ct values were transformed and used in three methods GeNorm, Comparative delta-Ct, and NormFinder, while the raw Ct values were used in the Excel-based tool, BestKeeper. RefFinder [32], an online platform, integrates results from these four methods and provides a comprehensive ranking of candidate gene expression stability. For gene stability assay [33], the mean cycling time (Ct) values of replicate samples of different tissues of guava for all 10 housekeeping genes were obtained from the Applied Biosystems™ 7500 system. Data were processed through Microsoft Excel Analysis ToolPak and analyzed using SPSS software version V.1.6. The Ct value, which is inversely proportional to the amount of template nucleic acid, was used to determine the overall mean Ct, standard deviation (SD), and coefficient of variation (CV) for each gene. The tissue level variations in Ct value were also assessed using a box whiskers plot (SPSS V.1.6), providing an early measure of the stability of the reference gene. Gene stability was measured across different tissues representing different developmental stages using the stability index based on linear regression analysis. The stability index was calculated as the product of the CV and the slope of linear regression. Genes with a lower stability index were considered more efficient reference genes for data normalization in the gene expression studies.

3. Results

3.1. PCR optimization and assessment of primer efficiency

Ten housekeeping genes were amplified using cDNA from the leaf tissue of guava cv. Allahabad Safeda for initial standard PCR optimization. Agarose gel electrophoresis of the PCR products

Table 1
Description of list of qRT-PCR primers for candidate housekeeping genes.

Gene	Function (Protein)	Primer sequence (5′-3′)	Tm (°C)	Product size (bp)
PgACT	Actin	F: CTGAGCGTGAAATTGTGAGAG	61.1	310
		R: GGAGTTGTAAGTGGTCTCATGG	62.2	
PgEF1G	Elongation factor 1 gamma	F: CAAAAGAGCCCAAAGCAGAAG	62.0	300
		R: CAGAACCACAGTGAGTATCCC	61.7	
PgEF1a	Elongation factor 1 alpha	F: CCCTCTGGATTGACAACTGAAG	62.5	320
		R: TTCAGATCCTTTACAGCCACG	62.1	
PgEF2	Elongation Factor2	F: AAATTCTCCGTGTCTCCAGTC	61.8	300
		R: ACTTCACAATCTCAGCTCCAC	62.1	
PgTUB1	Tubulin	F: CAAGAACTCGTCCTACTTCGTG	62.2	306
		R: AGGTCGTTCATGTTGCTCTC	62.0	
PgMON1	Monensin sensitivity1	F: TCTCCGCCATTCAACACTC	61.8	312
		R: CGTTTTCCACATCTTTCACCC	61.6	
PgH3	Histone 3	F: CTCTCCGCATCTCCAATCTC	61.5	265
		R: CACACTCGTACCACTTGCTC	62.3	
PgRBP47	RNA-binding protein 47	F: TGCCTAATACTGAACAGCCG	61.9	300
		R: CATCGCCAAACCTCACAAATC	62.0	
PgPOLX1	Polyubiquitin_X1	F: GGTTCCCCATCCCAAAGATC	62.2	300
		R: TCACGCACCATGTACCAAAG	62.6	
PgPOLX2	Polyubiquitin_X2	F: AACTCTACATCTCGTGCTTCG	61.7	300
		R: ATCTTCCAGTTGCTTCCCG	61.9	

Note: Tm refers to melting temperature and the actual annealing temperature (Ta) was 2°C lesser than the Tm.

revealed that each primer amplified a single fragment (approximately 300 bp) for the ten selected reference genes (Fig. 1a) in concordance with the expected product size (Table 1). Subsequent qRT-PCR analysis in the same sample demonstrated that the amplification plot showed a clear undisturbed trend and similarly, the melt curve analysis showed a single peak for all the ten candidate genes, clearly indicating that there was no non-specific amplification of stutters (Fig. 1b, **Fig. S1**). The amplification efficiency of the 10 housekeeping genes was evaluated using specific primer pairs. All the primer pairs demonstrated high reliability, achieving correlation coefficients exceeding 0.98. This indicates strong linearity in their amplification performance, ensuring accurate and consistent results in quantitative analysis (standard curve **Fig. S2**). The amplification efficiency (E) ranged between 74.353% and 124.9% (Table 2). These findings confirm the primers' efficiency in amplification of the target genes.

3.2. Expression levels of candidate housekeeping genes across different tissues

The qRT-PCR analyses using cDNA from different tissues were carried out to assess the consistency of the expression of target genes to determine the optimum basal level of transcription, a crucial factor in selecting the ideal reference gene. The expression levels were quantified in terms of threshold cycle values (Cq, or Ct) which ranged from 19.32 to 35.19. *PgACT* exhibited the highest expression levels, while *PgPOLX1* had the lowest. The line graph of Ct values across different tissues (Fig. 2a) revealed that *PgTUB* and *PgEF1G* had reasonably steady Ct values across all tissue types, indicating consistent expression commonly desired for reference genes in qRT-PCR. While PgRBP47 and *PgPOLX1* genes exhibited more varied results across tissues, implying that their expression is tissue-specific in nature. The box plot of Ct values of genes across different tissues (Fig. 2b) revealed that genes *PgTUB1*, *PgACT*, and *PgEF1G* have narrow interquartile ranges (IQRs) and low variability, indicating their eligibility as reference genes. Genes with a larger IQR and higher median Ct values, such as *PgPOLX2*, *PgRBP47*, and *PgPOLX1*, may be unsuitable as reference genes for normalization due to fluctuating expression. However, these results give only a basic idea about the basal transcription levels (expression patterns) of the housekeeping genes and could not be directly used to select an ideal reference gene before assessing the gene expression stability across tissues.

3.3. Assessment of stability in gene expression across tissues for selection of ideal reference gene

Four different methods were used for expression stability analysis which revealed variations in the stability pattern of the housekeeping genes. In GeNorm method employs the estimation of the M value as a key metric with an inverse relationship to the stability of the reference genes. In the present study, *PgACT* and *PgTUB1* exhibited the lowest M value of 0.304, indicating the highest stability. *PgEF2* followed closely with an M value of 0.373. In contrast, *PgEF1G* was identified as the least stable gene, with an M value of 0.944 (Fig. 3a). NormFinder evaluates the stability of each candidate gene based on stability value (SV) which is indirectly related to consistency of expression across samples. In this present study, *PgTUB1* showed the lowest SV (0.108), showing the highest stability. Conversely, *PgEF1G* had the highest SV (1.191) proving it to be the least stable (Fig. 3b). The BestKeeper method assesses gene expression stability by combining the coefficient of variation (CV) and standard deviation (SD) into a single stability value (S), where a lower S indicates greater stability. The results from BestKeeper differed from those of GeNorm and NormFinder. BestKeeper identified *PgEF2* (S = 0.237) as the most stable gene, while *PgRBP47* was the least stable. *PgTUB1* and *PgACT* were ranked third and fifth in stability, respectively (Fig. 3c). The Delta Ct method evaluates gene stability using the standard deviation (SD) of Ct values, with higher SDs indicating greater variability and less stability. The results

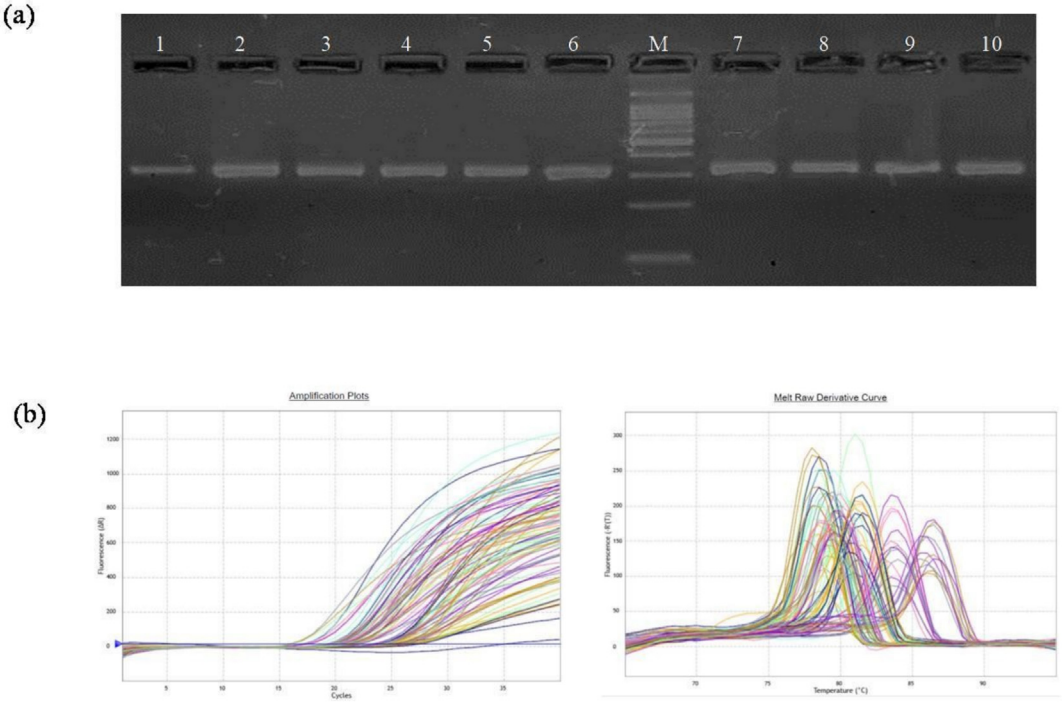


Fig. 1. Amplification of ten housekeeping genes in leaf tissue of guava (*Psidium guajava* L.) cv Allahabad Safeda. (a) Agarose gel electrophoresis profile of the amplicons of ten house-keeping genes. 1: *PgACT*, 2: *PgEF1G*, 3: *PgEF1a*, 4: *PgEF2*, 5: *PgTUB1*, 6: *PgMON1*, 7: *PgEF1*, 8: *PgRBP47*, 9: *PgPOLX1*, 10: *PgPOLX2*. M: 100 bp DNA ladder. (b) Amplification plot and melt curve plots of qRT-PCR analysis of the target genes.

Table 2
Assessment of amplification efficiency and power of reference genes for reliable gene expression analysis. Efficiency % defines the amplification efficiency, r^2: Correlation coefficient, Mean Ct: average of Ct value among all tested tissue samples, SD: Standard Deviation among all the tissue samples tested. Slope: Slope of regression of gene means. CV: Coefficient of variation = ((SD/Meant Ct)*100), and Stability index = CV*slope.

Reference gene	Efficiency %	r^2	Mean Ct	SD	Slope	CV%	Stability index
PgACT	105.35	0.99	18.71	0.54	3.2	2.87	9.18
PgEF1G	80.47	0.98	25.23	0.75	3.942	2.97	11.71
PgEF1a	103.09	0.99	22.50	0.32	3.258	1.42	4.63
PgEF2	105.35	0.99	19.69	0.31	3.22	1.57	5.06
PgTUB1	124.9	0.98	28.33	0.55	2.481	1.95	4.84
PgMON1	83.29	0.99	30.79	1.36	3.8	4.44	16.87
PgH3	97.04	0.99	20.65	0.48	3.395	2.33	7.91
PgRBP47L	74.35	0.98	26.1	1.60	4.142	6.12	25.35
PgPOLX1	100.92	0.99	33.45	1.04	3.3	3.11	10.26
PgPOLX2	89.41	0.99	32.48	1.21	3.605	3.73	13.45

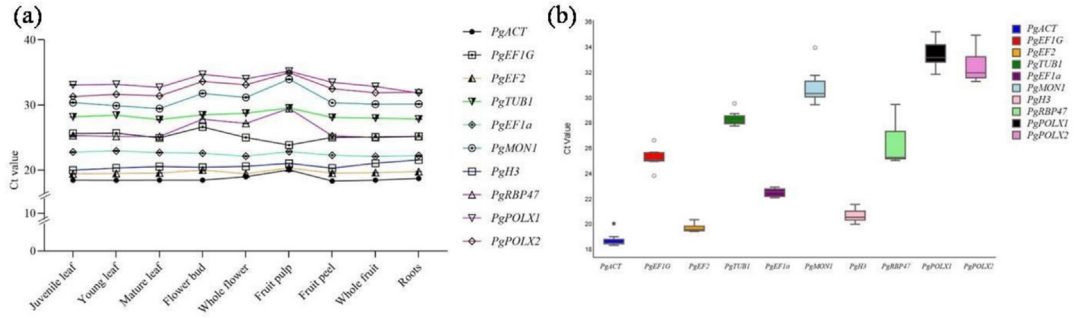


Fig. 2. Variations and uniformity in the gene amplification patterns among different tissues of guava (*Psidium guajava* L.) cv Allahabad Safeda. (a) Pattern of Ct values obtained in the different tissues for the 10 housekeeping genes represented in line graph. (b) Box and whiskers plots showing the tissue level variations and stability of Ct values of the candidate genes.

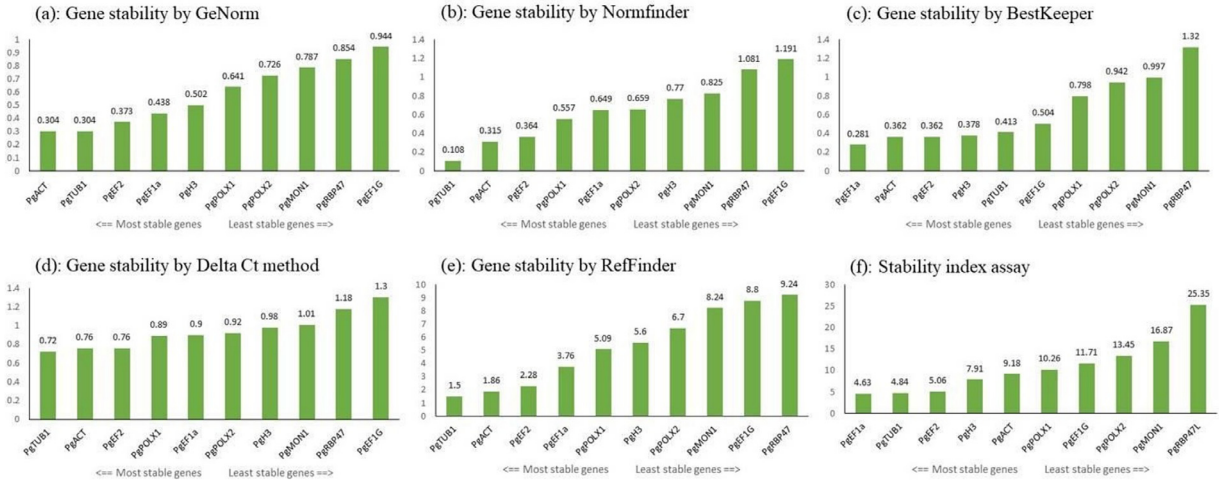


Fig. 3. Identification of stably expressed reference gene among the different housekeeping genes based on stability analysis using different methods. Histogram showing pattern of stability of gene expression assessed using different methods viz., (a) GeNorm, (b) Normfinder, (c) BestKeeper, (d) ΔCt, (e) RefFinder, and (f) stability index assay for the ten candidate genes.

from the Delta Ct method were consistent with those from GeNorm and NormFinder, with *PgTUB1* showing the lowest SD (0.72), indicating high stability, while *PgEF1G* had the highest SD (1.3), indicating lower stability (Fig. 3d).

RefFinder integrates the above four methods in which *PgTUB1* (stability score 1.5) and *PgACT* (stability score 1.86) were identified as the most stable reference genes. On the other hand, *PgRBP47*, with a stability score of 9.24, was found to be the least stable reference gene (Fig. 3e). The order of gene stability obtained in this assay could be ordered in the pattern as follows:

$$PgTUB1 > PgACT > PgEF2 > PgEF1a > PgPOLX1 > PgH3 > PgPOLX2 > PgMON1 > PgEF1G > PgRBP47.$$

Another method stability index (SI) incorporates the Ct values, and regression analysis to assess the stability of the candidate genes. In this study, the *PgEF1a* (SI: 4.63), *PgTUB1*(SI; 4.84), and

PgEF2 (SI: 5.06) were the most stable across all the different tissues. *PgRBP47* was found to be the least stable (Fig. 3f). The order of gene stability obtained in this assay is as follows:

$$PgEF1a > PgTUB1 > PgEF2 > PgH3 > PgACT > PgPOLX1 > PgEF1G > PgPOLX2 > PgMON1 > PgRBP47.$$

In comparison of all the six methods in gene stability and consistency, it is observed that *PgTUB1, PgEF1a,* and *PgEF2* were identified as the best reference genes occupying the first three ranks in different methods. Contrarily, *PgRBP47* could not be used for reference gene being least stable in all the methods evaluated for gene stability. *PgPOLX1* and *PgACT* could also be used as reference genes though their efficiency in qRT-PCR results would be moderate. Thus, it is clear from the results of the present study that *PgTUB1, PgEF1a,* and *PgEF2* are ideal reference genes for data normalization in gene expression studies of guava.

4. Discussion

Gene expression profiling and independent transcript-based reduced represented transcriptome analysis rely on the accuracy of relative quantification of expression levels, which depends on precise PCR conditions with optimum enzymes, initial RNA templates used in cDNA synthesis, cDNA concentration, inhibitors, and appropriate internal control, i.e., a stably, consistently and constitutively expressing housekeeping gene, often called as reference gene. A reliable internal control or reference gene should exhibit minimal fluctuations to the experimental conditions and external factors, in contrast to the potential significant variations in the gene being studied throughout the experiment [34].

The selection of an appropriate internal control is, therefore, crucial for accurately measuring gene expression levels. The most common housekeeping genes involved in fundamental cellular processes critical for plant cytoskeleton framework, metabolism, growth, and survival include Actin (*ACT*), tubulin (*TUB*), Polyubiquitin (*UBQ*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), translation elongation factor (*TEF*), ribosomal RNA (18S rRNA), eukaryotic initiation factor (*EIF4a*), cyclophilin (*CYP2*), protein phosphatase 2A (*PP2A*), and 25S ribosomal RNA (25S rRNA) [25]. Thus, these genes are expected to maintain stable expression levels across various conditions, such as different tissue types, developmental stages, and be less perturbed by environmental influences [7]. Unfortunately, all housekeeping genes are not universally stable and vary due to species-specific or even tissue-specific traits in different plant species [26,35].

The stability assay is also crucial because it assures that the reference gene used to normalize the data is dependable, resulting in more accurate and consistent results in the quantitation of mRNA expression. Various gene stability analysis-based algorithms use different principles for the estimation of gene stability across different tissues. Six methods of gene stability analysis viz., GeNorm, NormFinder, and BestKeeper, comparative delta cycle threshold (ΔCt) method, RefFinder, and stability index were examined to assess the stability of the 10 housekeeping genes in different tissues of guava.

PgRBP47 was detected to be least stable in all the methods evaluated indicating that this gene is unsuitable as a reference gene. *RBP47* is a heterogeneous nuclear ribonucleoprotein (hnRNP)-protein binding the poly(A) tail of mRNA and is associated with pre-mRNA maturation in plant cell nuclei [36]. This protein contains three RBD-type RNA-binding domains and a glutamine-rich N-terminus. It is highly conserved but diverse in its ability to interact with RNAs to regulate post-transcriptional events [37]. It was reported to be a key modulator in the stress granule formation and abiotic stress response in plants, especially with respect to heat, drought, and salinity stress [38]. From the above facts, it is evident that this housekeeping gene is expected to have variable gene expression patterns and is an inducible gene upon stress. The variation in the gene stability for PgRBP47 could be attributed to these reasons and thus, has proved to be less stable, indicating that this gene is not suitable to be used as a reference gene.

Several studies have indicated *ACT* gene encoding actin cytoskeleton protein as a housekeeping gene to quantify gene expression [4,15,17]. In the present study, we have identified *PgACT* with moderate gene stability, which could only provide moderate efficiency in the relative gene expression quantification.

Beta-tubulin and tubulin genes have emerged as robust choices due to their stable expression profiles across different plant tissues, developmental stages, and experimental conditions in different crop plants such as Arabidopsis [39], tomato [26], *Petunia hybrida* [40], cucumber [41], and *Oenanthe javanica* [42]. Tubulins are structural proteins that form the building blocks of microtubules, fundamental components of the plant cytoskeleton, and contribute to essential cellular processes such as cell division, intracellular transport, and cell shape maintenance. The stability of beta-tubulin and tubulin gene expression is attributed to their conserved sequences and crucial roles in cellular functions. As a result, they are widely used as reference genes in qRT-PCR for normalizing gene expression data. *PgTUB1* was identified to be the most stable in all the methods tested with the least M value = 0.304 (GeNorm), lowest SD = 0.72 (Delta Ct), least SV = 0.108 (NormFinder), rank 3 (RefFinder), and least SS = 1.5 (stability index assay). The GeNorm algorithm has the advantage that the expression intensity of the candidate genes minimally affects the stability [43]. RefFinder is an online platform that integrates GeNorm, NormFinder, BestKeeper, and the ΔCt method, providing a comprehensive and robust ranking system to help researchers select the most suitable reference genes for their experiments.

Eukaryotic elongation factors such as EF1α and EF2 or eEF2 are two important cellular translation machinery which is critical facilitators for the movement of the mRNA codons through the ribosomes during protein synthesis. EF-2 has five domains with an amino-terminal G, a twisted β-barrel, and three two-layer α-β sandwich domains [44]. During translation, EF2 mediates the translocation of the mRNA relative to the ribosome and allows codon shift from the aminoacyl (A) site to the peptidyl (P) site and finally the exit (E) site in a GTP-dependent reaction [45], while EF1α is responsible for the enzymatic delivery of aminoacyl tRNAs to the ribosome's acceptor A site [46]. In addition, eEF1A has also chaperone-like activity [47] and has also been reported to show antiviral response by interaction with a multifunctional protein Sgt1 [48] and is also a novel target for cancer drugs [49]. The abundant lysine methylations of eEF1A and the existence of multiple corresponding methyltransferases in various eukaryotes emphasize its physiological significance in vital functions [46]. Being constitutively expressed, they are considered to be the primary category of housekeeping genes which are involved in the protein synthesis of all kinds of genes. Their expression is expected to be stable across tissues and similar results have been obtained in our study. *PgEF2* has been found to be the most stable housekeeping gene and was identified in the top ranks in all 6 methods of stability assessment. For instance, BestKeeper identified *PgEF2* as the most stable with S = 0.237.

5. Conclusions

A stable reference gene is expected to show minimal variation across different tissues, thereby indicating that its expression is

consistent and reliable for normalization and quantification of relative mRNA/ transcript expression. Based on the present study, we report that three housekeeping genes viz., *PgTUB1*, *PgEF1a*, and *PgEF2* are potential candidates as reference genes for data normalization in gene expression studies of guava. The study also explains the power of different gene stability analysis tools in the selection of ideal reference genes.

CRediT authorship contribution statement

Sandeep Kumar: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Muthukumar Manoharan:** Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition, Conceptualization. **Anju Bajpai:** Writing – review & editing, Software, Resources. **Amar Kant Kushwaha:** Writing – review & editing, Software. **Israr Ahmad:** Writing – review & editing. **Yashi Bajpai:** Methodology. **Anshuman Singh:** Resources. **Thukkaram Damodaran:** Writing – review & editing, Visualization, Supervision. **Mala Trivedi:** Writing – review & editing, Supervision.

Financial support

This research was supported by Department of Science and Technology, New Delhi under DST-SERB in the form of Junior Research Fellowship to the first author and grant to the corresponding author (Dr. Muthukumar. M.) with grant no. DST-SERB EEQ/2018/000355.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are thankful to the DST-SERB for financial assistance under EMEQ project as Junior Research Fellow to the first author who has carried out this work under his Ph. D. research. We appreciate the Director, ICAR-Central Institute for Subtropical Horticulture (ICAR-CISH) for providing the necessary facilities for carrying out this research work. We also acknowledge the support from AMITY University Uttar Pradesh, Lucknow Campus, India.

Supplementary material

https://doi.org/10.1016/j.ejbt.2025.01.006.

Data availability

Data will be made available on request.

References

[1] Freitas FCP, Depintor TS, Agostini LT, et al. Evaluation of reference genes for gene expression analysis by real-time quantitative PCR (qPCR) in three stingless bee species (Hymenoptera: Apidae: Meliponini). Sci Rep 2019;9:17692. https://doi.org/10.1038/s41598-019-53544-0. PMid: 31776359.

[2] Garg R, Sahoo A, Tyagi A, et al. Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). Biochem Biophys Res Commun 2010;396(2):283–8. https://doi.org/10.1016/j.bbrc.2010.04.079. PMid: 20399753.

[3] Van-de-Moosdijk A, Van-Amerongen R. Identification of reliable reference genes for qRT-PCR studies of the developing mouse mammary gland. Sci Rep 2016;6:35595. https://doi.org/10.1038/srep35595. PMid: 27752147.

[4] Zhou P, Huang L, Wang Y, et al. Stepwise optimization of the RT-qPCR protocol and the evaluation of housekeeping genes in pears (*Pyrus bretschneideri*) under various hormone treatments and stresses. Horticulturae 2023;9(2):275. https://doi.org/10.3390/horticulturae9020275.

[5] Nguyen DQ, Nguyen NL, Nguyen VT, et al. Reliable reference genes for accurate gene expression profiling across different tissues and genotypes of rice seedlings (*Oryza sativa* L.) under salt stress. Russ J Plant Physiol 2023;70(5):104. https://doi.org/10.1134/S102144372360068X.

[6] Wang M, Bhullar NK. Selection of suitable reference genes for qRT-PCR gene expression studies in rice. In: Bandyopadhyay A, Thilmony R, editors. Rice genome engineering and gene editing. Methods in Molecular Biology. New York: Humana; 2021. p. 293–312. https://doi.org/10.1007/978-1-0716-1068-8_20. PMid: 33471340.

[7] Zhang L, Liu L, Cheng P, et al. Identification and validation of reference genes for RT-qPCR analysis in banana (*Musa* spp.) under Fusarium wilt resistance induction conditions. J Phytopathol 2017;165(11–12):746–54. https://doi.org/10.1111/jph.12614.

[8] Zhu L, Yang C, You Y, et al. Validation of reference genes for qRT-PCR analysis in peel and flesh of six apple cultivars (*Malus domestica*) at diverse stages of fruit development. Sci Hortic 2019;244:165–71. https://doi.org/10.1016/j.scienta.2018.09.033.

[9] Nocum JD, Manohar AN, Mendoza JV, et al. Identification of suitable internal control genes for gene expression analysis of banana in response to BBTV infection. Plant Gene 2022;32:100383. https://doi.org/10.1016/j.plgene.2022.100383.

[10] Kumar G, Singh AK. Reference gene validation for qRT-PCR based gene expression studies in different developmental stages and under biotic stress in apple. Sci Hortic 2015;197:597–606. https://doi.org/10.1016/j.scienta.2015.10.025.

[11] Zhu L, Lin Y, Yang W, et al. The identification of the banana endogenous reference gene *MaSPS1* and the construction of qualitative and quantitative PCR detection methods. Genes 2023;14(12):2116. https://doi.org/10.3390/genes14122116. PMid: 38136937.

[12] Song Y, Hanner RH, Meng B. Genome-wide screening of novel RT-qPCR reference genes for study of GLRaV-3 infection in wine grapes and refinement of an RNA isolation protocol for grape berries. Plant Methods 2021;17:110. https://doi.org/10.1186/s13007-021-00808-4. PMid: 34711253.

[13] Wei TL, Wang H, Pei MS, et al. Identification of optimal and novel reference genes for quantitative real-time polymerase chain reaction analysis in grapevine. Aust J Grape Wine Res 2021;27(3):325–33. https://doi.org/10.1111/ajgw.12483.

[14] Zha Q, Xi X, Jiang A, et al. Identification of the appropriate reference gene through using a real-time quantitative PCR in grapes. J Fruit Sci 2016;33(3):268–74. https://doi.org/10.13925/j.cnki.gsxb.20150327.

[15] Ni J, Liao Y, Zhang M, et al. Blue light simultaneously induces peel anthocyanin biosynthesis and flesh Carotenoid/Sucrose biosynthesis in mango fruit. J Agric Food Chem 2022;70(50):16021–35. https://doi.org/10.1021/acs.jafc.2c07137. PMid: 36484494.

[16] Hoang VL, Innes DJ, Shaw PN, et al. Sequence diversity and differential expression of major phenylpropanoid-flavonoid biosynthetic genes among three mango varieties. BMC Genomics 2015;16:561. https://doi.org/10.1186/s12864-015-1784-x. PMid: 26220670.

[17] Bajpai A, Khan K, Muthukumar M, et al. Molecular analysis of anthocyanin biosynthesis pathway genes and their differential expression in mango peel. Genome 2018;61(3):157–66. https://doi.org/10.1139/gen-2017-0205. PMid: 29338343.

[18] Hong K, Gong D, Xu H, et al. Effects of salicylic acid and nitric oxide pretreatment on the expression of genes involved in the ethylene signalling pathway and the quality of postharvest mango fruit. New Zeal J Crop Hort Sci 2014;42(3):205–16. https://doi.org/10.1080/01140671.2014.892012.

[19] Alghanem SM, Alnusairi GS, Alkhateeb MA, et al. Genome-wide identification and characterization of the *dof* gene family in mango (*Mangifera indica* L.). Genet Resour Crop Evol 2024;71:2749–65. https://doi.org/10.1007/s10722-023-01767-6.

[20] Karanjalker GR, Ravishankar KV, Shivashankara KS, et al. A study on the expression of genes involved in carotenoids and anthocyanins during ripening in fruit peel of green, yellow, and red colored mango cultivars. Appl Biochem Biotechnol 2018;184(1):140–54. https://doi.org/10.1007/s12010-017-2529-x. PMid: 28643121.

[21] Yao R, Huang X, Cong H, et al. Selection and identification of a reference gene for normalizing real-time PCR in mangos under various stimuli in different tissues. Horticulturae 2022;8(10):882. https://doi.org/10.3390/horticulturae8100882.

[22] Bai Y, Lv YN, Zeng M, et al. Selection of reference genes for normalization of gene expression in *Thermobia domestica* (Insecta: Zygentoma: Lepismatidae). Genes 2020;12(1):21. https://doi.org/10.3390/genes12010021. PMid: 33375665.

[23] Li X, Gong P, Wang B, et al. Selection and validation of experimental condition-specific reference genes for qRTPCR in *Metopolophium dirhodum* (Walker) (Hemiptera: Aphididae). Sci Rep 2020;10:21951. https://doi.org/10.1038/s41598-020-78974-z. PMid: 33319828.

[24] Medina-Lozano I, Arnedo MS, Grimplet J, et al. Selection of novel reference genes by RNA-Seq and their evaluation for normalising Real-Time qPCR expression data of anthocyanin-related genes in lettuce and wild relatives. Int J Mol Sci 2023;24(3):3052. https://doi.org/10.3390/ijms24033052. PMid: 36769376.

[25] Sarwar MB, Ahmad Z, Anicet BA, et al. Identification and validation of superior housekeeping gene(s) for qRT-PCR data normalization in *Agave sisalana* (a CAM-plant) under abiotic stresses. Physiol Mol Biol Plants 2020;26:567–84. https://doi.org/10.1007/s12298-020-00760-y. PMid: 32205931.
[26] Expósito-Rodríguez M, Borges AA, Borges-Pérez A, et al. Selection of internal control genes for quantitative real-time RTPCR studies during tomato development process. BMC Plant Biol 2008;8:131. https://doi.org/10.1186/1471-2229-8-131. PMid: 19102748.
[27] Untergasser A, Cutcutache I, Koressaar T, et al. Primer3–new capabilities and interfaces. Nucleic Acids Res 2012;40(15):e115. https://doi.org/10.1093/nar/gks596. PMid: 22730293.
[28] Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3(7):research0034.1. https://doi.org/10.1186/gb-2002-3-7-research0034. PMid: 12184808.
[29] 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 Res 2004;64(15):5245–50. https://doi.org/10.1158/0008-5472.CAN-04-0496. PMid: 15289330.
[30] Pfaffl MW, Tichopad A, Prgomet C, et al. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. Biotechnol Lett 2004;26:509–15. https://doi.org/10.1023/B:BILE.0000019559.84305.47. PMid: 15127793.
[31] Silver N, Best S, Jiang J, et al. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol Biol 2006;7:33. https://doi.org/10.1186/1471-2199-7-33. PMid: 17026756.
[32] Xie J, Alves AdS, von der Haar T, et al. Regulation of the elongation phase of protein synthesis enhances translation accuracy and modulates lifespan. Curr Biol 2019;29(5):737–49. https://doi.org/10.1016/j.cub.2019.01.029. PMid: 30773367.
[33] Brunner AM, Yakovlev IA, Strauss SH. Validating internal controls for quantitative plant gene expression studies. BMC Plant Biol 2004;4:14. https://doi.org/10.1186/1471-2229-4-14. PMid: 15317655.
[34] Dean JD, Goodwin PH, Hsiang T. Comparison of relative RT-PCR and northern blot analyses to measure expression of β-1,3-glucanase in *Nicotiana benthamiana* infected with *Colletotrichum destructivum*. Plant Mol Biol Rep 2002;20:347–56. https://doi.org/10.1007/BF02772122.
[35] Gutierrez L, Mauriat M, Guénin S, et al. The lack of a systematic validation of reference genes: A serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnol J 2008;6(6):609–18. https://doi.org/10.1111/j.1467-7652.2008.00346.x. PMid: 18433420.
[36] Lorković ZJ, Wieczorek Kirk DA, Klahre U, et al. RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting with poly(A)+ RNA in nuclei of plant cells. RNA 2000;11:1610–24. https://doi.org/10.1017/s1355838200001163. PMid: 11105760.
[37] Yan Y, Gan J, Tao Y, et al. RNA-Binding Proteins: The key modulator in stress granule formation and abiotic stress response. Front Plant Sci 2022;13:882596. https://doi.org/10.3389/fpls.2022.882596. PMid: 35783947.
[38] Weber C, Nover L, Fauth M. Plant stress granules and mRNA processing bodies are distinct from heat stress granules. Plant J 2008;56(4):517–30. https://doi.org/10.1111/j.1365-313X.2008.03623.x. PMid: 18643965.
[39] Czechowski T, Stitt M, Altmann T, et al. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol 2005;139(1):5–17. https://doi.org/10.1104/pp.105.063743. PMid: 16166256.
[40] Mallona I, Lischewski S, Weiss J, et al. Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. BMC Plant Biol 2010;10:4. https://doi.org/10.1186/1471-2229-10-4. PMid: 20056000.
[41] Wan H, Zhao Z, Qian C, et al. Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber. Anal Biochem 2010;399(2):257–61. https://doi.org/10.1016/j.ab.2009.12.008. PMid: 20005862.
[42] Jiang Q, Wang F, Tan HW, et al. *De novo* transcriptome assembly, gene annotation, marker development, and miRNA potential target genes validation under abiotic stresses in *Oenanthe javanica*. Mol Genet Genomics 2015;290:671–83. https://doi.org/10.1007/s00438-014-0953-y. PMid: 25416420.
[43] Rebouças EDL, Costa JJDN, Passos MJ, et al. Real time PCR and importance of housekeepings genes for normalization and quantification of mRNA expression in different tissues. Braz Arch Biol Technol 2013;56:143–54. https://doi.org/10.1590/S1516-89132013000100019.
[44] Xie F, Wang J, Zhang B. RefFinder: A web-based tool for comprehensively analyzing and identifying reference genes. Funct Integr Genomics 2023;23:125. https://doi.org/10.1007/s10142-023-01055-7. PMid: 37060478.
[45] Brönstrup M, Sasse F. Natural products targeting the elongation phase of eukaryotic protein biosynthesis. Nat Prod Rep 2020;37(6):752–62. https://doi.org/10.1039/D0NP00011F. PMid: 32428051.
[46] Xu B, Liu L, Song G. Functions and regulation of translation elongation factors. Front Mol Biosci 2022;8:816398. https://doi.org/10.3389/fmolb.2021.816398. PMid: 35127825.
[47] Lukash TO, Turkivska HV, Negrutskii BS, et al. Chaperone-like activity of mammalian elongation factor eEF1A: Renaturation of aminoacyl-tRNA synthetases. Int J Biochem Cell Biol 2004;36(7):1341–7. https://doi.org/10.1016/j.biocel.2003.11.009. PMid: 15109577.
[48] Novosylna O, Jurewicz E, Pydiura N, et al. Translation elongation factor eEF1A1 is a novel partner of a multifunctional protein Sgt1. Biochimie 2015;119:137–45. https://doi.org/10.1016/j.biochi.2015.10.026. PMid: 26545799.
[49] Losada A, Muñoz-Alonso MJ, García C, et al. Translation elongation factor eEF1A2 is a novel anticancer target for the marine natural product plitidepsin. Sci Rep 2016;6(1):35100. https://doi.org/10.1038/srep35100. PMid: 27713531.