

SCIENTIFIC REPORTS



OPEN

Selection and Validation of Reference Genes for mRNA Expression by Quantitative Real-Time PCR Analysis in *Neolamarckia cadamba*

Tian Huang^{1,2}, Jianmei Long², Si-Wen Liu^{1,2}, Zi-Wei Yang^{1,2}, Qi-Jin Zhu^{1,2}, Xiao-Lan Zhao² & Changcao Peng^{1,2}

Neolamarckia cadamba is an economically-important fast-growing tree species in South China and Southeast Asia. As a prerequisite first step for future gene expression studies, we have identified and characterized a series of stable reference genes that can be used as controls for quantitative real time PCR (qRT-PCR) expression analysis in this study. The expression stability of 15 candidate reference genes in various tissues and mature leaves under different conditions was evaluated using four different algorithms, i.e., geNorm, NormFinder, BestKeeper and RefFinder. Our results showed that *SAMDC* was the most stable of the selected reference genes across the set of all samples, mature leaves at different photosynthetic cycles and under drought stress, whereas *RPL10A* had the most stable expression in various tissues. *PGK* and *RPS25* were considered the most suitable reference for mature leaves at different developmental stages and under cold treatment, respectively. Additionally, the gene expression profiles of sucrose transporter 4 (NcSUT4), and 9-cis-epoxycarotenoid dioxygenase 3 (NcNCED3) were used to confirm the validity of candidate reference genes. Collectively, our study is the first report to validate the optimal reference genes for normalization under various conditions in *N. cadamba* and will benefit the future discovery of gene function in this species.

Quantitative real-time polymerase chain reaction (qRT-PCR) is a powerful tool which monitors the entire PCR process in real-time. Currently, northern blot, microarray, qRT-PCR and high-throughput sequencing are the four common methods for gene expression analysis. Among them, qRT-PCR is most widely used for accurate quantification of gene expression because of its high sensitivity, accuracy, specificity and reproducibility and low cost¹⁻³. However, the accuracy of gene expression is easily affected by the factors including RNA integrity, reverse transcription reaction efficiency, cDNA quality and genomic DNA contamination⁴⁻⁷. Hence, normalization is an essential step in qRT-PCR assay⁵. Among the multiple strategies proposed⁸, the use of reference gene is the most popular method for data normalization⁹. Udvardi *et al.*¹⁰ have illustrated eleven golden rules of qRT-PCR, among which a suitable reference gene should be stably expressed in the samples under a range of given experimental conditions, so as to ensure the accuracy and reproducibility of measurement in gene expression¹¹. In contrast, inappropriate reference gene selection can lead to discrepancies in interpreting the qRT-PCR data, potentially leading to inappropriate conclusions regarding expression of target genes. Therefore, selection of appropriate reference genes is of great importance for data normalization in qRT-PCR analysis.

Numerous studies have indicated that the housekeeping genes involved in maintaining the basic metabolism of cell, such as *ACT* (*Actin*), *TUA* (*Alpha-tubulin*), *GAPDH* (*Glyceraldehyde 3-phosphate dehydrogenase*) and *18S rRNA* (*18S ribosomal RNA*) have been routinely used as internal controls in qRT-PCR and as the reference genes for standardized analysis in many plants species¹²⁻¹⁴. They were assumed to be stably expressed in various tissues at different developmental stage and under a wide range of conditions¹⁵. However, several recent reports showed that these traditional reference genes are not always stable in some specific conditions^{16,17}. In addition, there is no

¹State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, South China Agricultural University, Guangzhou, 510642, China. ²Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, 510642, China. Tian Huang and Jianmei Long contributed equally to this work. Correspondence and requests for materials should be addressed to C.P. (email: ccpeng@scau.edu.cn)

reference gene universally expressed at constant level across all species under wide range of experimental conditions¹⁸. Thus, it is essential to systematically validate the stability of potential reference genes before their use in qRT-PCR assays in any new experimental organism.

Statistical algorithms such as geNorm¹, NormFinder¹⁹ and Bestkeeper²⁰ have been specifically developed for the evaluation of candidate reference genes stability and determination of the best reference in one or several specific experimental conditions. By using these programs, reference gene validation has been successfully carried out in many plant species, including *Arabidopsis*²¹, rice¹⁵, maize²², soybean²³, radish⁹, strawberry²⁴, orchardgrass⁷, *Lycoris aurea*²⁵, banana⁵, citrus²⁶, grape²⁷, poplar²⁸, litchi²⁹ and longan³⁰. The reference genes identified in each species have been frequently used in the subsequent studies, which accelerated to explore the mechanisms of key biological processes in plant species.

Neolamarckia cadamba, a member of the Rubiaceae family, is widely distributed in South China and South Asia, and is cultivated due to its inherent economic value in multiple aspects³¹. It has been praised as “a miraculous tree” since the World Forest Congress in 1972 due to its fast growth^{32,33}, reaching heights of 17 m and trunk diameters of 25 cm within 9 years of growth under normal conditions³⁴. Therefore, it is a good choice to use *N. cadamba* for the forest rehabilitation in tropical regions. In addition, *N. cadamba* lumber is also a useful source for raw materials in paper production, furniture, building and biomass utilization. *N. cadamba* has also attracted lots of attention on the medicinal value including application in the treatment of various ailments and extraction of bioactive compounds³⁵. Recently, *N. cadamba* was used as model plant to study the xylogenesis during wood formation³¹, although these studies were primarily physiological in nature, and did not include molecular details such as transcriptional regulation of these processes. In order to explore the mechanism of wood formation and fast growth in *N. cadamba*, gene expression analysis is one of the most important steps. However, so far there is no report on the identification of reference genes for normalization in gene expression detection in *N. cadamba*.

To identify suitable reference genes for accurate quantification of target genes in *N. cadamba*, fifteen potential reference genes including Actin 1 (ACT1), Actin 7 (ACT7), Actin 11 (ACT11), Tubulin alpha 2 (TUA2), Tubulin alpha 4 (TUA4), Tubulin alpha 5 (TUA5), Elongation factor 1-alpha (EF-1- α), Ribulose biphosphate carboxylase (Rubisco), Ribosomal protein S25 (RPS25), Ribosomal protein L10A (RPL10A), Malate dehydrogenase (MDH), Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), Phosphoglycerate kinase (PGK), S-adenosylmethionine decarboxylase (SAMDC), F-Box protein (F-BOX) were selected, and their sequences retrieved from our previously generated transcriptome data³¹. The expression stability of the selected candidates was evaluated by qRT-PCR using a set of cDNAs from 22 different samples of *N. cadamba*, which included various tissues, mature leaves at different developmental stages and photosynthetic cycles, and under cold and drought treatments. Optimal reference genes were suggested for each experimental condition by using four different gene expression analysis programs. Furthermore, sucrose transporter 4 (*NcSUT4*), responsible for sucrose transportation, and 9-cis-epoxycarotenoid dioxygenase 3 (*NcNCED3*), a key enzyme involved in the synthesis of ABA, were used to confirm the reliability and effectiveness of the selected reference genes. Our study provided a list of suitable reference genes for normalization of qRT-PCR analyses, which is anticipated to facilitate the gene expression analyses in *N. cadamba*.

Results

Primers specificity, PCR efficiencies and expression profile of the candidate reference genes.

A single PCR product from each primer pair of candidate reference genes was amplified with expected size by agarose gel electrophoresis analysis (see Supplementary Fig. S1). Specific amplifications were also confirmed by melting curves with a single peak in each candidate reference genes (see Supplementary Fig. S2). Standard curves were generated using serial dilution series, and high linear correlations ($R^2 > 0.99$) were detected for in all genes (see Supplementary Table S1). The PCR amplification efficiencies for fifteen genes varied from 94.5% for *ACT1* to 116.3% for *RPS25* (see Supplementary Table S1).

The expression levels of the candidate reference genes were determined by qRT-PCR across 5 subsets of samples including different tissues, developmental stages, photosynthetic cycles, cold and drought treatments. The candidate reference genes displayed wide range of accumulation level across all the tested samples, with threshold cycle (CT) values spanning 19.48–31.79 (Fig. 1). Among them, *Rubisco* exhibited the highest expression abundance, with median CT value of 21.93. Both *GAPDH* and *MDH* showed relative high expression, with median CT values of 22.44 and 23.69, respectively. In contrast, *RPL10A* expressed lowest, with a median CT value of 30.38 (Fig. 1). Additionally, candidate genes showed distinct expression variability, among which *SAMDC* and *RPS25* displayed relative narrower CT range values than others (Fig. 1), showing that these genes expressed more stably. Importantly, however, these results revealed that none of the selected reference genes were expressed constantly in all samples tested from *N. cadamba*. Hence, it was necessary to validate suitable reference genes for normalization under different experimental conditions in *N. cadamba*.

Expression stability of candidate reference genes. *geNorm analysis.* In order to rank the candidate reference genes under tested conditions, four commonly used gene expression analysis software programs, i.e., geNorm, NormFinder, BestKeeper, and RefFinder were applied to assess the expression stabilities of fifteen candidate reference genes. In geNorm analysis, the expression stability measure M was calculated for each gene based on the non-normalized expression level Q. A cut-off value of 1.5 was recommended and set for M to evaluate gene stability. For each sample groups in our study, M value in all the candidate genes was lower than 1.5 (Fig. 2a–e). All fifteen candidate genes in each sample group set were ranked by M value. Base on this analysis, a lower M value represents a higher degree of expression stability for the selected reference gene. For tissue group, *RPL10A* and *SAMDC* were the top ranked candidates, followed by *ACT11* and *RPS25* (Fig. 2a). *SAMDC*, *PGK*, *ACT7* and *RPL10A* were the four most stable reference genes in the mature leaves at different developmental stage (Fig. 2b). Among the samples of different photosynthetic cycles in a day, *ACT7*, *SAMDC* and *ACT1* exhibited the three

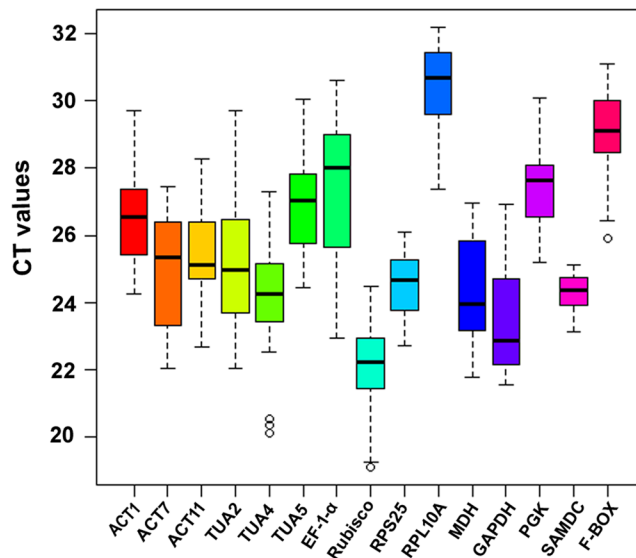


Figure 1. Distribution of threshold cycle (CT) values of 15 candidate reference genes across all 22 samples. The solid line within each box represents the 50th percentile. The lower boundary and upper boundary of each box represents the 25th and 75th percentile, respectively. The circles represent potential outliers.

most stable reference genes (Fig. 2c). *SAMDC* and *RPS25* with *M* value of about 0.3 ranked as the most stable genes in both groups of drought and cold treatment (Fig. 2d–e). When all samples were mixed and tested, most of the candidate genes had higher *M* values than their respective calculation in each group, four of which including *EF-1- α* , *TUA2*, *GAPDH* and *ACT7* even exceeded the cut-off value of 1.5. *SAMDC* and *RPS25* were the two most stable candidates with *M* value of 0.78 (Fig. 2f).

In certain conditions, multiple references might be more reliable than a single one for normalization in qRT-PCR assay¹. To determine the minimum number of reference genes for accurate normalization, GeNorm was applied through calculating the pairwise variation ($V_{n/n+1}$) between two sequential normalization factors to determine whether additional reference gene is necessary. Utilizing this software program, 0.15 was set as a cut-off for *V* value to determine whether additional reference genes be required¹. In our study, for each sample group, a pairwise variation value of $V_{2/3}$ was lower than 0.15 (Fig. 3), suggesting that only two reference genes were sufficient for normalization by qRT-PCR. Specifically, combination of *RPL10A* and *SAMDC* were the optimal reference genes for different tissues, while *PGK* and *SAMDC* were optimal for mature leaves at distinct development stages. Interestingly, combination of *RPS25* and *SAMDC* were the best choice for normalization under both drought and cold stress conditions, whereas *ACT7* and *SAMDC* were the best combination for samples under different photosynthetic cycles. However, when all the samples were analyzed simultaneously, nine reference genes were required since $V_{8/9}$ was higher than 0.15 whereas $V_{9/10}$ was below the cut-off value (Fig. 3). This suggested that no reasonable number of reference genes would be suitable for comparing expression levels between large numbers of tissues/growth conditions since the requirement to use nine references to calculate gene expression by qRT-PCR would be unreasonably time consuming. Alternatively, it may have been possible that a pairwise analysis cut-off value of 0.15 was too strict for this case, and this result also suggested that it is hard to find out universal reference genes for samples under different experimental conditions. Despite this limitation, we were able to identify suitable reference genes that exhibited specificity under a certain conditions.

NormFinder analysis. To further confirm the stability of the reference genes obtained by the geNorm software, we applied the NormFinder software for optimal normalization genes identification among the candidates. The statistic algorithm of NormFinder is different from geNorm, the former considers intra- and intergroup variations for the calculation of normalization factors (NF), which were used to estimate the stability values of each reference gene. For the sample group of different tissues, *RPL10A*, *ACT7*, *SAMDC* and *ACT11* were the four top ranked candidates (Fig. 4a). Five stable references including *F-BOX*, *PGK*, *SAMDC*, *RPL10A*, and *ACT7* were top ranked in the sample set of mature leaves at different development stages, which showed good agreement with geNorm analysis but with slight changes in the ranking order (Fig. 4b). For the sample set of different photosynthetic cycle, *ACT7* and *SAMDC* were the most stable candidates (Fig. 4c). *SAMDC* was the most suitable references under drought stress (Fig. 4d), whereas it was ranked second according to geNorm. Consistent with geNorm analysis, *Rubisco* and *RPS25* were the best candidates for normalization under cold condition (Fig. 4e). For all the sample groups, the three top ranked references were *SAMDC*, *F-BOX* and *RPL10A*, of which *SAMDC* was the best single candidate with stability value of 0.548.

BestKeeper analysis. The expression stability of the candidate reference genes were reanalyzed by another program BestKeeper. CT value and efficiencies of each primer were used as input data to calculate the standard deviation (SD), in which lower SD suggested higher stability of reference gene. *ACT7*, which had the lowest SD value

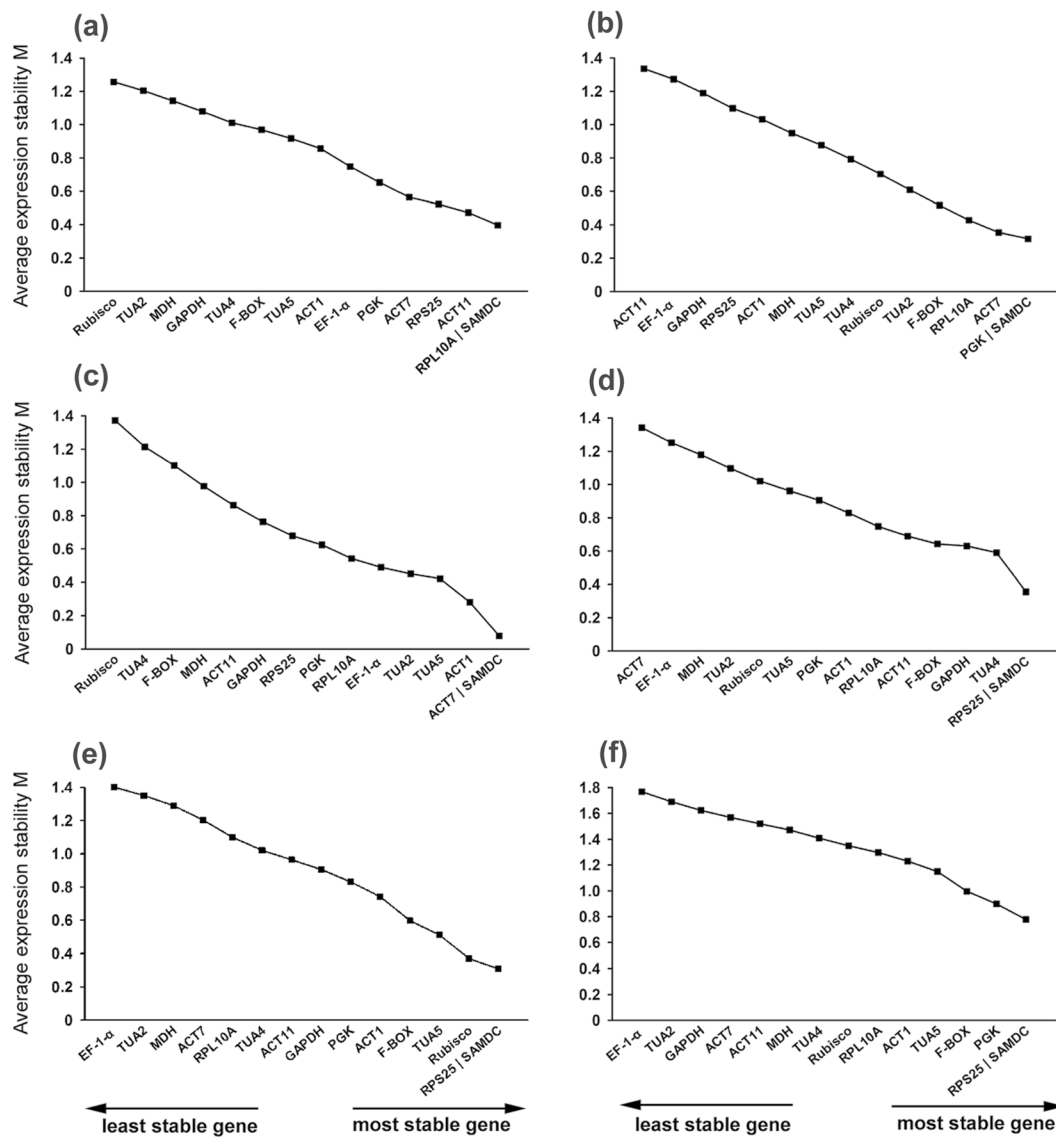


Figure 2. Average expression stability values (M) of 15 candidate reference genes calculated by geNorm. (a) Tissue; (b) developmental stage; (c) photosynthetic cycles; (d) drought treatment; (e) cold treatment; (f) all of the samples in our given conditions.

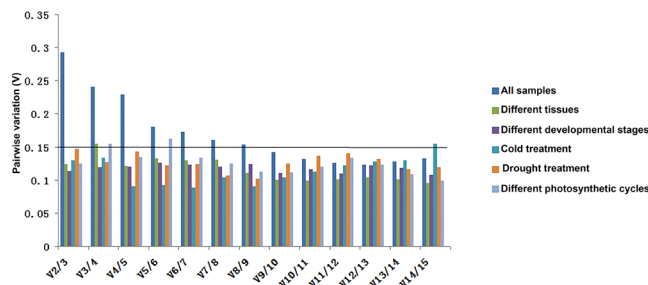


Figure 3. Determination of the optimal number of reference genes for normalization in the tested experimental conditions. geNorm was used to calculate the normalization factor (NF) from at least two genes; the variable V defines the pair-wise variation between two sequential NF values.

of 0.36, was the most stable candidate in the sample group of different tissues, whereas *MDH* was the least stable gene with the highest SD (Table 1). *PGK*, *RPL10A* and *SAMDC* had SD value of less than 0.6, which were the three top ranked references in the sample set of mature leaves at different developmental stage. *SAMDC* were the most

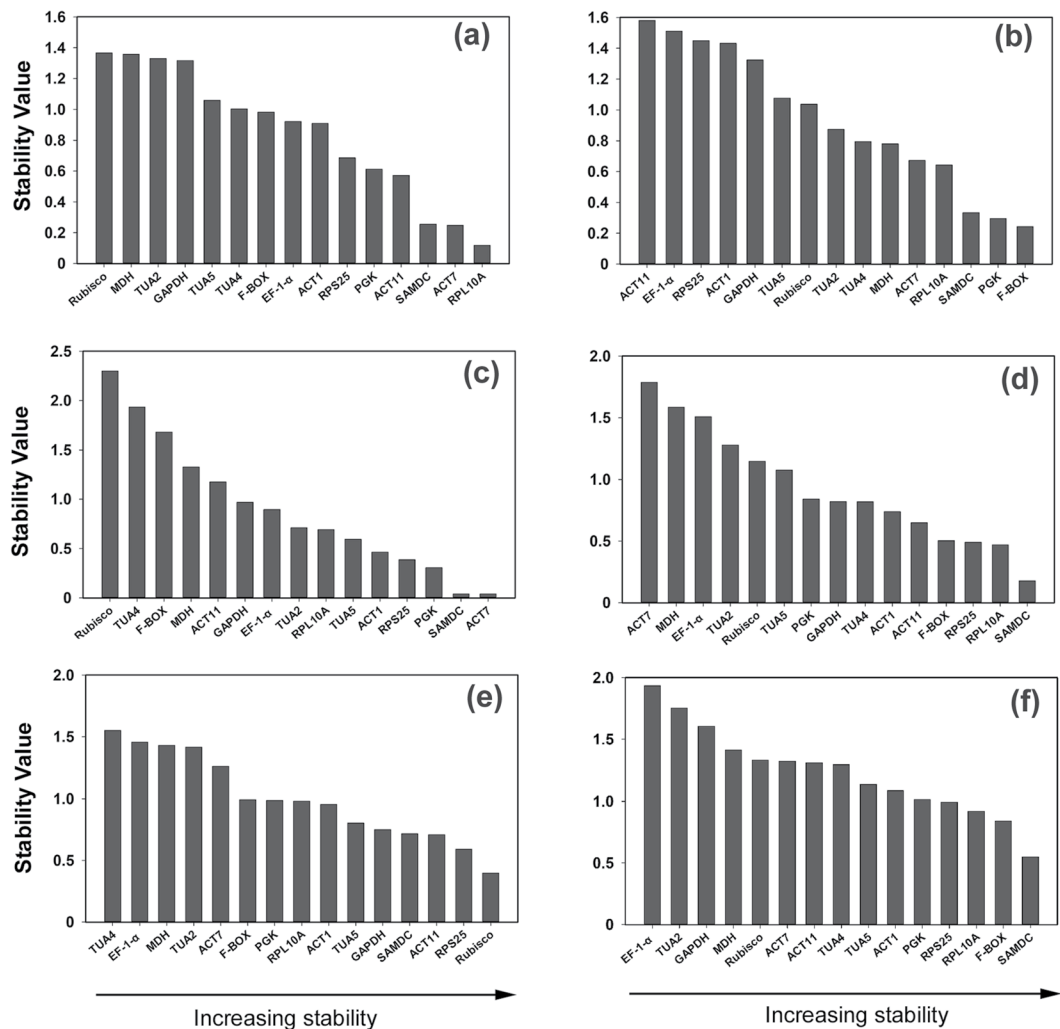


Figure 4. The stability value of 15 candidate reference genes calculated by NormFinder. The lower value indicated the higher stability of gene expression. **(a)** Tissue; **(b)** developmental stage; **(c)** photosynthetic cycles; **(d)** drought treatment; **(e)** cold treatment; **(f)** all of the samples in our given conditions.

stable gene in the sample set of different photosynthetic cycles. Similarly, for the abiotic stress, *SAMDC* was the most suitable reference gene, followed by *RPS25* and *RPL10A* under both cold and drought conditions, respectively. When considering all samples in our study, *F-BOX* and *SAMDC* were the top two ranked reference genes.

RefFinder analysis. The three different programs used for stability analysis showed different ranking orders of candidate reference genes. Therefore, to obtain a consensus result, we used RefFinder program to reorder the selected reference genes comprehensively (Table 2), and the integrated comparison of the evaluation by four programs was shown in Supplementary Table S2. The ranking order by RefFinder showed that the top two ranked reference gene were in agreement with the list yielded by geNorm under all given experimental conditions. *Rubisco* was ranked at the bottom for the group of different tissues and photosynthetic cycle; *EF-1 α* was at the bottom position for cold treatment and all samples tested. *ACT11* and *ACT7* were the least stable genes in the samples of different developmental stage and drought stress, respectively.

Validation of reference genes. In order to verify the stability of reference gene in this study, we analyzed the expression levels of sucrose transporter 4 (*NcSUT4*) and 9-cis-epoxycarotenoid dioxygenase 3 (*NcNCED3*) using the selected reference genes. Sucrose transporters are responsible for phloem loading, transport and unloading of sucrose from source to sink³⁶. *NCED* genes encode key enzymes for the synthesis of ABA, and are induced under drought stress^{37,38}. The transcript levels of *NcSUT4* were quantified in the samples of different developmental stages and different tissues, using two stable reference genes and two unstable reference genes as endogenous controls, respectively (Fig. 5a,b). As shown in Fig. 5a, when using the two stable reference genes (*SAMDC* or *PGK*) for normalization, the expression level of *NcSUT4* was increased in the mature leaves from March to December, with high peak in June. However, when using the least stable reference genes *ACT7* and *EF-1 α* as normalization factors, *NcSUT4* exhibited totally different expression pattern, with gradually decrease across different development stages of mature leaves. In various tissues, relative higher transcript abundance of

Ranking	Tissue	Developmental stage	photosynthetic cycle	Drought treatment	Cold treatment	Total
1	ACT7 (1.37 ± 0.36)	PGK (1.83 ± 0.51)	SAMDC (0.35 ± 0.09)	RPL10A (0.61 ± 0.19)	RPS25 (0.67 ± 0.17)	SAMDC (1.43 ± 0.35)
2	RPL10A (1.43 ± 0.45)	RPL10A (1.91 ± 0.58)	ACT7 (0.64 ± 0.14)	SAMDC (1.04 ± 0.25)	SAMDC (1.35 ± 0.33)	F-BOX (2.71 ± 0.79)
3	SAMDC (1.53 ± 0.37)	SAMDC (2.1 ± 0.51)	PGK (1.32 ± 0.37)	RPS25 (1.38 ± 0.33)	TUA5 (1.81 ± 0.51)	RPS25 (2.78 ± 0.68)
4	EF-1-α (1.96 ± 0.58)	F-BOX (2.67 ± 0.78)	RPS25 (1.57 ± 0.39)	ACT1 (1.41 ± 0.38)	Rubisco (1.87 ± 0.41)	PGK (2.85 ± 0.78)
5	PGK (2 ± 0.56)	ACT7 (2.79 ± 0.68)	ACT1 (1.6 ± 0.41)	ACT11 (1.75 ± 0.46)	RPL10A (2.6 ± A0.79)	RPL10A (3.33 ± 1.01)
6	F-BOX (2.11 ± 0.62)	TUA5 (2.94 ± 0.75)	TUA5 (1.6 ± 0.41)	F-BOX (1.79 ± 0.52)	ACT11 (2.64 ± 0.64)	TUA5 (3.58 ± 0.96)
7	TUA5 (2.17 ± 0.6)	TUA2 (3.09 ± 0.76)	RPL10A (1.94 ± 0.55)	TUA5 (1.97 ± 0.53)	F-BOX (2.94 ± 0.87)	ACT11 (3.86 ± 0.98)
8	ACT11 (2.26 ± 0.57)	RPS25 (3.09 ± 0.76)	TUA2 (2.13 ± 0.49)	TUA4 (2.37 ± 0.59)	TUA4 (3.1 ± 0.77)	ACT1 (4 ± 1.06)
9	RPS25 (2.46 ± 0.6)	MDH (3.45 ± 0.83)	EF-1-α (2.54 ± 0.61)	PGK (2.52 ± 0.67)	ACT1 (3.26 ± 0.88)	TUA4 (4.41 ± 1.07)
10	ACT1 (2.54 ± 0.66)	TUA4 (3.52 ± 0.86)	ACT11 (3.11 ± 0.79)	MDH (3.76 ± 0.91)	PGK (3.61 ± 0.99)	MDH (4.94 ± 1.2)
11	TUA4 (3.22 ± 0.77)	EF-1-α (3.97 ± 1.11)	GAPDH (3.17 ± 0.8)	TUA2 (4.97 ± 1.36)	ACT7 (3.76 ± 0.95)	Rubisco (5.04 ± 1.1)
12	TUA2 (4.48 ± 1.09)	ACT11 (4.46 ± 1.17)	MDH (4.11 ± 1.02)	EF-1-α (5.12 ± 1.37)	EF-1-α (3.84 ± 1.06)	ACT7 (5.26 ± 1.32)
13	GAPDH (4.69 ± 1.09)	ACT1 (4.74 ± 1.3)	F-BOX (4.28 ± 1.21)	ACT7 (5.33 ± 1.36)	GAPDH (4.44 ± 1.02)	GAPDH (5.77 ± 1.35)
14	Rubisco (5.2 ± 1.16)	Rubisco (5.18 ± 1.13)	TUA4 (5.89 ± 1.31)	Rubisco (6.16 ± 1.32)	TUA2 (5.07 ± 1.32)	TUA2 (6.21 ± 1.57)
15	MDH (5.28 ± 1.24)	GAPDH (5.31 ± 1.21)	Rubisco (7.17 ± 1.56)	GAPDH (6.23 ± 1.48)	MDH (6.25 ± 1.56)	EF-1-α (6.44 ± 1.77)

Table 1. Expression stability values for *Neolamarckia cadamba* candidate reference genes calculated by BestKeeper. Note: Fifteen candidate reference genes are evaluated by the lowest values of the coefficient of variance (CV) and standard deviation (SD). The number in the bracket indicated that CV ± SD.

Ranking	Tissue	Developmental stage	Photosynthetic cycle	Drought treatment	Cold treatment	Total
1	RPL10A	PGK	SAMDC	SAMDC	RPS25	SAMDC
2	SAMDC	SAMDC	ACT7	RPS25	SAMDC	RPS25
3	ACT7	F-BOX	ACT1	RPL10A	Rubisco	F-BOX
4	ACT11	RPL10A	PGK	F-BOX	TUA5	PGK
5	PGK	ACT7	TUA5	ACT11	ACT11	RPL10A
6	RPS25	TUA2	RPS25	ACT1	F-BOX	TUA5
7	EF-1-α	MDH	TUA2	TUA4	GAPDH	ACT1
8	ACT1	TUA4	RPL10A	GAPDH	ACT1	TUA4
9	F-BOX	TUA5	EF-1-α	TUA5	RPL10A	ACT11
10	TUA5	Rubisco	GAPDH	PGK	PGK	Rubisco
11	TUA4	RPS25	ACT11	Rubisco	TUA4	MDH
12	GAPDH	GAPDH	MDH	TUA2	ACT7	ACT7
13	TUA2	ACT1	F-BOX	MDH	TUA2	GAPDH
14	MDH	EF-1-α	TUA4	EF-1-α	MDH	TUA2
15	Rubisco	ACT11	Rubisco	ACT7	EF-1-α	EF-1-α

Table 2. Comprehensive ranking of 15 candidate reference genes integrated by RefFinder.

NcSUT4 was found in root and phloem when using the two most stable reference genes (*SAMDC* and *RPL10A*) singly or in combination. Though the expression pattern was similar when the unstable gene *Rubisco* served as internal control, the relative expression of *NcSUT4* was much higher in root, xylem and phloem than that using the most stable references. In contrast, the highest expression of *NcSUT4* was found in xylem rather than phloem when using another unstable endogenous gene *GAPDH* (Fig. 5b). The relative transcript abundance of *NcNCED3* under cold and drought stress condition was also calculated in the way similar with *NcSUT4* (Fig. 5c,d). Under cold treatment, the expression level of *NcNCED3* was not affected significantly when normalized by the most stable reference genes *SAMDC* and *RPS25*, whereas the transcript abundance was overestimated when using the least stable references *TUA2* and *EF-1-α*. With drought treatment, the relative expression level of *NcNCED3* increased to two peaks at the PEG concentration of 5% and 20%. On the contrary, normalization by the two least stable genes (*GAPDH*, *Rubisco*) led to the nearly unchanged pattern of *NcNCED3* expression level. Taken together, the substantial divergence was found in the expression levels of *NcSUT4* and *NcNCED3* when normalized by the most stable and least stable reference genes.

Discussion

Gene expression analysis has become an important aspect in the functional investigation of genes during the growth and development of various living organisms. The accuracy of genes relative expression mainly depends on the reference genes. Therefore, the selection of inappropriate reference gene can give rise to the erroneous data and misinterpretation of experimental results³⁹. An ideal reference gene for normalization in qRT-PCR analysis should be stably expressed at moderate level in a variety of test samples and has similar amplification efficiency

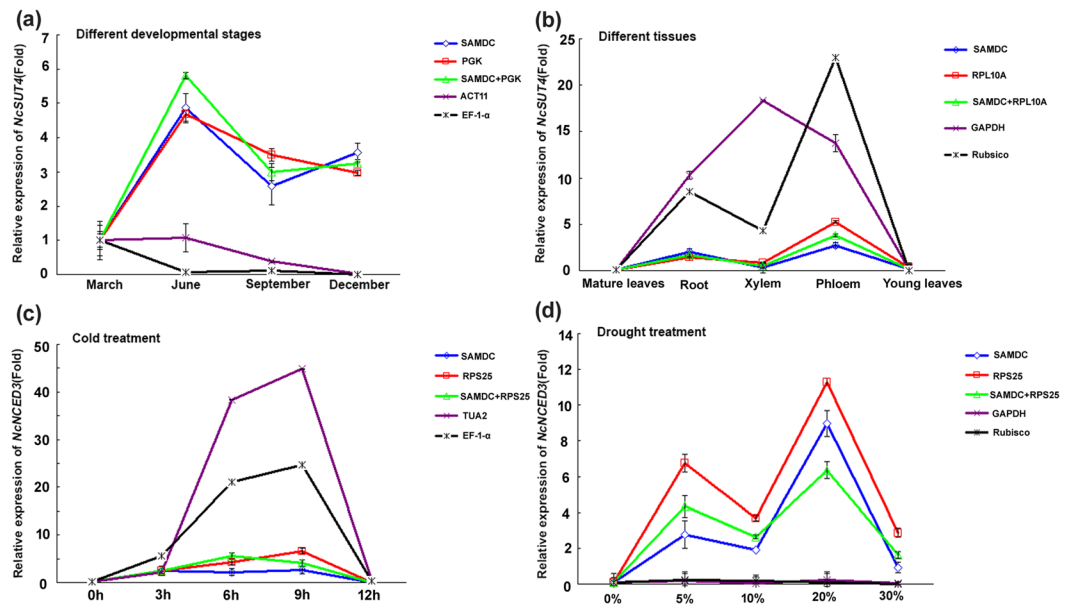


Figure 5. Relative quantification of *NcSUT4* and *NcNCED3* expression levels using the most and least stable reference genes for normalization in the given experimental conditions. **(a)** Expression level of *NcSUT4* at different developmental stages. Most stable reference genes (*SAMDC*, *PGK*) and least stable reference genes (*ACT11*, *EF-1- α*) were used, respectively; **(b)** expression level of *NcSUT4* in different tissues. Most stable reference genes (*SAMDC*, *RPL10A*) and least stable reference genes (*GAPDH*, *Rubisco*) were used, respectively; **(c)** expression level of *NcNCED3* under cold treatment. Most stable reference genes (*SAMDC*, *RPS25*) and least stable reference genes (*TUA2*, *EF-1- α*) were used, respectively; **(d)** expression level of *NcNCED3* under drought treatment. Most stable reference genes (*SAMDC*, *RPS25*) and least stable reference genes (*GAPDH*, *Rubisco*) recommended by BestKeeper were used, respectively. The error bars represent the mean of three biological replicates \pm SD.

to target genes^{25,40}. However, in our analysis of potential candidate reference genes in *Neolamarckia cadamba*, we determined that there is no single reference gene that exhibited constant expression level in all the samples of various tissues and under different experimental conditions, similar with the findings in cotton⁴¹. Even though housekeeping genes have been commonly used for normalization in qRT-PCR assay, the expression levels of several genes like *ACT*, *18S rRNA* and *GAPDH* were not always constant in the given conditions. Therefore, it is necessary to assess the expression stability of candidate reference genes systematically when performing qRT-PCR analysis for different tissue and experimental conditions.

In this study, 15 candidate genes were selected for identifying the best reference among 22 samples under different experimental conditions and a variety of various tissues from *N. cadamba*. By the use of four computational gene expression analysis programs, i.e. geNorm, NormFinder, BestKeeper and RefFinder, the expression stability of the selected candidate reference genes were evaluated. The statistical algorithms to assess the stability of references of three programs vary greatly. geNorm estimates the average pairwise variation of a reference gene versus all other genes among the given samples^{1,11}. However, for NormFinder analysis, the expression stability value was calculated based on intra- and inter-group variation¹⁹. In BestKeeper, the CV and SD values determine the stability ranking of a candidate reference gene²⁰. Owing to the different algorithms of the three software programs used in this study, the ranking order of the selected candidates showed variable degrees of agreement (Figs 2, 4, Table 1 and Supplementary Table S2). The reference genes, *SAMDC* exhibited the most stable expression based on analysis by all three programs across all 22 samples under the five given conditions (see Supplementary Table S2). In addition, *RPL10A* was one of top two most stable reference genes in various tissues by all three programs. However, *PGK*, one top ranked gene by geNorm and BestKeeper in the sample group of different developmental stages, was ranked at a medium position in NormFinder (Fig. 4). Similarly, under cold stress condition, expression of *RPS25* was the least stable according to NormFinder, whereas it was the most stable gene by geNorm and BestKeeper. Increasing evidences have demonstrated that two or three algorithms were widely used to evaluate expression stability of the reference genes for normalizing mRNA and miRNA expression^{5,24,41,42}. Specially, we also noticed that the unstable reference genes identified by three programs were similar. For example, *TUA2*, *GAPDH*, *Rubisco* and *MDH* were expressed unstably in different tissues by all three algorithms. Furthermore, a web-based tool RefFinder was employed to integrate and generate the comprehensive ranking of the candidate reference genes based on the geometric mean of the weights of every gene calculating by each program⁴³. The result produced by RefFinder provides us the overall ranking order, which has been widely used in the previous studies on exploring the suitable reference genes in the certain conditions^{7,9,18,24}.

Our study showed that *SAMDC* was the top ranked gene when all samples were analyzed, since it was recommended by all three programs as the most stable reference gene (Figs 2, 4, and Table 2). Interestingly, the integrated ranking order generated by RefFinder suggested that *SAMDC* was also the optimal reference gene

in five independent sample sets, including the samples under cold and drought stresses, which was similar with the result of *Brachypodium distachyon* plant grown under various stresses conditions⁴⁴. *SAMDC*, encoding S-adenosylmethionine decarboxylase, is a rate-limiting enzyme in the biosynthesis of the polyamines spermidine and spermine^{45,46}. *SAMDC* expressed ubiquitously in different plant organs in *Brachypodium*⁴⁴, and highly induced under various abiotic stress treatments in rice⁴⁷, indicating it might play essential roles in plant growth and tolerance response. *SAMDC* has never been used as internal reference gene for normalization *N. cadamba*, and thereby would be considered as a novel reference candidate. However, the superior stability performance of *SAMDC* was not found in some species under certain conditions. For example, *SAMDC* was considered one of the least stable reference genes in orchardgrass under various abiotic conditions and different tissues of banana^{3,7}. In addition to *SAMDC*, *PGK* was also found to be the best stable candidate as reference gene in the leaves at different developmental stages in our study. *PGK*, encoding phosphoglycerate kinase, plays pivotal roles in the glycolytic pathway⁴⁸. *PGK* was found to be one of the superior reference genes for normalization in chrysanthemum of cross-ploidy levels and in tomato leaves during light stress^{49,50}, suggesting that *PGK* had potential to be a suitable reference in certain experimental conditions. Likewise, our current study showed that *PGK* was the most stable for different stages, which will be a new reference candidate in *N. cadamba*.

Previous reports have demonstrated that using multiple reference genes rather than a single one for was more robust and accurate for normalizing qRT-PCR data^{1,49}. Accordingly, more and more studies in various species were tended to apply multiple references^{51–53}. In our study, the combination of two reference candidates was sufficient for normalization in the five different experimental sets (Fig. 2). Thus, additional genes were not required for gene expression normalization. However, it was not always to use multiple reference genes due to the cost, and also it would be dependent on the variability of the selected reference genes.

The housekeeping genes have long been used as internal control to quantify gene expression for their requirement in maintaining the basal cellular functions. They were supposed to be highly conserved and stably expressed under different experimental conditions and various developmental stages^{15,54}. *ACTIN*, *GAPDH*, *TUB* and *EF*, considered as classical housekeeping genes, have been commonly used as reference genes in qRT-PCR analysis in the model plants, as well as the non-model species^{55–58}. Unfortunately, increasing studies demonstrated the housekeeping genes would variably expressed in many species under given experimental sets, which would not be suitable to serve as reference genes for normalization. *ACT* and *GAPDH*, for instance, were not appropriate candidate reference under many experimental conditions in papaya¹⁸, similar with the result of chicory in seedlings and cell cultures⁴². In the present study, three *ACT* genes were chosen to evaluate their stabilities in 22 samples from *N. cadamba*. When all samples were tested, *ACT7* was ranked at the bottom position while *ACT1* was ranked intermediate position, contrast with the result in samples at different developmental stages with top-ranked *ACT7* and bottom-ranked *ACT1* (Table 2). For the separate sample groups, *ACT7* and *ACT1* ranked among the top three most stable reference genes under different photosynthesis cycle, while *ACT7* had poor performance under drought stress condition. Liu *et al.*²⁶ have proposed that genes with similar function might have various performances in expression stability. In our work, the ribosomal protein *RPS25* and *RPL10A* were considered as the appropriate reference genes when all sample were tested, whereas they performed variably with top ranking in *RPL10A* and middle-bottom ranking in *RPS25*. Another classical housekeeping gene *TUA*, encoding one of the major components of microtubules⁵⁹, had a moderate or least stable level in most of our given experimental sets, but considered as the most stable candidate in poplar⁶⁰ and least stable reference in *Caragana korshinskii* Kom.⁶¹ These results further confirmed that there is no universal reference gene across all species under various experimental conditions. Accordingly, stability assessment of the traditional reference genes is necessary prior to use in qRT-PCR.

To verify the suitability of the reference genes identified in this study, the relative expression levels of *NcSUT4* and *NcNCED3* under different experimental conditions have been detected. The expression profile of *NcSUT4* and *NcNCED3* were normalized by best suitable genes as well as least unstable genes recommended by RefFinder in their own experimental sets. The results revealed that expression pattern of *NcSUT4* and *NcNCED3* were obviously over- or underestimated when using the unstable genes for normalization (Fig. 5), suggesting that selection of appropriate stable reference genes is critically important for the correct normalization for qRT-PCR data in *N. cadamba*.

In conclusion, our study presented the systematic evaluation of a set of candidate reference genes as normalization factors in qRT-PCR analysis in the samples under a wide range of experimental conditions in *N. cadamba*. Four widely used programs (geNorm, NormFinder, BestKeeper and RefFinder) were applied to estimate the expression stability of the selected candidate genes. The results of comprehensive ranking order showed that *SAMDC* displayed highest stability across the set of all samples, mature leaves at different photosynthetic cycles and under drought stress, whereas *RPL10A* had the most stable expression in various tissues. *PGK* and *RPS25* were the most stable in mature leaves at different developmental stages and under cold conditions, respectively. The expression analysis of *NcSUT4* and *NcNCED3* emphasized the importance of suitable reference gene to get accurate and reliable quantitation results by qRT-PCR. For the first time in *N. cadamba*, our study provided the optimal reference under different experimental conditions, which would be of great importance in further analysis of gene expression and facilitate the understanding of underlying mechanisms in the aspects of development and stress response.

Materials and Methods

Plant materials. Fresh young leaves, mature leaves, phloem, xylem, roots of 4–5 year-old *Neolamarckia cadamba* from South China Agricultural University Yuejin North experimental field were harvested. For samples of developmental stages, mature leaves were collected on March, June, September and December. For samples of different photosynthetic cycle, mature leaves were sampled at 7:00, 13:00 and 19:00 on October 28, 2016. For drought treatment samples, the seedlings after seven days of transplantation from tissue culture were treated with 0%, 5%, 10%, 20% and 30% concentration of PEG 6000 and the mature leaves were collected. Meanwhile,

the seedlings were also transferred to 4 °C for chilling stress, and the mature leaves were taken after 0, 3, 6, 9 and 12 h. Each sample was collected with three biological replicates. Information of all five sample sets mentioned above is summarized in Supplementary Table S3. Samples were frozen immediately in liquid nitrogen and stored at −80 °C until RNA extraction.

RNA isolation, quality control and cDNA synthesis. Total RNA was extracted from the samples using OmniPlant RNA Kit (CWBIO) with DNase I treatment to avoid genomic DNA containment. The RNA purity and integrity were estimated by calculating their A260/280 and A260/A230 absorbance ratios and agarose gel electrophoresis analysis. cDNA was synthesized from 0.5 µg of total RNA by Hiscript II Reverse Transcriptase kit (Vazyme) according to the manufacturer's instruction. The cDNA were diluted 30-fold for subsequent qRT-PCR analysis.

Selection of candidate reference gene and primers design. Fifteen candidate reference genes, including Actin 1 (ACT1), Actin 7 (ACT7), Actin 11 (ACT11), Tubulin alpha 2 (TUA2), Tubulin alpha 4 (TUA4), Tubulin alpha 5 (TUA5), Elongation factor 1-alpha (EF-1-α), Ribulose biphosphate carboxylase (Rubisco), Ribosomal protein S25 (RPS25), Ribosomal protein L10A (RPL10A), Malate dehydrogenase (MDH), Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), Phosphoglycerate kinase (PGK), S-adenosylmethionine decarboxylase (SAMDC), F-Box protein (F-BOX) were selected, based on a preliminary *in silico* evaluation of gene expression stability using *in house* RNA-Seq libraries from various tissues of *N. cadamba* including leaves, xylem, cambium and phloem (unpublished data). Although gene expression study is common on plant functional genomic research, *N. cadamba* is still a less studied species. To date, only two studies on gene expression have been carried out in *N. cadamba*, and cyclophilin was used as the internal reference for normalization^{31,33}. However, according to the transcriptome data of leaves, xylem, cambium and phloem, the expression of cyclophilin varied greatly in these samples (unpublished data). Therefore, cyclophilin was not selected for further evaluation of expression stability in our study. Due to lacking the genome sequence of *N. cadamba*, the primers were designed based on sequences extracted from the released transcriptome data³¹. Primers for qRT-PCR were designed using web-based Primer-BLAST tool in NCBI (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) with default parameters. Gene ID, primers and the expected length of each gene were indicated in Supplementary Table S4. All the primers were designed across introns except *F-BOX* and *RPL10A*. To check the specificity of the primers, general PCR was carried out and the products was verified by electrophoresis on 1% agarose gels.

Quantitative Real-time PCR. qRT-PCR was performed on LightCycler480 (Roche Molecular Biochemicals, Mannheim, Germany) with optical 96-well plate. 10 µL of the reaction mix containing 5 µL AceQ qPCR SYBR GREEN PCR Master Mix, 0.5 µL diluted cDNA template and 0.5 µL each primer were added into each well. The thermal cycling profile was recommended by the manufacturer: 95 °C for 10 min, 40 cycles of 95 °C for 15 s, 60 °C for 30 s. To confirm the specificity of the primers, melting curves were included after amplification. All the samples for qRT-PCR analysis were conducted with three biological replicates each comprising three technical replicates. To calculate the gene-specific PCR efficiency (E) and correlation coefficient (R²) for each gene, standard curves were generated using tenfold dilution series from the mixed cDNA.

Analysis of the stabilities of reference genes. Three software tools including geNorm¹ (version 3.5, <http://medgen.ugent.be/~jvdesomp/genorm/>), NormFinder¹⁹ (<http://www.mdl.dk/publicationsnormfinder.htm>) and BestKeeper²⁰ (<http://www.gene-quantification.de/bestkeeper.html>) were used to evaluate the stability of the 15 selected candidate reference genes across all the experimental sets. Expression levels of the candidate reference genes were determined by cycle threshold (CT) values. All the procedures of statistical analyses by the above packages were conducted according to the manufacturer's instructions.

The geNorm software calculates of gene expression stability value (M), which lower M value suggested higher gene's expression stability. Moreover, geNorm also generated a pairwise stability measurement, which can be used to evaluate the suitable number of reference genes for normalization. NormFinder focuses on finding the two genes with the least intra- and inter-group expression variation or the most stable reference gene in intra-group expression variation. The principle of the BestKeeper program is to determine the stability of the reference gene by using the two parameters to calculate the correlation coefficient (CV) and the standard coefficient of variation (SD). Finally, the stability rankings of the reference genes from the three different algorithms were integrated, generating an overall ranking according to the geometric mean by RefFinder⁴³.

Validation of reference genes. To evaluated the reliability of the reference genes ranked by three algorithms, we analyzed the expression profiles of two target genes *9-Cis-epoxycarotenoid dioxygenase 3 (NcNCED3)*, and *Sucrose transporter 4 (NcSUT4)* under the two most stable and two least stable reference genes. The expression level of *NcNCED3* was determined under drought and cold treatment with forward primer TTTCGCGATAACTGAGAACT and reverse primer ACCAAACCTCGAACTTTGT, while *NcSUT4* was quantified in various tissues and at different developmental stages of mature leaves with forward primer GGCTTTGTTTATAGGTTT and reverse primer CTCGAGTCCTCCTGTGATC. The E^{−ΔΔCT} method was used to calculate the relative expression levels⁶².

References

- Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome. Biol.* **3**, (research0034). 1–0034.11 (2002).
- Rebrikov, D. & Trofimov, D. Y. Real-time PCR: a review of approaches to data analysis. *Appl. Biochem. Micro.* **42**, 455–463 (2006).
- Phillips, M. A., D'Auria, J. C., Luck, K. & Gershenzon, J. Evaluation of candidate reference genes for real-time quantitative PCR of plant samples using purified cDNA as template. *Plant Mol. Biol. Rep.* **27**, 407–416 (2009).
- Bustin, S. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* **29**, 23–39 (2002).

5. Chen, L. *et al.* Validation of reference genes for RT-qPCR studies of gene expression in banana fruit under different experimental conditions. *Planta* **234**, 377–390 (2011).
6. Hao, X. *et al.* Identification and evaluation of reliable reference genes for quantitative real-time PCR analysis in tea plant (*Camellia sinensis* (L.) O. Kuntze). *Int. J. Mol. Sci.* **15**, 22155–22172 (2014).
7. Huang, L. *et al.* Reference gene selection for quantitative real-time reverse-transcriptase PCR in orchardgrass subjected to various abiotic stresses. *Gene* **553**, 158–165 (2014).
8. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* **6**, 279 (2005).
9. Duan, M. *et al.* Identification of optimal reference genes for expression analysis in Radish (*Raphanus sativus* L.) and its relatives based on expression stability. *Front. Plant Sci.* **8**, 1605 (2017).
10. Udvardi, M. K., Czechowski, T. & Scheible, W. R. Eleven golden rules of quantitative RT-PCR. *Plant Cell* **20**, 1736–1737 (2008).
11. Kou, S. J. *et al.* Selection and validation of suitable reference genes for miRNA expression normalization by quantitative RT-PCR in citrus somatic embryogenic and adult tissues. *Plant Cell Rep.* **31**, 2151–2163 (2012).
12. Pillitteri, L. J., Lovatt, C. J. & Walling, L. L. Isolation and characterization of a *TERMINAL FLOWER* homolog and its correlation with juvenility in citrus. *Plant Physiol.* **135**, 1540–1551 (2004).
13. Gamuyao, R. *et al.* The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* **488**, 535–539 (2012).
14. Saraiva, K. D., Fernandes de Melo, D., Morais, V. D., Vasconcelos, I. M. & Costa, J. H. Selection of suitable soybean *EF1 α* genes as internal controls for real-time PCR analyses of tissues during plant development and under stress conditions. *Plant Cell Rep.* **33**, 1453–1465 (2014).
15. Jain, M., Nijhawan, A., Tyagi, A. K. & Khurana, J. P. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* **345**, 646–651 (2006).
16. Gutierrez, L. *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.* **6**, 609–618 (2008).
17. Maroufi, A., Van Bockstaele, E. & De Loose, M. Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. *BMC Mol. Biol.* **11**, 15 (2010).
18. Zhu, X. *et al.* Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. *PLoS One* **7**, e44405 (2012).
19. Andersen, C. L., Jensen, J. L. & 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.* **64**, 5245–5250 (2004).
20. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
21. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W. R. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* **139**, 5–17 (2005).
22. Manoli, A., Sturaro, A., Trevisan, S., Quaggiotti, S. & Nonis, A. Evaluation of candidate reference genes for qPCR in maize. *J. Plant Physiol.* **169**, 807–815 (2012).
23. Hu, R., Fan, C., Li, H., Zhang, Q. & Fu, Y. F. Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol. Biol.* **10**, 93 (2009).
24. Galli, V. *et al.* Validation of reference genes for accurate normalization of gene expression for real time-quantitative PCR in strawberry fruits using different cultivars and osmotic stresses. *Gene* **554**, 205–214 (2015).
25. Ma, R., Xu, S., Zhao, Y., Xia, B. & Wang, R. Selection and validation of appropriate reference genes for quantitative Real-Time PCR analysis of gene expression in *Lycoris aurea*. *Front. Plant Sci.* **7**, 536 (2016).
26. Liu, Z. *et al.* Selection and validation of suitable reference genes for mRNA qRT-PCR analysis using somatic embryogenic cultures, floral and vegetative tissues in citrus. *Plant Cell Tiss. Org. Cult.* **113**, 469–481 (2013).
27. Reid, K. E., Olsson, N., Schlosser, J., Peng, F. & Lund, S. T. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biol.* **6**, 27 (2006).
28. Xu, M., Zhang, B., Su, X., Zhang, S. & Huang, M. Reference gene selection for quantitative real-time polymerase chain reaction in *Populus*. *Anal. Biochem.* **408**, 337–339 (2011).
29. Zhong, H. Y. *et al.* Selection of reliable reference genes for expression studies by reverse transcription quantitative real-time PCR in litchi under different experimental conditions. *Plant Cell Rep.* **30**, 641–653 (2011).
30. Lin, Y. L. & Lai, Z. X. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. *Plant Sci.* **178**, 359–365 (2010).
31. Ouyang, K. *et al.* Transcriptomic analysis of multipurpose timber yielding tree *Neolamarckia cadamba* during xylogenesis using RNA-seq. *PLoS One* **11**, e0159407, <https://doi.org/10.1371/journal.pone.0159407> (2016).
32. Ouyang, K., Liu, M., Pian, R., Liu, S. & Chen, X. Isolation and analysis of α -expansin genes in the tree *Anthocephalus chinensis* (Rubiaceae). *Genet. Mol. Res.* **12**, 1061–1073 (2013).
33. Li, J. *et al.* Functional identification of an *EXPA* gene (*NcEXPA8*) isolated from the tree *Neolamarckia cadamba*. *Biotechnol. Biotech. Eq.* **31**, 1116–1125 (2017).
34. Zayed, M. Z., Zaki, M. A., Ahmad, F. B., Ho, W. S. & Pang, S. L. Comparison of mimosine content and nutritive values of *Neolamarckia cadamba* and *Leucaena leucocephala* with *Medicago sativa* as forage quality index. *Int. J. Sci. Technol. Res.* **3**, 146–150 (2014).
35. Dubey, A., Nayak, S. & Goupale, D. *Anthocephalus cadamba*: A review. *Phcog. J.* **2**, 71–76 (2011).
36. Kuhn, C. & Grof, C. P. Sucrose transporters of higher plants. *Curr. Opin. Plant Biol.* **13**, 288–298 (2010).
37. Peleg, Z. & Blumwald, E. Hormone balance and abiotic stress tolerance in crop plants. *Curr. Opin. Plant Biol.* **14**, 290–295 (2011).
38. Pedrosa, A. M. *et al.* Effect of overexpression of citrus 9-cis-epoxycarotenoid dioxygenase 3 (*CsNCED3*) on the physiological response to drought stress in transgenic tobacco. *Genet. Mol. Res.* **16**, gmr16019292 (2017).
39. Dheda, K. *et al.* The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal. Biochem.* **344**, 141–143 (2005).
40. Yang, Q. *et al.* Evaluation and validation of the suitable control genes for quantitative PCR studies in plasma DNA for non-invasive prenatal diagnosis. *Int. J. Mol. Med.* **34**, 1681–1687 (2014).
41. Artico, S., Nardeli, S. M., Brillhante, O., Grossi-de-Sa, M. F. & Alves-Ferreira, M. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biol.* **10**, 49 (2010).
42. Delporte, M., Legrand, G., Hilbert, J. L. & Gagneul, D. Selection and validation of reference genes for quantitative real-time PCR analysis of gene expression in *Cichorium intybus*. *Front. Plant Sci.* **6**, 651, <https://doi.org/10.3389/fpls.2015.00651> (2015).
43. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* **80**, 75–84 (2012).
44. Hong, S. Y., Seo, P. J., Yang, M. S., Xiang, F. & Park, C. M. Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR. *BMC Plant Biol.* **8**, 112 (2008).
45. Hu, W. W., Gong, H. & Pua, E. C. Modulation of *SAMDC* expression in *Arabidopsis thaliana* alters *in vitro* shoot organogenesis. *Physiol. Plant* **128**, 740–750 (2006).

46. Kumar, A., Taylor, M. A., Arif, S. A. M. & Davies, H. V. Potato plants expressing antisense and sense S-adenosylmethionine decarboxylase (SAMDC) transgenes show altered levels of polyamines and ethylene: Antisense plants display abnormal phenotypes. *Plant J.* **9**, 147–158 (1996).
47. Basu, S., Roychoudhury, A. & Sengupta, D. N. Identification of trans-acting factors regulating *SamDC* expression in *Oryza sativa*. *Biochem. Biophys. Res. Commun.* **445**, 398–403 (2014).
48. Anderson, L. E. & Carol, A. A. Enzyme co-localization in the pea leaf cytosol: 3-P-glycerate kinase, glyceraldehyde-3-P dehydrogenase, triose-P isomerase and aldolase. *Plant Sci.* **169**, 620–628 (2005).
49. Lovdal, T. & Lillo, C. Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. *Anal. Biochem.* **387**, 238–242 (2009).
50. Wang, H., Chen, S., Jiang, J., Zhang, F. & Chen, F. Reference gene selection for cross-species and cross-ploidy level comparisons in *Chrysanthemum* spp. *Sci. Rep.* **5**, 8094, <https://doi.org/10.1038/srep08094> (2015).
51. Mahboubi, A. *et al.* Aspen SUCROSE TRANSPORTER3 allocates carbon into wood fibers. *Plant Physiol.* **163**, 1729–1740 (2013).
52. Liu, M. Y., Wu, X. M., Long, J. M. & Guo, W. W. Genomic characterization of miR156 and SQUAMOSA promoter binding protein-like genes in sweet orange (*Citrus sinensis*). *Plant Cell Tiss. Org. Cult.* **130**, 103–116 (2017).
53. Plötner, B. *et al.* Chlorosis caused by two recessively interacting genes reveals a role of RNA helicase in hybrid breakdown in *Arabidopsis thaliana*. *Plant J.* **91**, 251–262 (2017).
54. Jian, B. *et al.* Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol. Biol.* **9**, 59, <https://doi.org/10.1186/1471-2199-9-59> (2008).
55. de Oliveira, T. M. *et al.* Analysis of the NAC transcription factor gene family in citrus reveals a novel member involved in multiple abiotic stress responses. *Tree Genet. Genom.* **7**, 1123–1134 (2011).
56. Li, H. *et al.* ZmWRKY33, a WRKY maize transcription factor conferring enhanced salt stress tolerances in *Arabidopsis*. *Plant Growth Regul.* **70**, 207–216 (2013).
57. Lv, P., Li, N., Liu, H., Gu, H. & Zhao, W. E. Changes in carotenoid profiles and in the expression pattern of the genes in carotenoid metabolisms during fruit development and ripening in four watermelon cultivars. *Food Chem.* **174**, 52–59 (2015).
58. Vij, S. & Tyagi, A. K. Genome-wide analysis of the stress associated protein (SAP) gene family containing A20/AN1 zinc-finger(s) in rice and their phylogenetic relationship with *Arabidopsis*. *Mol. Genet. Genomics* **276**, 565–575 (2006).
59. Silflow, C. D. *et al.* Plant tubulin genes: structure and differential expression during development. *Genesis* **8**, 435–460 (1987).
60. Brunner, A. M., Yakovlev, I. A. & Strauss, S. H. Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biol.* **4**, 14 (2004).
61. Yang, Q. *et al.* Reference gene selection for qRT-PCR in *Caragana korshinskii* Kom. under different stress conditions. *Mol. Biol. Rep.* **41**, 2325–2334 (2014).
62. Wong, M. L. & Medrano, J. F. Real-time PCR for mRNA quantitation. *Biotechniques* **39**, 75–85 (2005).

Acknowledgements

This research is supported by grants from the National Natural Science Foundation of China (Grant Nos 30771759, 31170636 and 31470681), the National Key Research and Development Program of China (Grant No. 2016YFD0600104), Natural Science Foundation of Guangdong Province of China (Grant No. 2016A030311032), Guangzhou Science and Technology Program (Grant No. 201607020024) and Foundation of Young Creative Talents in Higher Education of Guangdong Province(2017KQNCX017).

Author Contributions

C.P. and T.H. designed the research; T.H., S.-W.L., Z.-W.Y., Q.-J.Z. and X.-L.Z. conducted the experiment; J.L. and T.H. analyzed the data; J.L. and C.P. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-27633-5>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018