Selection and validation of suitable reference genes for gene expression studies in *Callerya speciosa* (Champ. ex Benth.) Schot under different experimental conditions

Linchan Yu  
Guangxi University of Chinese Medicine

Ruhong Ming  
Guangxi University of Chinese Medicine

Ding Huang  
Guangxi University of Chinese Medicine

Chunmei Qin  
Guangxi University of Chinese Medicine

Liangbo Li  
Guangxi University of Chinese Medicine

Yong Tan  
Guangxi University of Chinese Medicine

Rongshao Huang  
Guangxi University of Chinese Medicine

Shaochang Yao  (yaoshaochang@163.com)  
Guangxi University of Chinese Medicine

**Research Article**

**Keywords:** reference gene validation, RT-qPCR, Callerya speciosa, different experimental conditions

**Posted Date:** April 11th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1291888/v2

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. 
[Read Full License]
Abstract

The accuracy of reverse transcription quantitative real-time PCR (RT-qPCR) is strongly depended on the stability of reference gene. Callerya speciosa, as a traditionally Chinese medicine, has a long cultivation history in south China. It is essential to select the suitable reference gene to obtain reliable RT-qPCR results when gene expression changes were evaluated. However, suitable reference genes in C. speciosa have not yet been investigated for accurate gene expression quantification under different experimental conditions. In this study, eight candidate reference genes (GAPDH, 60S, ACTIN, TUA2, TUB1, TIF5, UBQ, EF2) were selected from the transcriptome databases, and their expression stability under six experimental conditions (developmental stages, tissues, MeJA treatment, GA3 treatment, CPPU and PP333 treatment) was evaluated using ΔCT, geNorm, NormFinder, BestKeeper, RefFinder programs. The results showed that GAPDH was the optimal reference gene for all different experimental conditions, whereas ACTIN showed the most stability under the hormone treatments in C. speciosa. GAPDH and EF2 were proved to be the most stable genes for developmental stages, while different genes (GAPDH and TUB1) were stable reference genes for tissues. For treatments, ACTIN was identified as the most stable gene under most of hormone treatments. TUBI and ACTIN were at the beginning of the ranking order in PP333 treatment, while GAPDH and ACTIN were adequate for normalization in MeJA and GA3 treatments. TUBI and GAPDH were the most stable genes for CPPU treatment, while ACTIN was proved to be the most stable gene for three different treatments (MeJA, GA3 and PP333). Validation of reference genes was carried out by the target gene CsMYB36, which further confirmed their reliability. These results provided a theoretical basis for subsequent research on the regulation of functional gene expression in C. speciosa.

Introduction

Gene expression analysis has become an important method to reveal gene function and molecular mechanism. Due to its cost-effectiveness and relative simplicity, reverse transcription quantitative real-time PCR (RT-qPCR) is a popular tool in molecular biology to study gene expression [1,2]. However, RT-qPCR analysis is highly sensitive, affected by several factors, including the intrinsic variability of RNA [3], the specificity of the reaction, the stability of the applied reference gene. The result of RT-qPCR is accurate, only when the experimental settings are properly performed and appropriate normalization methods are employed [4]. In addition, the lack of validation of reference genes questions the accuracy of results obtained by RT-qPCR using non-valid references, implying the need for a systematic validation of reference genes [5]. Housekeeping genes are not expected to change under the experimental conditions serve as internal standard, generally used as reference gene [6,7]. Traditionally, housekeeping genes include Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), eukaryotic translation initiation factor 5A (EF5), 60S ribosomal protein L34 (60S) ubiquitin-conjugating enzyme E2 (UBC), polyubiquitin (UBQ), tubulin beta-1 chain (TUBI), elongation factor (EF2) ACTIN, α-Tubulin, β-Tubulin, 18S rRNA, 25S rRNA [8,9]. Therefore, selecting suitable reference genes is crucial to analyze the gene expression level under different experimental conditions by RT-qPCR. Several modern statistical algorithm tools have been
recommended to find the suitable reference genes, such as $\Delta CT$ [10], BestKeeper [11], geNorm [12], Normfinder [13], and RefFinder [14].

*Callerya speciosa* (Champ. ex Benth.) Schot is a perennial medicinal plant belonging to the Fabaceae family, and has a long cultivation history in south China. *C. speciosa*, commonly called Niu Da Li, possesses high medicinal value, and has been used as a Chinese medicine since the Ming Dynasty [15]. In recent decades, the pharmacology and phytochemistry researches have proved that *C. speciosa* exhibited a number of pharmacological effects of enhancing human immunity, anti-fatigue, expectorant effects, clinically used for invigorating vital energy and relieving asthma and strengthening the bones and muscles [16]. (Iso)flavonoids are the main components in *C. speciosa*, and the index compounds maackiain and formononetin are considered to be the important indicators to control the quality [17]. At present, more attentions are paid to *C. speciosa* because of its health benefits. However, the current molecular mechanisms of the (iso)flavonoids biosynthesis in *C. speciosa* is still limited.

In recent years, several progresses have been made with respect to reference gene normalization for RT-qPCR, demonstrating that expression of reference genes changed considerably in different developmental stages and different experimental conditions. For example, in Gossypium hirsutum L. salt stress, UBQ7, GAPDH and EF1A8 were the better reference genes in leaves, while the better reference genes were TUA10, UBQ7, CYP1, GAPDH and EF1A8 in roots [18]. For Bromus sterilis, three developmental stages (2-leaves stage, 3-leaves stage and 4-leaves stage), 18S rRNA and ACCase were the better reference gene [19]. Under ABA treatment, the ideal reference genes of Nitraria tangutorum were EF1-α and His [20]. Our previous studies revealed that several R2R3-MYB TFs and structural genes might be involved regulating isoflavonoid biosynthesis in *C. speciosa*, and CsMYB36 has the highest expression in R2R3-MYB TFs [21]. To understand these related gene expression changes will be facilitated to explain their functions in *C. speciosa* under different experimental conditions. Plant growth regulators play a vital role in the adaptation of plants to the environment, which can regulate the growth, development and reproduction of plants. Methyl jasmonic (MeJA), gibberellic acid (GA$_3$), N-phenyl-N'-(2-chloro-4-pyridyl) urea (CPPU) and paclobutrazol (PP$_{333}$) are important abiotic elicitors involved in abiotic stresses [22], accelerating cell division and plant growth [23]. Our previous research found that MeJA might be linked to the initial development of tuberous roots, GA$_3$ was considered essential for rapid thickening of tuberous roots [24]. With the development of molecular biology, gene expression analysis may help reveal the molecular mechanism of stress responses and root thickening in *C. speciosa*. Therefore, screening for *C. speciosa* reference genes under different experimental conditions are of great significance.

In this study, the expression stabilities of eight candidate reference genes ((GAPDH (glyceraldehyde-3-phosphate dehydrogenase), 60S (60S ribosomal protein L34), ACTIN, TUA2 (tubulin alpha chain), TUB1 (tubulin beta-1 chain), TIF5 (eukaryotic translation initiation factor 5A), UBQ (polyubiquitin), EF2 (elongation factor 2)) under different experimental conditions were evaluated. Their expression levels under a serious of experimental conditions, including different tissues (root, stem, leaf, flower and seed), roots at four different developmental stages (6, 12, 18, 30 months after germination (MAG)), and growth regulator treatments (MeJA, GA$_3$, CPPU, PP$_{333}$), were detected by RT-qPCR. Five different software
packages (ΔCT, geNorm, NormFinder, BestKeeper, and RefFinder) were used to evaluate the expression stability of eight genes. In order to further verify the reliability of the reference gene selection results, we analyzed the expression patterns of *CsMYB36*, a gene related to the regulation of the isoflavonoid synthesis. These results provide a theoretical basis for subsequent research on the regulation of functional gene expression in *C. speciosa*.

Results

2.1 Verification of primer specificity. A set of 48 pooled samples were selected to evaluate the primer specificity, including five different tissues (roots, stems, leaves, flowers, and seeds), four different developmental stages (6, 12, 18, 30 MAG), and nine different treatments by plant growth regulators. For eight reference genes, the melting curve was a single peak by RT-qPCR analysis, and the no-template control in RT-qPCR reactions are no signal (Supplementary Fig. S1 online). These results suggested that there were no genomic DNA contaminants and non-specific amplification. The amplification efficiencies \( E \) ranged from 123% to 166.9%, and correlation coefficients \( R^2 \) was higher than 0.981 (Supplementary Table S2 online).

2.2 Analysis of the Ct value of reference genes. The Ct value usually is inversely proportional to the transcription level of a gene, so that lower expression abundance had higher Ct values. In this study, the Ct values of eight candidate reference genes changed under a series of experimental conditions, but the change trends were different. The lowest and highest Ct values were found in *UBQ* and *EF2* at 20.15 and 32.14, respectively, indicating a large fluctuation range of expression. Five candidate reference genes (*GAPDH, 60S, ACTIN, TIF5* and *UBQ*) had relatively high expression abundances, and the mean Ct of *60S* is the lowest (21.88). Another three genes (*GAPDH, TIF5* and *UBQ*) showed the highest stability, followed by *60S* with the mean Ct values of 22.27, 22.77, and 22.79, respectively (Fig. 2). These results suggested that it’s necessary to evaluate the stability of candidate reference genes under different experimental conditions.

2.3 ΔCT analysis. The method of ΔCT were evaluated the most stability reference genes based on Ct values. As shown in Table 1, the most stable gene in the eight different sample groups (all samples, all treatments, developmental stages, tissues, MeJA, CPPU, GA₃, and PP₃₃₃) recommended by ΔCT analysis was *GAPDH, TUB1, GAPDH, GAPDH, ACTIN, UBQ, ACTIN*, and *TUB1*, respectively.

2.4 GeNorm analysis. Using the geNorm, the M value of candidate reference genes are ranked from lowest to highest. The lower M value represents more stable reference genes. In all case, the M values were < 1.5, suggesting that the candidate reference genes all had relatively acceptable stability values.
For all samples, the *GAPDH* have the lowest M value. For development stages and tissues, *TIF5* and *GAPDH* were the most stable reference genes, respectively. For treatments, the optimal candidate reference genes were *TIF5*, *GAPDH*, *60S*, and *ACTIN* in all treatments, MeJA treatment, CPPU treatment, *GA₃* treatment and PP333 treatment groups, respectively (Table 2). Notably, *geNorm* analysis indicated that *GAPDH* might be the most stable reference gene for all samples, while the stability of the conference genes under different treatments was not constant.

The optimal number of reference genes could also be recommended by *geNorm* program based on the pairwise variation value ($V_{n/n+1}$). The criterion is when $V_{n/N}/V_{n+1}<0.15$, then the most suitable reference gene number is $n$. For all samples and treatments, the $V_{7/8}>0.15$, suggesting that more than seven reference genes were needed to evaluate the gene expression. For developmental stages and PP333 treatments, the $V_{3/4}$ values were $<0.15$, indicating that the three recommended reference genes would be adequate for gene normalization. However, gene normalization in tissues, MeJA treatments, CPPU treatments and *GA₃* treatments, only needed two reference genes, since the $V_{2/3}$ values were all $<0.15$ (Fig. 4).

### 2.5 NormFinder analysis

The calculation principle of *NormFinder* is similar to *geNorm*, the expression stability values (M) were calculated according to the CT value of the reference genes. The gene with the smallest M value is the most stable reference gene. As shown in Table 2, the stability of reference genes was ranked by *NormFinder* algorithm. For all samples and tissues, *GAPDH* was the lest expressed gene. *ACTIN* was the most expressed stable reference gene under MeJA and *GA₃* treatment. However, the most stable gene in all treatments, developmental stages, CPPU treatment, and PP333 treatment groups, was *TUB1, EF2, UBQ*, and *TUB1*, respectively. Consistent with the results of *geNorm* analysis, *GAPDH* was recommended to be the three-top stable reference gene in several groups. However, differences were also observed between the *geNorm* and *NormFinder* algorithms. For example, *TIF5* gene was the most stable one by *geNorm* while ranked it as the first in all treatments group (Table 2).

### 2.6 BestKeeper analysis

Distinct from the *geNorm* and *NormFinder* algorithms, the *BestKeeper* method evaluates gene stability based on the standard deviation (SD) and percentage covariance (CV) of candidate reference genes. The larger the SD value, the worse the stability of the gene. When SD $>1$, the expression of the internal reference gene is unstable. The analysis results are shown on Table 3. In all samples, the SD values of *GAPDH, ACTIN, TUA2* and *EF2* were $>1$, indicating that they were not suitable candidate reference genes. In developmental stages, tissues, MeJA treatments, CPPU treatments, *GA₃* treatments and PP333 treatments, the gene with the smallest SD value was *GAPDH, TUBI, TUBI, TIF5, UBQ, UBQ*, respectively. Interestingly, SD value $>1$ was generated in many gene under different experimental conditions, such as *ACTIN* in all treatments and CPPU treatments, *TUA2* in all treatments, developmental...
stages, and CPPU treatments (Table 3). These results indicated that the BestKeeper analysis differed obviously from the geNorm and NormFinder algorithms.

2.7 RefFinder analysis. Based on the results of geNorm, NormFinder, BestKeeper and Delta CT, the online tool RefFinder is used to calculate the geometric mean to obtain the final comprehensive index ranking of gene stability. As shown in Table 4, the ranking of the eight candidate genes for the entire sample was as follow: 60S > TIF5 > GAPDH > TUBI > EF2 > UBQ > ACTIN > TUA2. GAPDH was almost always the top-ranked one for all samples, developmental stages, tissues, MeJA treatment, CPPU treatment, GA3 treatment, while TUBI and ACTIN were at the beginning of the ranking order for PP333 treatment. TUBI and UBQ were the most stable genes for all treatment, while ACTIN was proved to be the most stable gene for three different treatments (MeJA, GA3 and PP333).

2.8 Reference gene validation. CsMYB36 is a key gene in R2R3-MYB transcription factor family of C. speciosa, which might play an important role in regulating isoflavone synthesis. Therefore, CsMYB36 was selected as target gene to investigate the expression profile using the two most stable reference genes and the least stable reference gene in this study, in order to validate the reliability and stability of the candidate reference genes determined by the applied algorithms. Our results showed that the relative expression of CsMYB36 increased significantly in root and flower when using the two most stable reference genes (GAPDH and TUB1) and the combination of GAPDH + TUB1 (p < 0.001), while the expression levels did not have dramatically different when using the least stable reference gene (EF2). The expression trends of CsMYB36 were similar when the most stable reference gene or the combination of the two most stable reference genes. Consistent with different tissues, the expression trends of CsMYB36 were also similar when using the single and combination of the two most stable reference genes in different developmental stages (6, 12, 18, 30 MAG) and MeJA treatments (0, 3, 24, 48 h) (Fig. 5). Therefore, the results indicated that the stable reference genes strongly affected the relative expression levels of the target gene. If the reference gene has high stability, one gene is adequate for normalization.

Discussion

Gene expression patterns help to in-depth study of the mechanism and function of some unknown genes in animals, plants and microorganisms. There are many methods in molecular biology for gene expression, including Northern and Southern hybridizations, scintillation proximity assay, PCR-ELISA and RNase protection assay [6]. However, most of these methods have some shortcomings as follow: time consuming, labor intensive, nonquantitative, insufficiently sensitive, and cross contamination. Real-time polymerase chain reaction (RT-qPCR) is easy to perform with high sensitivity and specificity, which provides the accuracy and produces reliable as well as rapid quantification results [25,26], therefore it has emerged as a robust and widely used methodology for gene expression. To date, some plants have been
selected for research on RT-qPCR analysis, including *Glycyrrhiza* [27], *Nitraria tangutorum* [20], *Fucus distichus* [28], *Bromus sterilis* [19], *Allium tuberosum* [29], and *Suaeda glauca* [30]. In addition, RT-qPCR is widely used in diagnosing diseases, such as preterm birth [31], diabetic kidney disease [32], Bronchopulmonary dysplasia [33]. Due to its rapidity and accuracy in informing on active coronavirus (CoV) infection, RT-qPCR has become the assay of choice for COVID-19 diagnosis [34,35]. However, RT-qPCR requires a stable internal standard as a reference gene. Previous studies reported that the expression profiles of reference genes might be unstable under different species, tissues, developmental stages, and abiotic stress [9,18,36]. And an unstable reference gene will cause misleading and even contradictory results [37,38]. Therefore, it’s always necessary to validate the optimal reference genes prior to their applications.

*C. speciosa* has attracted more attention own to its medicinal and health care values. Unfortunately, there has been no research on the normalization of appropriate reference gene for RT-qPCR analysis in *C. speciosa*. In our study, eight typical candidate reference genes (GAPDH, 60S, ACTIN, TUA2, TUB1, TIF5, UBQ, EF2) were selected from the transcriptome database of *C. speciosa*, and their expression stability under six different experimental conditions (developmental stages, tissues, MeJA treatment, GA3 treatment, CPPU treatment, and PP333 treatment) was evaluated using ΔCT, geNorm, NormFinder, BestKeeper, and RefFinder programs. However, the most stable genes recommended by the four different programs were not always the same due to their different calculations. Similar to geNorm algorithm, the most suitable reference genes selected by NormFinder are generally the same (Table 2). Notably, the ranking order of reference genes when using BestKeeper was quite different from other algorithms (Table 3), which was consistent with other studies [39,40]. GAPDH is widely used as a reference gene, which converts glyceraldehyde-3-phosphate into 1, 3-bisphosphoglycerate and is an essential component of the glycolytic pathway. It was reported that GAPDH has been selected as the most stable gene in different tissues of *Gossypium hirsutum* L. [18] and *Dendrocalamus latiflorus* Munro [41]. GAPDH displayed the maximum stability for most of single abiotic stresses in carrot [42]. Similarly, GAPDH was almost always the top-ranked gene for all different experimental conditions, and TUA2 was generally the least stable one in *C. speciosa* (Table 4). Additionally, ACTIN, which is the essential components of the eukaryotic cytoskeleton, was the most stable reference gene under the selected abiotic stress and hormone treatments in celery [43], and carrot [42]. In *C. speciosa*, GAPDH and EF2 proved to be the most stable genes for developmental stages, while different genes (GAPDH and TUB1) were stable reference genes for tissues. For treatments, TUB1 and ACTIN were at the beginning of the ranking order in PP333 treatment, while GAPDH and ACTIN were adequate for normalization in MeJA and GA3 treatments. TUB1 and GAPDH were the most stable genes for CPPU treatment, while ACTIN was proved to be the most stable gene for three different treatments (MeJA, GA3 and PP333) (Fig. 3 and Table 4). Therefore, we suggested that GAPDH was the optimal reference gene for all different experimental conditions, whereas ACTIN showed the most stability under the hormone treatments in *C. speciosa*.

To validate the accuracy of the experimental results, an isoflavonoid synthesis gene CsMYB36 was chosen to be the target gene, and two most suitable reference genes and the least stable reference gene
were chosen as reference genes. Our work revealed that the expression profiles of the *CsMYB36* gene were similar when normalized by single reference genes with very high stability (Fig. 5). In contrast, there was a significant difference in target gene expression with the least stable gene *TUA2* as reference gene. These results suggested that no significant difference was found when using the stable reference gene in RT-qPCR normalization in *C. speciosa*. However, many previous studies reported that the application of one reference gene was not adequate for normalization [12,30]. For instance, the combinations of *ACTIN*/TUB or *ACTIN*/EFIA were recommended for their use in the pooled analysis responded to biotic and abiotic stress in *Vigna mungo* [2]. In agreement with this, the combination of other housekeeping genes also showed high expression stabilities, such as *PP2A* and *TUA5* in *Suaeda glauca* seeds [30], as well as *UBQ* and *EF-1α* in *Tectona grandis* L.f. [44]. According to the pairwise variation parameters by geNorm program, two reference genes were recommended for four different experimental conditions (tissues, MeJA, CPPU, and GA₃ treatments), while the V3/4 value was below the threshold of 0.15 for samples in developmental stages and under PP₃₃₃ treatment in this study (Fig. 4). As shown in Fig. 5, the expression level of the target gene were significantly upregulated in five different tissues and under MeJA treatment when a single or a combination of reference genes were used, while it was remarkably upregulated in the 18 MAG developmental stages. These results suggested that one reference gene might be sufficient for accurate normalization in *C. speciosa*.

**Conclusions**

For the first time, our work aimed to identify the most stable reference genes for RT-qPCR analysis in different experimental conditions in *C. speciosa*. Eight candidate reference genes (*GAPDH, 60S, ACTIN, TUA2, TUB1, TIF5, UBQ, EF2*) were selected from the transcriptome database of *Callerya speciosa*, and their expression stability under six experimental conditions (developmental stages, tissues, MeJA treatment, GA₃ treatment, CPPU and PP₃₃₃ treatment) was evaluated using ΔCT, geNorm, NormFinder, BestKeeper, RefFinder programs. Meanwhile, the target gene *CsMYB36* was used to validate the results recommended by programs. The results showed that *GAPDH* was the optimal reference gene for all different experimental conditions, whereas *ACTIN* showed the most stability under the hormone treatments in *C. speciosa*. *GAPDH* and *EF2* were proved to be the most stable genes for developmental stages, while different genes (*GAPDH* and *TUB1*) were stable reference genes for tissues. For treatments, *ACTIN* was identified as the most stable gene under most of hormone treatments. *TUB1* and *ACTIN* were at the beginning of the ranking order in PP₃₃₃ treatment, while *GAPDH* and *ACTIN* were adequate for normalization in MeJA and GA₃ treatments. *TUB1* and *GAPDH* were the most stable genes for CPPU treatment, while *ACTIN* was proved to be the most stable gene for three different treatments (MeJA, GA₃ and PP₃₃₃). Validation of reference genes was carried out by the target gene *CsMYB36*, which further confirmed their reliability. In a word, our work contributes a suitable reference for selecting stable internal reference gene candidates to investigate gene expression in *C. speciosa*, and provides a theoretical basis for subsequent research on the regulation of functional gene expression and the understanding of molecular mechanisms in *C. speciosa*. 
Materials And Methods

5.1 Plant Materials and Treatment. The samples in this experiment were collected from the planting base of Guangxi Guangze Health Industry Co., Ltd., Nanning, China. There different tissues (root, stem, leaf) were collected from the 30-month seedlings, flowers are collected at the bud stage, seeds are collected at ripe. tuberous roots at four developmental ages were also collected, including 6, 12, 18, 30 months after germination (MAG) (Fig. 1). Mature seeds were selected to germinate on wet filter under dark conditions at 26 ± 2°C for 10 days. Two days after germination, seedlings were cultured in plastic pots filled with a 3:1 (v:v) mixture of loamy soil and pearlite at 26 ± 2°C with 16-h/8-h light/dark photoperiod, 300 μmol m⁻² s⁻¹ light intensity, and 60% relative humidity in an artificial weather box (MGC-400H, Shanghai, China) in GuangXi University of Traditional Chinese Medicine. For CPPU, GA₃, and PP₃₃₃ treatment, the 3-month-old seedlings were sprayed with 10 mg·L⁻¹, 50 mg·L⁻¹, and 100 mg·L⁻¹ for 0, 12, 48, 72 h, respectively. The seedlings were sprayed with clean water as control condition. Meanwhile, three different MeJA concentrations (200, 400, 600 μmol L⁻¹) for 48 h and six different time intervals (0, 3, 6, 24, 48 and 72 h) for 200 μmol L⁻¹ MeJA were also treated on the 3-month-old seedlings. After treatment, the seedlings were cultured in the above-mentioned environments. The roots of seedlings were sampled after washing with distilled water. Thereafter, all samples are frozen immediately in liquid nitrogen and stored at −80°C until further use.

5.2 Total RNA extraction and reverse transcription. Total RNA was extracted from samples using the Plant Total RNA Extraction Kit BSC65S1 (Biospin Technology Co., Ltd, Hangzhou, China), according to the manufacturer’s instructions. The RNA purity and concentration were measured with spectrophotometer (Tiangen Biotech Co., Ltd.) and integrity was checked on 1% agarose gel electrophoresis. RNA samples with a concentration higher than 300 ng/μL and A₂₆₀/A₂₈₀ ratio between 2.0 and 2.3 were required for cDNA preparation (Supplementary Table S1 online). cDNA was synthesized using the PrimeScript™ RT reagent Kit with gDNA Eraser (Perfect Real Time) (Takara, Dalian, China) according to the manufacturer’s instructions. The products were stored at −20°C until further analyses.

5.3 Selection of candidate reference genes and primer design. Eight candidate reference genes (GAPDH, 60S, ACTIN, TUA2, TUB1, TIF5, UBQ, EF2) were selected from the transcriptome databases (SRA accession No. SRP223620), which have stable expressions (|log₂ ratio|≤0.5) in the roots of C. speciosa and often selected as housekeeping genes for RT-qPCR analysis in different plant species. The primers were designed by Primer 5.0 software according to their CDS sequences. The design Principles were as follows: the length of primers and amplification product size were 15-25 bp and 80-120 bp, respectively; the melting temperature (Tm), 58°C; GC content, 40%-60% [45]. All primers were synthesized by the BGI Tech Solutions Co., Ltd. (BGI Tech) (BGI, Shenzhen, P. R. China). The primer information used in this study was listed in Supplementary Table S1 online.
5.4 RT-qPCR analyses. The RT-qPCR reaction was performed using a CFX96™ real-time system (Bio-Rad, America). The 10 µL reaction system consisted of SYBR® Premix Ex Taq™ (5 µL), both the forward and reverse primers (0.5 µL, 5000 ng), 0.5 µL of the cDNA template (diluted five-fold with RNase-free ddH2O, and 3.5 µL of RNase-freed ddH2O. All the PCRs were performed under the following conditions: one cycle of 15 min at 95°C for activation, followed by 40 cycles of 10 sec at 95°C, 30 sec at 60°C, and 30 sec at 72°C. Three technical replicates were conducted for each of the three biological replicates.

5.5 Primer amplification efficiency and linearity inspection. The standard curve was constructed with a five-step dilution series of cDNA pool (equal volume of the cDNA from all the samples to be the pool). The amplification efficiency ($E$) and correlation coefficient ($R^2$) of primers were calculated using the standard curve. The amplification efficiency ($E$) of each primer pair was calculated by the curve slope using $E = 10^{(-1/k)} - 1 \times 100$% [46]. The $E$ and $R^2$ of primer were listed in Supplementary Table S2 online. Every candidate gene in each sample was performed in three biological replicates with three technological replicates. PCR reactions with no-template controls were also conducted for each primer pair, to check the absence of primer dimers and random contaminations.

5.6 Stability analysis of candidate reference genes. The stability of eight candidate reference genes was evaluated by four Microsoft Excel-based computational programs (geNorm, NormFinder, BestKeeper and $\Delta CT$), and an online data analysis program RefFinder. For geNorm analysis, the Ct values needed to be converted into a relative quantity according to the formula: $2^{-\Delta CT}$, where $\Delta CT$ is equal to the corresponding Ct value minus the minimum Ct value [47]. The geNorm algorithm was applied to evaluate the average expression stability values (M) and the pairwise variation (V) values of reference genes. Gene with an M value below the threshold 1.5 was thought to be the stable one. In addition, this program could also recommend the optimal number of reference genes by calculating the pairwise variation ($V_n/V_{n+1}$). When the pairwise variation was $< 0.15$, the optimal number of reference genes is n. The NormFinder was used to analyze the stable values of the expression stability by calculating the variance, including intra group and inter group variance. The lowest values indicated the highest stability. The BestKeeper program, which used the Ct values as row data calculated the values of variation (CV) and standard deviation (SD) to analyze gene stability. Generally, the more stable genes have the lower SD and CV values. The stability of eight candidate reference genes was comprehensively evaluated using RefFinder tool. To execute, the ranking of each gene by $\Delta CT$, geNorm, NormFinder and BestKeeper was multiplied by different weighting coefficients and calculated the geometric mean to obtain an overall final ranking.
5.7 Reference gene validation. In order to further verify the reliability of the reference genes, an isoflavone synthesis gene CsMYB36 was chosen to be the target gene. CsMYB36 has a high-level expression level in the C. speciosa R2R3-MYB transcription factor family [21]. The expression pattern of CsMYB36 in different samples was investigated using the two most stable reference genes and the least stable reference gene ranked by RefFinder program. The primers of CsMYB36 (forward primer: 5’ CCACGGTTGATGAGCTTGC 3’ and reverse primer: 5’ CTGATTGCTGGACGGTTGC 3’) were designed by Primer 5.0 software and synthesized by BGI-Shenzhen. The relative expression of CsMYB36 was calculated by the $2^{-\Delta\Delta CT}$ formula [2].

5.8 Statistical analyses. Data were analyzed using IBM SPSS Statistics 19.0 software. The statistical significance was determined using Duncan's multiple range test. Values in figures marked with different lowercase were significantly different at 0.05 probability levels.

Three repetitions were performed to determine means and standard deviations (SD). Statistical analyses were performed using SPSS 24.0 software (Ehningen, Germany), and presented as the means ± SD. The statistical significance was determined using Duncan's multiple range test. "*", "**", and "***" in figures means the significantly different at 0.05, 0.01, and 0.001 probability levels, respectively.

Declarations

Funding

This work was supported by the National Science Foundation of China (Grant No. 82060685 and 81760689), the National Science Foundation of Guangxi (No. 2018GXNSFAA281287), the Program of Guangxi Key Laboratory of Zhuang and Yao Ethnic Medicine (No. 20-065-14), Youth Innovation research team project of Guangxi University of traditional Chinese medicine (2018QT001), and the Training Program for 1000 Young and Middle-aged Backbone Teachers of Guangxi Higher Education Institution in 2020.

Author statement

A preprint has been published [48].

Data Availability

The database generated for this study can be found in the transcriptome databases (SRA accession No. SRP223620).

Competing of interest

The authors declare no conflict of interest.
**Author contributions**

SY designed the research; LY, CQ, DH, and RM performed the sample preparation, total RNA extraction, RT-qPCR analysis and data collection; LY and LL performed the statistical analysis; LY wrote the manuscript; SY, RH and YT revised the manuscript. All authors read and approved the final manuscript.

**Ethics approval**

Not applicable.

**Consent to participate**

Not applicable.

**Consent for publish**

Not applicable.

**References**

1. Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R., and Mathieu, C. An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. *Methods*, Vol. 25, no. 4, pp. 386-401, 2001.

2. Kundu, A., Patel, A., and Pal, A. Defining reference genes for qPCR normalization to study biotic and abiotic stress responses in *Vigna mungo*. *Plant Cell Rep*, Vol. 32, no. 10, pp. 1647-1658, 2013.

3. Huggett, J., Dheda, K., Bustin, S., and Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun*, Vol. 6, no. 4, pp. 279-284, 2005.

4. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*, Vol. 55, no. 4, pp. 611-622, 2009.

5. Gutierrez, L., Mauriat, M., Guenin, S., Pelloux, J., Lefebvre, J.F., Louvet, R., Rusterucci, C., Moritz, T., Guerineau, F., Bellini, C., 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*, Vol. 6, no. 6, pp. 609-618, 2008.

6. Valasek, M.A., and Repa, J.J. The power of real-time PCR. *Adv Physiol Educ*, Vol. 29, no. 3, pp. 151-159, 2005.

7. Wong, M.L., and Medrano, J.F. Real-time PCR for mRNA quantitation. *Biotechniques*, Vol. 39, no. 1, pp. 75-85, 2005.
8. Lin, J., and Redies, C. Histological evidence: housekeeping genes beta-actin and GAPDH are of limited value for normalization of gene expression. *Dev Genes Evol*, Vol. 222, no. 6, pp. 369-376, 2012.

9. Joseph, J.T., Poolakkalody, N.J., and Shah, J.M. Plant reference genes for development and stress response studies. *J Biosci*, Vol. 43, no. 1, pp. 173-187, 2018.

10. Silver, N., Best, S., Jiang, J., and Thein, S.L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol Biol*, Vol. 7, no. pp. 33, 2006.

11. Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T.P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pairwise correlations. *Biotechnol Lett*, Vol. 26, no. 6, pp. 509-515, 2004.

12. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, Vol. 3, no. 7, pp. RESEARCH0034, 2002.

13. Andersen, C.L., Jensen, J.L., and Ørntoft, T.F. 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*, Vol. 64, no. 15, pp. 5245-5250, 2004.

14. Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant molecular biology*, Vol. no. pp. 2012.

15. Zeng, C., Cao, H., Dai, W., and Mei, Q. Herbal verification study of *Callerya speciosa* (Champ. ex Benth.) Schot. *Journal of Chinese Medicinal Materials*, Vol. 42, no. 6, pp. 1433-1437, 2019.

16. WANG, C.-W. Chemical constituents from roots of *Millettia speciosa*. *Chinese Traditional and Herbal Drugs*, Vol. no. pp. 1515-1520, 2014.

17. Chen, Y., Xie, Z., Wu, F., Wei, Y., Lu, S., and Zeng, H. Determination of maackiain and formononetin in root of *Millettia Speciosa* by HPLC. *World Science and Technology/Modernization of Traditional Chinese Medicine and Materia Medica*, Vol. 15, no. 2, pp. 260-264., 2013.

18. Wang, M., Wang, Q., and Zhang, B. Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.). *Gene*, Vol. 530, no. 1, pp. 44-50, 2013.

19. Sen, M.K., Hamouzova, K., Kosnarova, P., Roy, A., and Soukup, J. Identification of the most suitable reference gene for gene expression studies with development and abiotic stress response in *Bromus sterilis*. *Sci Rep*, Vol. 11, no. 1, pp. 13393, 2021.

20. Wang, B., Duan, H., Chong, P., Su, S., Shan, L., Yi, D., Wang, L., and Li, Y. Systematic selection and validation of suitable reference genes for quantitative real-time PCR normalization studies of gene
expression in *Nitraria tangutorum*. *Sci Rep*, Vol. 10, no. 1, pp. 15891, 2020.

21. Yu, L., Huang, D., Gu, J., Pan, D., Tan, Y., Huang, R., and Yao, S. Identification of Isoflavonoid Biosynthesis-Related R2R3-MYB Transcription Factors in *Callerya speciosa* (Champ. ex Benth.) Schot Using Transcriptome-Based Gene Coexpression Analysis. *Int J Genomics*, Vol. 2021, no. pp. 9939403, 2021.

22. Yu, X., Zhang, W., Zhang, Y., Zhang, X., Lang, D., and Zhang, X. The roles of methyl jasmonate to stress in plants. *Funct Plant Biol*, Vol. 46, no. 3, pp. 197-212, 2019.

23. Camara, M.C., Vandenberghe, L.P.S., Rodrigues, C., de Oliveira, J., Faulds, C., Bertrand, E., and Soccol, C.R. Current advances in gibberellic acid (GA3) production, patented technologies and potential applications. *Planta*, Vol. 248, no. 5, pp. 1049-1062, 2018.

24. Yao, S., Lan, Z., Huang, R., Tan, Y., Huang, D., Gu, J., and Pan, C. Hormonal and transcriptional analyses provides new insights into the molecular mechanisms underlying root thickening and isoflavonoid biosynthesis in *Callerya speciosa* (Champ. ex Benth.) Schot. *Sci Rep*, Vol. 11, no. 1, pp. 9, 2021.

25. Reischl, U., Youssef, M.T., Kilwinski, J., Lehn, N., Zhang, W.L., Karch, H., and Strockbine, N.A. Real-time fluorescence PCR assays for detection and characterization of Shiga toxin, intimin, and enterohemolysin genes from Shiga toxin-producing Escherichia coli. *J Clin Microbiol*, Vol. 40, no. 7, pp. 2555-2565, 2002.

26. Reischl, U., Youssef, M.T., Wolf, H., Hyytia-Trees, E., and Strockbine, N.A. Real-time fluorescence PCR assays for detection and characterization of heat-labile I and heat-stable I enterotoxin genes from enterotoxigenic Escherichia coli. *J Clin Microbiol*, Vol. 42, no. 9, pp. 4092-4100, 2004.

27. Li, Y., Liang, X., Zhou, X., Wu, Z., Yuan, L., Wang, Y., and Li, Y. Selection of Reference Genes for qRT-PCR Analysis in Medicinal Plant *Glycyrrhiza* under Abiotic Stresses and Hormonal Treatments. *Plants (Basel)*, Vol. 9, no. 11, pp. 2020.

28. Linardic, M., and Braybrook, S.A. Identification and selection of optimal reference genes for qPCR-based gene expression analysis in *Fucus distichus* under various abiotic stresses. *PLoS One*, Vol. 16, no. 4, pp. e0233249, 2021.

29. Tong, J., Hu, M., Han, B., Ji, Y., Wang, B., Liang, H., Liu, M., Wu, Z., and Liu, N. Determination of reliable reference genes for gene expression studies in Chinese chive (*Allium tuberosum*) based on the transcriptome profiling. *Sci Rep*, Vol. 11, no. 1, pp. 16558, 2021.

30. Wang, M., Ren, T., Marowa, P., Du, H., and Xu, Z. Identification and selection of reference genes for gene expression analysis by quantitative real-time PCR in Suaeda glauca's response to salinity. *Sci Rep*, Vol. 11, no. 1, pp. 8569, 2021.
31. Hieronymus, K., Dorschner, B., Schulze, F., Vora, N.L., Parker, J.S., Winkler, J.L., Rosen-Wolff, A., and Winkler, S. Validation of reference genes for whole blood gene expression analysis in cord blood of preterm and full-term neonates and peripheral blood of healthy adults. *BMC Genomics*, Vol. 22, no. 1, pp. 489, 2021.

32. Hosni, N.D., Anauate, A.C., and Boim, M.A. Reference genes for mesangial cell and podocyte qPCR gene expression studies under high-glucose and renin-angiotensin-system blocker conditions. *PLoS One*, Vol. 16, no. 7, pp. e0246227, 2021.

33. Linge, M., Mobius, M.A., Rosen-Wolff, A., and Winkler, S. Systematic analysis of candidate reference genes for gene expression analysis in hyperoxia-based mouse models of bronchopulmonary dysplasia. *Am J Physiol Lung Cell Mol Physiol*, Vol. no. pp. 2021.

34. Sule, W.F., and Oluwayelu, D.O. Real-time RT-PCR for COVID-19 diagnosis: challenges and prospects. *Pan Afr Med J*, Vol. 35, no. Suppl 2, pp. 121, 2020.

35. Tahamtan, A., and Ardebili, A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn*, Vol. 20, no. 5, pp. 453-454, 2020.

36. Ling, H., Wu, Q., Guo, J., Xu, L., and Que, Y. Comprehensive selection of reference genes for gene expression normalization in sugarcane by real time quantitative rt-PCR. *PLoS One*, Vol. 9, no. 5, pp. e97469, 2014.

37. Pan, X., and Zhang, B. Identification of Stable Reference Genes for Toxicogenomic and Gene Expression Analysis. *Methods Mol Biol*, Vol. 2326, no. pp. 67-94, 2021.

38. Fragoulis, A., Biller, K., Fragoulis, S., Lex, D., Uhlig, S., and Reiss, L.K. Reference Gene Selection for Gene Expression Analyses in Mouse Models of Acute Lung Injury. *Int J Mol Sci*, Vol. 22, no. 15, pp. 2021.

39. Tajti, J., Pal, M., and Janda, T. Validation of Reference Genes for Studying Different Abiotic Stresses in Oat (*Avena sativa L.*) by RT-qPCR. *Plants (Basel)*, Vol. 10, no. 7, pp. 2021.

40. Shen, J., Wu, Y., Jiang, Z., Xu, Y., Zheng, T., Wang, J., Cheng, T., Zhang, Q., and Pan, H. Selection and validation of appropriate reference genes for gene expression studies in *Forsythia*. *Physiology and molecular biology of plants : an international journal of functional plant biology*, Vol. 26, no. 1, pp. 173-188, 2020.

41. Liu, M., Jiang, J., Han, X., Qiao, G., and Zhuo, R. Validation of reference genes aiming accurate normalization of qRT-PCR data in *Dendrocalamus latiflorus* Munro. *PloS one*, Vol. 9, no. 2, pp. e87417, 2014.

42. Tian, C., Jiang, Q., Wang, F., Wang, G.L., Xu, Z.S., and Xiong, A.S. Selection of suitable reference genes for qPCR normalization under abiotic stresses and hormone stimuli in carrot leaves. *PloS one*, Vol. 10, no. 2, pp. e0117569, 2015.
43. Feng, K., Liu, J.X., Xing, G.M., Sun, S., Li, S., Duan, A.Q., Wang, F., Li, M.Y., Xu, Z.S., and Xiong, A.S. Selection of appropriate reference genes for RT-qPCR analysis under abiotic stress and hormone treatment in celery. PeerJ, Vol. 7, no. pp. e7925, 2019.

44. Galeano, E., Vasconcelos, T.S., Ramiro, D.A., De Martin Vde, F., and Carrer, H. Identification and validation of quantitative real-time reverse transcription PCR reference genes for gene expression analysis in teak (Tectona grandis L.f.). BMC Res Notes, Vol. 7, no. pp. 464, 2014.

45. Peirson, S.N., and Butler, J.N. Quantitative polymerase chain reaction. Methods Mol Biol, Vol. 362, no. pp. 349-362, 2007.

46. Radonić, A., Thulke, S., Mackay, I.M., Landt, O., Siegert, W., and Nitsche, A. Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun, Vol. 313, no. 4, pp. 856-862, 2004.

47. Livak, K.J., and Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, Vol. 25, no. 4, pp. 402-408, 2001.

48. Linchan Yu, Ruhong Ming, Ding Huang, Chunmei Qin, Liangbo Li, Yong Tan, Rongshao Huang and Shaochang Yao. Selection and validation of suitable reference genes for gene expression studies in Callerya speciosa (Champ. ex Benth.) Schot under different experimental conditions. Int J Genomics, DOI: 10.21203/rs.3.rs-1291888/v1.

Tables

Tables 1-4 are in the supplementary files section.

Figures

Figure 1

The different tissues and developmental stages of C. speciosa. (a-d) Stem, leaf, flower, and seed, respectively. (e-h) Root at 6, 12, 18, 30 MAG, respectively.

Figure 2

Violin chat of CT value of 8 candidate reference genes. A violin column represents the data distribution state of a sample. The curve of the violin shape is the data distribution probability curve. The more data is distributed under this value, the wider the curve. The upper and lower ends of the thin black line
represent the maximum and minimum values of the non-outliers of the data, the upper and lower edges of the thick black line represent the 3/4 digits and 1/4 digits of the data, and the white dot in the center represents the median of the data.

Figure 3

Average expression stability (M) and ranking of 8 candidate reference genes by geNorm software. (a) all samples; (b) all treatments; (c) developmental stages; (d) tissues; (e) MeJA treatment; (f) CPPU treatment; (g) GA3 treatment; (h) PP333 treatment.

Figure 4

Pairwise variation of 8 candidate reference genes by geNorm software. The pairwise variation \((V_n/V_{n+1})\) was calculated between the normalization factors \(NF_n\) and \(NF_{n+1}\), with a recommended cut off threshold of 0.15.

Figure 5

Relative expression of the \(CsMYB36\) gene in tissues (a), developmental stages (b) of \(C.\ speciosa\). MeJA treated with 200 \(\mu\)mol L\(^{-1}\) for 0, 3, 24 and 48h (c), respectively. We selected the most stable gene and the worst one identified by RefFinder. Significant difference according to Student’s t-test at \(p<0.05(*)\), \(p<0.01(**)\), \(p<0.001(***)\).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.docx
- ESM2.docx
- ESM3.docx
- Table1.pdf
- Table2.pdf
- Table3.pdf
- Table4.pdf