Pharmacogn. Mag.
A multifaceted peer reviewed journal in the field of Pharmacognosy and Natural Products
www.phcog.com | www.phcog.net

ORIGINAL ARTICLE

Selection of Suitable Reference Genes for Reverse Transcription-quantitative Polymerase Chain Reaction Normalization in *Artemisia annua* L. Plants at Different Stages of Growth and Development

Javed Ahmad, Nazima Nasrullah, Naved Quadri, Umara Nissar, Shashi Kumar[1], Malik Zainul Abdin

Centre for Transgenic Plant Development, Department of Biotechnology, School of Chemical and Life Sciences, Jamia Hamdard, [1]International Centre for Genetic Engineering and Biotechnology, New Delhi, India

Submitted: 01-12-2018 Revised: 06-03-2019 Published: 28-11-2019

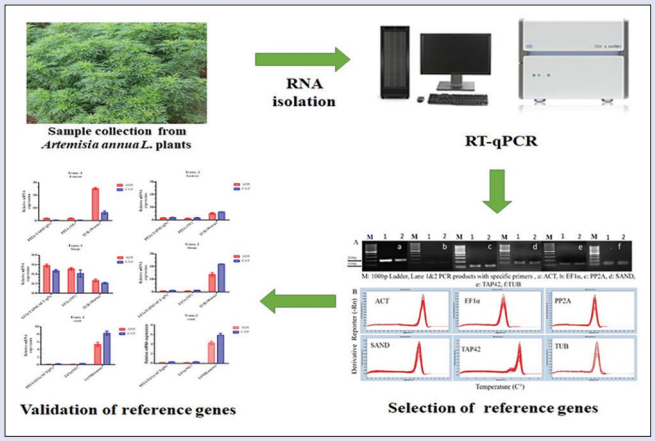
ABSTRACT

Background: Artemisinin, an antimalarial compound suggested by the WHO to treat drug-resistant malaria, was obtained from *Artemisia annua* L. plants. However, due to the low level of artemisinin in the plant causes limitation to its commercialization, so to increase the concentration of artemisinin, two transgenic lines were developed by us, overexpressing the key genes of artemisinin biosynthetic pathway, namely, 3 S-hydroxy-3-methyl glutaryl-CoA reductase (*HMGR*), amorpha-4,11-diene synthase (*ADS*) (Trans. 1) and *HMGR*, *ADS*, and *CYP71AV1* (cytochrome P450 monooxygenase) (Trans. 2). **Objectives:** Our main aim for this study was to select the suitable reference gene for the normalization of reverse transcription-quantitative polymerase chain reaction (Reverse transcription RT-qPCR) data in different tissues at various developmental stages in *A. annua* L. plants. **Materials and Methods:** Six candidate reference genes, namely; β-actin (*ACT*), elongation factor 1-alpha (*EF1α*), TAP-42 interacting protein (*TAP42*), SAND family protein, β-tubulin, and protein phosphatase 2A (*PP2A*) for their expression stability in the root, stem, leaf, and flower of *A. annua* L. plants at vegetative, preflowering, and flowering stages were analyzed using geNorm, NormFinder and BestKeeper, the Excel-based research tools. **Results:** The genes *ACT/PP2A, PP2A/TAP42,* and *EF1α/PP2A* were appropriate as reference genes in the leaf tissues at vegetative, preflowering, and flowering stages, respectively. In addition, *EF1α/PP2A* genes at vegetative and flowering stage, while *EF1α/TAP42* gene at preflowering stage was found suitable reference genes for normalization of expression data in the stem. In the root samples, *ACT/EF1α*, *EF1α/PP2A,* and *ACT/TAP42* sets were found to be reliable reference genes at vegetative, preflowering, and flowering stages, respectively, whereas, *PP2A/TAP42* gene set was found suitable for flower tissues at flowering stage. **Conclusion:** These results will be helpful in the normalization of expression data in RT-qPCR to find the reliable outcome.

Key words: *Artemisia annua* L., BestKeeper, geNorm, norm finder, reference gene

SUMMARY

- The study provided suitable reference genes for reverse transcription-quantitative polymerase chain reaction (RT-qPCR) normalization and emphasized the importance of validating reference genes for gene expression analysis in *Artemisia annua* plants
- Different reference gene sets were required for normalization of RT-qPCR expression data in different tissues and developmental stages
- The geNorm analysis shows that a minimal number of reliable pair of the reference gene for normalization of RT-qPCR data also vary in different tissues and developmental stages.



Abbreviations used: HMGR: 3 S-hydroxy-3-methyl glutaryl-CoA reductase; ADS: Amorpha-4,11-diene synthase; CYP71AV1: Cytochrome P450 monooxygenase; ACT: β-actin; EF1α: Elongation factor 1-alpha; TAP42:TAP-42-interacting protein; SAND: SAND family protein;TUB: β-tubulin; PP2A: Protein phosphatase 2A; TAIR: The Arabidopsis Information Resource; EST: Expressed sequence tag.

Correspondence:

Prof. Malik Zainul Abdin,
Centre for Transgenic Plant Development,
Department of Biotechnology, School of Chemical and Life Sciences, Jamia Hamdard,
New Delhi - 110 062, India.
E-mail: malikzabdin@gmail.com
DOI: 10.4103/pm.pm_600_18

Access this article online
Website: www.phcog.com
Quick Response Code:

INTRODUCTION

Artemisia annua L. plant (Family: *Asteraceae*), a traditional Chinese medicinal herb, accumulates numerous pharmacologically active compounds.[1] Artemisinin is the most important therapeutic compound present in this plant. It is used in combination with other drugs for the treatment of malaria. Besides antimalarial property, artemisinin has antiviral,[2] anticancer,[3] and antischistosomal activities.[4] The concentrations of artemisinin vary widely depending on the genotype, growth conditions, and developmental stages of plants.[5] Due to its

Cite this article as: Ahmad J, Nasrullah N, Quadri N, Nissar U, Kumar S, Abdin MZ. Selection of suitable reference genes for reverse transcription-quantitative polymerase chain reaction normalization in *Artemisia annua* L. plants at different stages of growth and development. Phcog Mag 2019;15:S377-85.

unique antimalarial property, *Artemisia* has attracted immense scientific interests among plant breeders, agricultural scientists, developmental geneticists, and molecular biologists. Consequently, numerous studies are underway to know the artemisinin biosynthetic pathway and the mechanisms of its biosynthesis in this plant.

In the synthesis of artemisinin both pathways; plastidial 2-C-methyl-D-erythritol-4 phosphate and cytosolic mevalonate (MVA)[6] contribute carbon, but the major contributor of carbon (80%) is MVA pathway. The conversion of 3 S-hydroxy-3-methyl glutaryl-CoA (HMG-CoA) to mevalonate (MVA) is catalyzed by 3 HMG-CoA Reductase (HMGR) activity which is a rate-limiting step in MVA pathway, thus limits the artemisinin biosynthesis and its accumulation in *A. annua* L. plants.[7] Further, mevalonate is converted to farnesyl pyrophosphate (FPP) through a sequence of biochemical reactions. Amorpha-4,11-diene synthase (ADS), thereafter, converts FPP to amorpha-4, 11-diene, the first committed precursor of artemisinin biosynthesis. This step hence links the mevalonate pathway to artemisinin biosynthesis. Amorpha-4, 11-diene, is then converted to artemisinic acid through two-step reactions, and these steps are catalyzed by cytochrome P450 monooxygenase (CYP71AV1). Modulation of the expression of different genes of artemisinin biosynthesis pathway has been considered an important strategy to increase the concentration of artemisinin. In our laboratory, we, therefore, developed two transgenic lines of *A. annua* L. by transforming wild-type plants with *HMGR* and *ADS* genes (Trans. 1) and *HMGR, ADS,* and *CYP* genes (Trans. 2), respectively, through *Agrobacterium tumefaciens*-mediated genetic transformation. To understand the expression patterns of these genes, reverse transcription-quantitative polymerase chain reaction (RT-qPCR) assays were carried out. This technique is commonly used due to its high sensitivity and specificity in the quantification of the transcriptional abundance of various individual genes. It also leads to synchronized quantification of gene expression in different samples, with a wide quantification range of up to several orders of magnitude in comparison to common techniques such as RT-PCR or northern hybridization.[8] Hence, it is necessary to normalize the expression data of the target genes first by reference genes, since the expression stability of reference genes has a high impact on the expression levels of the target genes analyzed using RT-qPCR.[9] It has been seen that the expressions of most commonly used reference genes themselves are influenced by a number of factors, namely across the tissue samples, growth and developmental stages of plants, different biotic and abiotic stresses, and other experimental conditions. This leads to the serious implications for the gene expression studies that have used invalidated reference genes. The significant downregulations of the actin and ubiquitin, two commonly used reference genes were reported in a transcriptome analysis of *Arabidopsis* infected by *Agrobacterium.*[10] It is thus necessary to demonstrate that the reference gene of choice is suitable for the experiment in question.[9] Suitable reference genes have been established for the various model, commercially and agronomically important plants such as rice,[11] *Arabidopsis,*[12] coffee,[13] carrot,[14] cotton[15] and tobacco.[16] This practice of gene normalization is, however, rare in medicinal plants. Further, the studies on gene expression in wild-type and transgenic *A. annua* L. plants with RT-qPCR have used actin, ubiquitin, *16S rRNA, GAPDH*, etc., as reference genes without establishing their suitability across different tissue samples, developmental stages, and different experimental conditions in order to get trustworthy data. In the present study, therefore, several software-based algorithms, for instance, geNorm,[17] NormFinder,[18] BestKeeper,[19] and the delta Cq[20] were employed to assess the stability of six candidate reference genes in both wild-type and transgenic *A. annua* L. plants. These algorithms have been used in previous studies, for normalization of the gene expression data.[12,21-24]

MATERIALS AND METHODS

Plant materials

Two lines of transgenic plants of *A. annua* L. (Trans. 1 and Trans. 2) were developed by *Agrobacterium*-mediated transformation using strain EHA105 which contain recombinant pCambia expression vectors. Trans. 1 lines contain expression cassette comprised two genes, *HMGR* and *ADS* drove by ubiquitin and cauliflower mosaic virus (CaMV) 35S promoter, respectively, and Trans. 2 lines contain expression cassette comprised three genes; *HMGR, ADS,* and *CYP71AV1* drove by ascorbate peroxidase (APX), CaMV 35S, and cassava vein mosaic virus promoters, respectively (data not published). The transgenic plants were hardened and grown in Transgenic Containment Facility of Jamia Hamdard, New Delhi, India, under controlled conditions at 25°C ± 2°C for 16/8 h light/dark conditions with a light intensity of 300 μmol/m^2/s and relative humidity of 60%–70%. Samples of different tissues from these transgenic as well as wild-type plants were taken in triplicate. The stages of growth were considered from the date of transfer (DOT) of hardened plants to the larger pots for growth and development. Samples were collected at vegetative (90 days from DOT), preflowering (180 days from DOT-budding stage), and flowering stages (200 days from DOT) of growth. These tissue samples were frozen in liquid nitrogen before storage at −80°C for RNA extraction.

Total RNA isolation and first-strand cDNA synthesis

Total RNA extraction was performed using Plant RNeasy Mini kit (Qiagen, Germany) according to the manufacturer's instruction from the frozen samples (100 mg each) of the root, stem, leaves, and flowers of the transgenic and wild-type plants at vegetative, preflowering, and flowering stages. The extracted RNA samples may contain an impurity of genomic DNA which was removed by treatment with DNase I (Sigma Aldrich). The RNA quality was validated using NanoDrop (ND1000) spectrophotometer and 1.5% (w/v) agarose gel electrophoresis. First-strand cDNA was synthesized from total RNA (1 μg) through RT using 20 μl Verso cDNA kit (Thermo Scientific), following the manufacturer's protocol. cDNA was diluted in a 1:10 ratio, which was then used in the RT-qPCR analysis. Six commonly used reference genes, β-actin (*ACT*), β-tubulin (*TUB*), elongation factor 1-alpha (*EFIα*), protein phosphatase 2A (*PP2A*), SAND family protein (*SAND*) and TAP-42 interacting protein (*TAP42*) were selected and their sequences were obtained from the The Arabidopsis Information Resource database (http://www.arabidopsis.org). The sequences of these reference genes identified have potential homologs, therefore, from the expressed sequence tag database (taxid: 35608) of *A. annua* L. from the NCBI were aligned using BLAST Sequence Alignment tool. Primers of these reference genes were designed using the Integrated DNA Technologies PrimerQuest tool (http://eu.idtdna.com/primerquest/home/index) and analyzed in OligoAnalyser (https://eu.idtdna.com/calc/analyzer) according to the developer's guidelines. The primer sequence along with their Tm (°C) and amplicon size are provided in the table [Table 1]. Amplicon specificity and size were verified by semiquantitative RT-PCR and gel electrophoresis, respectively.

Real-time quantitative polymerase chain reaction

Real-time-qPCR was carried out in a 96-well plate using SYBR Green I on a real-time-PCR Light Cycler® 480 System (Roche Diagnostics). The RT-qPCR experiment was designed according to the minimum information for publication of quantitative real-time-PCR experiment guidelines.[25] The 20 μL reaction mixtures consist of 10 μL SYBR Green I Mix, 2 μL diluted cDNA, double-distilled water, and a final primer concentration of 0.4 μM. In Light Cycler following experimental

Table 1: Description of candidate reference gene primers used for the quantitative real-time polymerase chain reaction

Gene symbol	Gene name	Gene bank accession number	Primer sequence	Tm (°C)	Amplicon size (bp)
ACT	β-actin gene	EY039526.1	AGCTCCTGCTCATAGTCAAG CCTATCTACGAAGGGTATGC	60.0	191
TUB	β-tubulin gene	EY115109.1	CTTACAACGCCACCCTTTCT TGGATTGGCGAGCTTTAGTG	60.0	116
EFIα	Elongation factor 1-alpha gene	FJ874734.1	GCCACTACACCAAAGTACTCAA GGGACTTTGTCAGGGTTGTATC	60.0	98
PP2A	Protein phosphatase 2A gene	EY073872.1	CGGCAGATTACCCAAGTGTATG GCGCTGTAAGAGGAAGGTAATC	60.0	112
SAND	SAND family protein gene	EY082985.1	CACTCTCATTCACCACCTCAC CATCATCGTCATTCCTCCATACA	60.0	102
TAP42	TAP-42-interacting protein	EY039843.1	GAGTGGGAAGCAGATGAGAAAG GAAGTGAGACAAGAGCGGTTAG	60.0	122
ADS	Amorpha-4,11-diene synthase	AB530988.1	ATGAAGCACGCCAACTTG CCCGTTCTTGTCCTTGTAG	60.0	219
CYP71AV1	Cytochrome P450 monooxygenase	AB706288.1	AGCAATGGCACTCTCACTG CTTTCTGGCTAAATCCCTAACC	60.0	190

ACT: β-actin; TUB: β-tubulin; EFIα: Elongation factor 1-alpha; PP2A: Protein phosphatase 2A; SAND: SAND family protein; TAP42: TAP-42 interacting protein; ADS: Amorpha 4,11-diene synthase; CYP71AV1: Cytochrome P450 monooxygenase

protocol was used; preincubation (95°C for 10 min), amplification, and quantification program (94°C for 50 s, 60°C for 60 s, 72°C for 60 s with a single fluorescence measurement) for 35 times; melting curve program (55°C–97°C with a heating rate of 0.1°C/s and a continuous fluorescence measurement); and finally a cooling step to 37°C for 10 min.

Data analysis

Excel-based softwares, geNorm,[17] NormFinder,[18] and BestKeeper,[17] were used to calculate the relative expression levels of reference genes. Data were used directly for stability calculations through BestKeeper, converted into relative quantities using the formula $2^{-\Delta cq}$ and then imported in geNorm and NormFinder. The geNorm algorithm first calculates an expression stability value (M) for each gene, then the level of pair-wise variation (V) for each reference gene with others is calculated. Most stable expression of the reference gene is shown by the lowest value of M within the set of genes examined. NormFinder works in a similar manner as that of geNorm. It identifies the genes with optimal normalization among a set of different reference genes. Lowest stability value indicates the most stable expression within the set of genes examined. BestKeeper algorithm is based on the pair-wise correlation analysis of raw C_q value and establishes the best-suited reference gene.

Reference gene validation

The best and worst reference genes identified by software-based tools were validated in the root, stem, and leaf of wild-type and transgenic *A. annua* L. plants at the preflowering stage using them to normalize the expression data of the transgenes.

RESULTS

Quality control and selection of the candidate reference gene

Six potential candidate reference genes, namely, *ACT, TUB, EF1α, PP2A, SAND,* and *TAP42* were used for gene expression studies using RT-qPCR in wild-type and transgenic *A. annua* L. full-length sequences were retrieved from *A. annua* L. transcriptome on account of homology analysis with *Arabidopsis* sequences. Specific primers for the candidate reference genes were designed and confirmed on the basis of the amplification specificity. The amplification of reference genes using these primers on PCR, yielded products of the expected sizes on 1.5% (w/v) agarose gel electrophoresis confirms the primer specificity [Figure 1a]. The presence of a single band in the agarose gel electrophoresis in each case and the single-peak melting curves in all RT-qPCR amplifications ruled out the presence of primer dimers [Figure 1b].

Variation in the expression level of the candidate reference genes

The extracted RNA samples, in triplicate from the root, stem, leaf, and flower of wild-type and the transgenic lines of *A. annua* L. at different stages, namely, vegetative, preflowering, and flowering were used to synthesize cDNAs by RT approach. These cDNAs were further used as a template for RT-qPCR analysis. The threshold cycle value (Cq) denotes the expression level upon which the fluorescence signal reaches over the baseline threshold during the RT-qPCR analysis of candidate reference genes. Baseline thresholds were standardized to mean 75.55. The distribution of raw data is shown by a Box and Whiskers plot [Figure 2]. Higher Cq value corresponds to lower expression level and lower Cq value corresponds to higher expression level. Out of the selected candidate reference genes, at least two genes, *ACT* and *EF1α,* showed the highest expression levels (15<Cq<27), while genes *TUB, PP2A, SAND,* and *TAP42* showed lowest expression levels (19<Cq<32). The genes *ACT, TUB, EF1α,* and *PP2A* (standard deviation [SD] = 2.11–2.72; Cq value varied from 15.75 to 32.74) showed maximum variability in their expression levels, but the genes *SAND* and *TAP42* with Cq value varying from 20 to 28.90 showed minimum variability (SD = 1.55–1.70). In the transgenic together with wild-type *A. annua* L. plants, the expression profile of candidate reference genes was studied in various tissues and developmental stages [Supplementary Figure S1]. The expression of *ACT* gene in the leaves of wild-type and transgenic plants decreased from vegetative to flowering stage. The similar trend was also observed in the case of *TUB* gene expression in wild-type plants. In transgenic plants, however, its expression decreased from vegetative to the preflowering stage and became almost constant at the flowering stage. Expression of *EF1α* decreased from vegetative to the preflowering stage and then increased in the flowering stage. The contrary result was, however, observed in the expression of *SAND*. The expression of *TAP42* in wild-type plants remained almost constant from vegetative to preflowering and then increased at the flowering stage. Instead, in transgenic plants, the expression of *TAP42* increases in the leaves from vegetative to the preflowering stage but decreased afterward

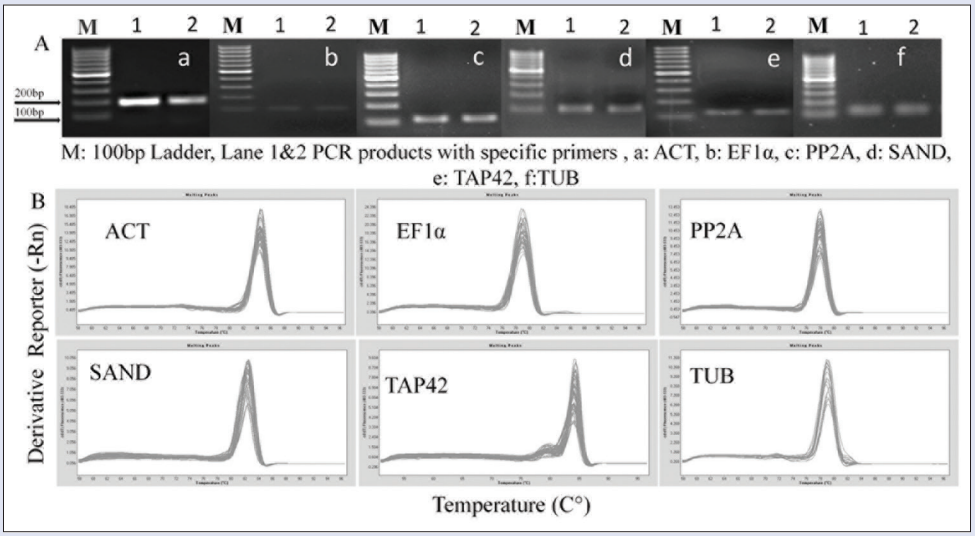


Figure 1: Reverse transcription-quantitative polymerase chain reaction amplification specificity. (A) Agarose gel (1.5%) showing amplification of specific polymerase chain reaction products of the expected size. (B) The melting curve of six reference genes showing a single peak

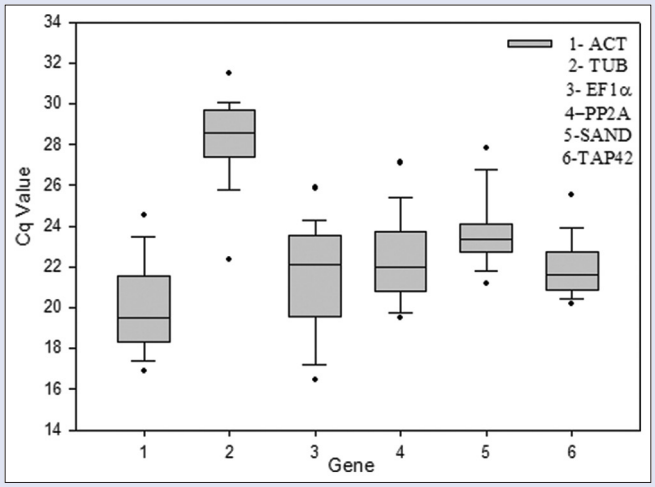


Figure 2: Quantitation cycle (Cq) values of the candidate reference genes in the experimental samples. Each box indicated 25–75 percentiles and the line across box depicts median values Whiskers represent percentiles from 5th to 95th and outliers by dots

at the flowering stage. In the leaves of wild-type plants, the expression of *PP2A* decreased from vegetative to the preflowering stage and then became almost constant. However, in the transgenic plants, the leaves have shown an unchanged expression level of *PP2A* from vegetative to preflowering, which later was decreased at flowering stage in the stem, the expression of *ACT, TAP42,* and *PP2A* decreased from vegetative to flowering stage in wild-type and transgenic plants. Expression of *TUB* and *EF1α* decreased from vegetative to the preflowering stage and then increased. The expression of *SAND* in the wild-type plants decreased from vegetative to flowering stage but increased from vegetative to the preflowering stage and remained almost constant at the flowering stage of transgenic plants. The roots have shown the similar trends in the expression levels of *ACT, PP2A, EF1α, SAND,* and *TAP42* in the wild-type as well as transgenic plants. Conversely, from vegetative to preflowering, these genes have shown an increased level of expression which got decreased at the flowering stage. The expression of *TUB* gene in wild-type plants decreased from vegetative to preflowering and remained constant at the flowering stage, whereas its expression increased from vegetative to preflowering and remained constant at flowering stage of the development in transgenic plants. In transgenic plants, the changes observed in some of the above-mentioned reference genes could be due to the overexpression of genes involved in artemisinin biosynthesis. A similar pattern of expression of the reference genes observed in transgenic plants in our study had also been reported earlier as well.[26]

geNorm analysis

The expression stabilities of the six candidate reference genes in different tissues of wild-type and transgenic *A. annua* L. plants at various developmental stages were analyzed by a statistical algorithm, geNorm. geNorm calculates gene expression stability (M) as the average pair-wise variation between all tested genes in a given set of samples. On account of geNorm analysis, the cutoff range of stability value (M) is <1.5, and the stability of reference genes is related inversely proportional to M value of that reference gene.[17] When "total sample" sets were considered, *ACT* and *PP2A* gene pairs were two most stable expressed reference genes out of all candidates, calculated by step-wise exclusion of genes with a stability value of 1.11. Leaf samples in wild and transgenic plants at vegetative, preflowering, and flowering stages had M = 0.73, 0.36, and 0.71 for *ACT/PP2A*, *PP2A/TAP42,* and *EF1α/PP2A,* respectively. This gene sets, therefore, had the highest expression stability in the leaf tissue at various developmental stages of transgenic together with wild-type *A. annua* L. plants. In the wild and transgenic *A. annua* L. plants, the lowest M values obtained for stem were 0.38 and 0.90 at vegetative and flowering stages, respectively, for the gene set *EF1α/PP2A*. Furthermore, in the preflowering stage, it was 0.50 for the gene set *EF1α/TAP42*. The most stable reference gene pair in the root of wild and transgenic plants was found to be *ACT/EF1α* with M value of 0.68 at vegetative stage, whereas at preflowering and flowering stages, the gene sets *EF1α/PP2A* and *ACT/TAP42* had lowest M values (0.19 and 0.50 respectively). In the flowers, *PP2A* and *TAP42* were highly stable reference genes in the flowers with an M = 0.78 [Figure 3]. Nonetheless, in all the developmental stages and in different tissues, the M values of selected reference genes were found to be less than one that indicates its relatively adequate range of expression stabilities. According to above geNorm analysis, it

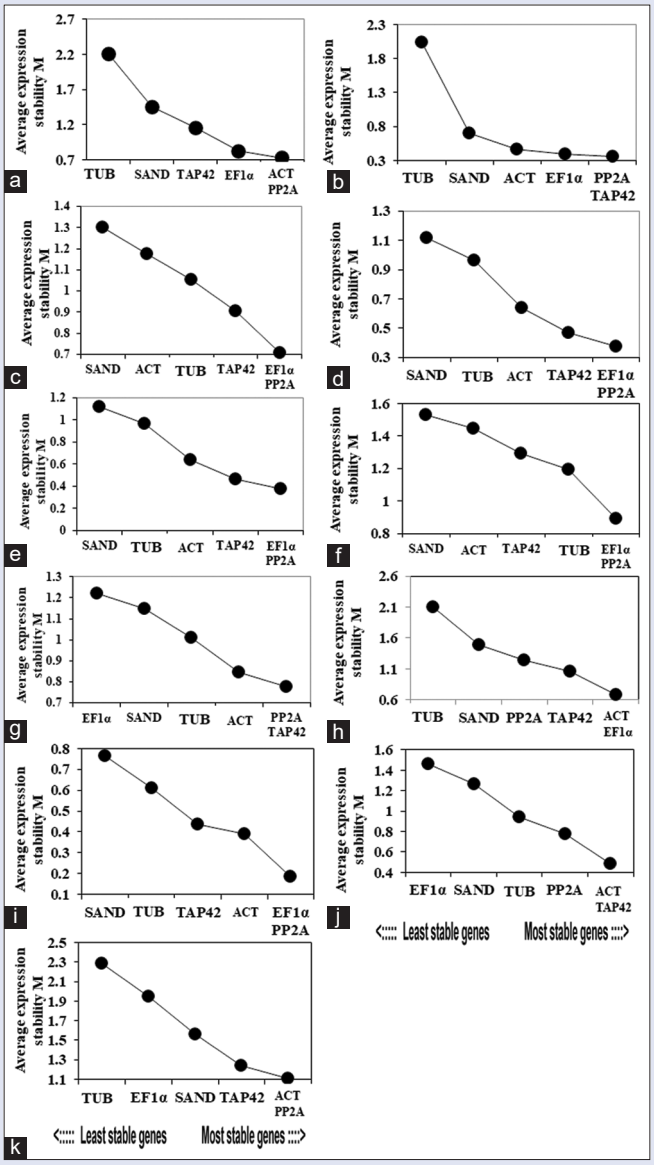


Figure 3: M values of six reference genes in different sample groups by using geNorm analysis: Leaf: (a) vegetative (b) preflowering (c) flowering, Stem: (d) vegetative (e) preflowering (f) flowering, (g) Flower total, root: (h) vegetative, (i) preflowering (j) flowering, (k) Total sample

is clear that based on expression stability; different sets of genes could be taken as reference genes for different tissues as well as stages of the development. A pair-wise variation (V_n/V_{n+1}) was estimated in order to identify the minimal number of reliable pair of the reference gene. The accurate normalization of *A. annua* L. was obtained on the basis of the cutoff value of 0.15.

The pair-wise variation at $V_{2/3}$ was below 0.15 signifying that two reference genes are sufficient for normalization at the preflowering stage of leaf and the adding of the third reference gene had no noteworthy effect on the result. In all the other experimental sets, pair-wise variations were >0.15 [Figure 4] and hence lowest V_n/V_{n+1} was selected as a minimal number of reference genes for normalization as reported by several other researchers.[27-29] There were no significant differences found in our study in the stability of the expression of reference genes for the wild-type as well as for the transgenic *A. annua* L. plants.

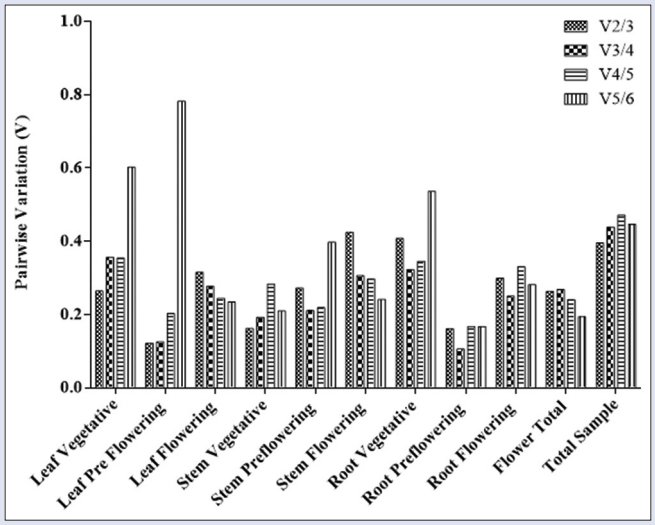


Figure 4: Pair-wise variation (V) for the optimal number of reference genes for normalization in a different set of samples

NormFinder analysis

In NormFinder, a Microsoft Excel-based tool, the reference genes are ranked on the basis of their minimal combined inter- and intragroup variations of expression levels. Stable expression of the genes showed an inverse relation with the average expression stability of that gene. The wild-type and transgenic *A. annua* L. plants, the samples were grouped in the total root, total stem, total leaf, total flower, and in individual, root, stem, and leaf at vegetative, preflowering, and flowering stages. The gene expression data without subgroupings were also separately estimated as "total samples" (pooled root, stem, leaf, along with flower samples collected at various developmental stages) of wild type with transgenic *A. annua* L. plants [Table 2].

The grouping of samples was done as one main and several subgroups, moreover, TAP42 (stability value of 0.54) was best among all the reference genes if all the samples were gathered together. However, ACT was found to be the most stable if the study was carried out on individual groups including total leaf and flower samples. In total stem and root samples, *TAP42* was categorized as the most stable reference gene. In the leaf samples of wild-type as well as transgenic *A. annua* L. plants at the vegetative stage, *PP2A* gene whereas at preflowering and flowering stages *TAP42* gene was found most stable reference gene. *TAP42, EF1α,* and *PP2A* were found most stable genes at vegetative, preflowering, and flowering stages, respectively, in stem samples of wild-type and transgenic *A. annua* L. plants. *TAP42, EF1α,* along with *TUB* were most stable genes at vegetative, preflowering, and flowering stages, respectively, in the root samples of wild-type as well as transgenic *A. annua* L. plants. Similar to the NormFinder, the stability ranking of reference was generated by the geNorm, but it was slightly different from the earlier as both the tools use a different algorithm as the mathematical approaches.[15,30] The above result suggests that the suitability of the reference genes vary in various developmental stages and in different tissues in the *A. annua* L. plants.

BestKeeper analysis

The raw Cq values were used to calculate inter- and intragene relation between the reference gene pairs by estimating the numerous pair-wise correlation by the BestKeeper tool. The SD, the coefficient of correlation, and percentage covariance of individual reference gene were used to evaluate gene expression stability.

Table 2: Expression stability of six candidate reference genes of *Artemisia annua* analyzed by norm finder

Rank	Total sample	Total flower	Total root	Total stem	Total leaves	Root (vegetative)	Root (preflowering)	Root (flowering)	Stem (vegetative)	Stem (preflowering)	Stem (flowering)	Leaf (vegetative)	Leaf (preflowering)	Leaf (flowering)
1	TAP42	ACT	TAP42	TAP42	ACT	TAP42	EF1α	TUB	TAP42	EF1α	PP2A	PP2A	TAP42	TAP42
M	0.537	0.416	0.356	0.239	0.319	0.36	0.072	0.304	0.199	0.172	0.548	0.213	0.123	0.359
2	ACT	PP2A	ACT	PP2A	PP2A	ACT	TAP42	TAP42	PP2A	TAP42	TAP42	ACT	ACT	TUB
M	0.707	0.453	0.741	0.781	0.682	0.662	0.193	0.517	0.2	0.172	0.58	0.252	0.157	0.393
3	PP2A	TAP42	SAND	ACT	TAP42	SAND	PP2A	PP2A	EF1α	ACT	EF1α	EF1α	EF1α	EF1α
M	0.756	0.486	0.909	0.814	0.874	0.733	0.228	0.552	0.296	0.435	0.649	0.295	0.336	0.46
4	SAND	TUB	PP2A	SAND	SAND	PP2A	ACT	ACT	ACT	PP2A	ACT	SAND	PP2A	ACT
M	1.112	0.703	0.94	1.252	1.306	0.849	0.333	0.554	0.696	0.673	0.86	1.154	0.516	0.704
5	EF1α	SAND	EF1α	TUB	EF1α	EF1α	TUB	SAND	SAND	SAND	TUB	TAP42	SAND	PP2A
M	1.543	0.728	1.327	1.256	1.587	1.108	0.562	1.087	0.841	0.866	0.907	1.235	0.822	0.888
6	TUB	EF1α	TUB	EF1α	TUB	TUB	SAND	EF1α	TUB	TUB	SAND	TUB	TUB	SAND
M	1.799	0.764	1.395	1.592	2.31	2.201	0.681	1.132	0.844	1.635	0.948	2.481	3.249	0.935

ACT: β-actin; TUB: β-tubulin; SAND: SAND family protein; PP2A: Protein phosphatase 2A; TAIR: The Arabidopsis Information Resource; EF1α: Elongation factor 1-alpha; ADS: Amorpha-4,11-Diene Synthase; TAP42: TAP-42 interacting protein

BestKeeper analysis showed that *TAP42* is the gene with the least variation, while *EF1α* with the maximum variation in overall sample sets. *TAP42, ACT,* along with *EF1α* were found stable at vegetative, preflowering, and flowering stages, respectively, in the leaf samples of wild-type as well as transgenic *A. annua* L. plants, whereas *TAP42* gene was found to be stable in "total leaf" and "total stem" samples. *TUB* gene was highly stable reference gene in the "total root" sample. In the stem, the *PP2A* reference gene was found to be stably expressed in at vegetative and flowering stages and *EF1α* at the preflowering stage. In root samples, *TUB, EF1α,* and *ACT* were stably expressed at vegetative, preflowering, and flowering stages, respectively. *TUB* was found stably expressed reference gene in "total flower" samples of *A. annua* L. plants [Table 3].

Reference gene validation

The reference genes selected by geNorm and NormFinder algorithms were examined for its expression stability and the suitability by investigating the expression profiles of transgenes *CYP* and *ADS* in different tissues, namely, root, stem, and leaf in transgenic as well as wild-type *A. annua* L. plants at preflowering stage. We have chosen the preflowering stage because of the maximum accumulation of artemisinin in the leaves at preflowering stage was reported earlier.[31-34] The gene *ADS* was overexpressed in both Trans. 1 and Trans. 2 plants, whereas *CYP* was overexpressed only in Trans. 2 plants of *A. annua* L. geNorm and NormFinder determined similar expression stability ranking of the six candidate reference genes in the samples of different tissues collected at the preflowering stage of development. In the leaf samples, the expression of *ADS* and *CYP* were validated using the gene set *PP2A/TAP42* and *PP2A* as best reference calculated by geNorm and NormFinder, respectively, while *TUB* as the worst reference gene. *ADS* gene showed a similar pattern of expression in both transgenic lines, whereas *CYP* showed higher expression in Trans. 2 as compared to Trans. 1 on normalization by the best reference genes (*PP2A/TAP42* and *PP2A),* whereas, on normalization by the worst reference gene (*TUB),* their expressions were increased several folds in both the transgenic lines. In the stem, the expression levels of the genes *ADS* and *CYP,* on normalization by gene set *EF1α/TAP42/ACT* (determined by geNorm) and *EF1α* (by NormFinder) were almost similar, whereas, on normalization by *TUB* (the worst gene), these genes showed decreased expression levels in Trans. 1. In stem of Trans. 2, the expression of *ADS* and *CYP* was found higher in comparison to Trans. 1 on normalization by *EF1α/TAP42/ACT* (determined by geNorm) and *EF1α* (by NormFinder), but on normalization by *TUB* (the worst gene), the expression levels of *ADS* and *CYP* were found enormously enhanced by several folds. In root, at the preflowering stage, on normalization by gene set *PP2A/EF1α/ACT* determined by geNorm and *EF1α* by NormFinder, *ADS* and *CYP* genes showed decreased expression when compared with the leaf of both transgenic lines, which is supported by the previous study.[35] On the other hand, when normalized with *SAND* (the worst gene) determined by geNorm, the *ADS* and *CYP* genes showed higher expression in the root than the leaves, which is not expected [Figure 5].

DISCUSSION

During the past decades, different potent techniques have been developed to differentiate expression levels of genes among different tissues, organs, plant species, and under different experimental conditions. Among these, RT-qPCR is an important technique due to its reliability and sensitivity, for quantification of gene expression covering different biological processes. However, to avoid biased quantification results, a set of reference genes is required to normalize the expression data. Therefore, for different experimental sets, each candidate reference genes requires its evaluation. In the present study, different software, namely,

Table 3: BestKeeper analysis ranking

Rank	Total sample	Total flower	Total root	Total stem	Total leaves	Root (vegetative)	Root (preflowering)	Root (flowering)	Stem (vegetative)	Stem (preflowering)	Stem (flowering)	Leaf (vegetative)	Leaf (preflowering)	Leaf (flowering)
1	TAP42	TUB	TUB	TAP42	TAP42	TUB	EF1α	ACT	PP2A	EF1α	PP2A	TAP42	ACT	EF1α
SD (±Cq)	1.19	0.36	1.31	0.71	0.79	0.72	0.27	0.98	0.19	0.59	0.47	0.62	0.27	0.61
Cv (%Cq)	5.41	1.22	4.63	3.29	3.68	2.45	1.15	4.32	0.94	2.56	1.99	2.83	1.42	2.89
2	SAND	ACT	EF1α	SAND	SAND	SAND	PP2A	SAND	TAP42	TAP42	TUB	PP2A	TAP42	PP2A
SD (±Cq)	1.29	0.68	1.53	0.94	1.09	1.34	0.3	1.13	0.27	0.6	0.72	0.95	0.39	0.84
Cv (%Cq)	5.43	3.32	6.48	4.05	4.71	5.05	1.4	4.71	1.28	2.79	2.55	4.55	1.86	3.58
3	TUB	TAP42	SAND	TUB	PP2A	TAP42	TAP42	TAP42	EF1α	SAND	TAP42	EF1α	PP2A	TAP42
SD (±Cq)	1.69	0.8	1.87	1.32	1.27	1.78	0.39	1.17	0.45	0.71	0.72	1	0.41	0.86
Cv (%Cq)	6.02	3.56	7.67	4.6	5.82	7.27	1.92	5.03	2.65	3.11	3.25	5.39	1.95	3.89
4	PP2A	PP2A	TAP42	PP2A	ACT	ACT	ACT	TUB	ACT	PP2A	EF1α	ACT	EF1α	TUB
SD (±Cq)	1.8	0.84	1.91	1.56	1.33	1.98	0.39	1.44	0.62	0.78	0.81	1.24	0.47	0.96
Cv (%Cq)	8	3.62	8.4	7.27	6.83	8.39	2.03	5.18	3.6	3.65	3.85	6.62	1.98	3.35
5	ACT	SAND	ACT	ACT	EF1α	EF1α	TUB	PP2A	SAND	ACT	ACT	SAND	SAND	SAND
SD (±Cq)	1.87	0.85	2.25	1.58	1.9	2.14	0.68	1.55	0.97	0.8	1.05	1.51	0.73	1.04
Cv (%Cq)	9.3	3.5	10.34	8.41	9	8.98	2.46	6.13	4.09	4.27	5.1	6.29	3.24	4.53
6	EF1α	EF1α	PP2A	EF1α	TUB	PP2A	SAND	EF1α	TUB	TUB	SAND	TUB	TUB	ACT
SD (±Cq)	2.27	1.41	2.43	2.45	2.72	2.18	0.83	2.22	1.03	2.02	1.2	3.24	2.81	1.4
Cv (%Cq)	10.52	6.8	10.24	12.04	10.11	8.75	3.66	9.41	3.65	6.88	5.16	12.91	1.42	6.67

SD: Standard deviation; Cq: Quantitation cycle; CV: Coefficients of Variance; ACT: β-actin; TUB: β-tubulin; SAND: SAND family protein; PP2A: Protein phosphatase 2A; TAIR: The Arabidopsis Information Resource; EF1α: Elongation factor 1-alpha; ADS: Amorpha-4,11-Diene Synthase; TAP42: TAP-42 interacting protein

geNorm, NormFinder, and BestKeeper were applied to determine the exact number of reference genes required, which expression is stable across experimental conditions. These selected reference genes were used for accurate normalization of gene expression data from *A. annua* L. Two transgenic lines developed by transforming with two different gene constructs one containing *ADS* and *HMGR* genes (Trans. 1) and second consist of *ADS*, *HMGR*, and *CYP* (Trans. 2) were analyzed for the gene normalization. In our study, we found that the suitable reference genes could vary among different developmental stages and tissues. However, the change in expression levels of reference genes between wild-type and transgenic plants was statistically insignificant. For example, in leaf samples of *A. annua* L. plants, there is variation in gene expression of reference genes at all the three developmental stages, i.e., vegetative (*ACT*/*PP2A*), preflowering (*PP2A*/*TAP42*), and flowering (*EF1α*/*PP2A*) stages but remain same in wild and transgenic plants.

The variations in reference gene expression in different tissues at different developmental stages have been reported earlier in *Linum usitatissimum* L. as well.[36] RT-qPCR, arguably the most widely used molecular technique, is however far from being a "gold standard" because of the lack of standardization. The average difference in expression level of a gene of interest after normalization with any of the two randomly selected nonvalidated reference genes is between 3- and 6-fold among 10%–25% of the case studies.[37] Such variation makes it difficult to draw a conclusion with biological relevance. To avoid biased normalization, ideas of using multiple reference genes for analysis of gene expression data of target genes have picked up the pace among researchers. In our study, geNorm pair-wise variation at $V_{2/3}$ was found below 0.15 at the preflowering stage of leaf samples of *A. annua* L. which support the use of the two reference genes for attaining better accuracy in normalization. Hence, the addition of other reference genes would not affect the expression results in this case. However, for the remaining experimental sets, pair-wise variation was >0.15, showed that it would require more than two reference genes for normalization. In *Lipaphis erysimi* and chicory plants under different experimental conditions, more than two reference genes were used earlier.[38,39] Gene normalization is incomplete without validating the best and the worst reference genes obtained by a different algorithm. Thus, we validated the best and the worst reference gene by normalizing the expression data of target genes, *ADS* and *CYP* in Trans. 1 and Trans. 2 of *A. annua* L. plant. In our study, PP2A/*TAP42* and *PP2A,* the best reference genes determined by geNorm and NormFinder, respectively, whereas, *TUB* the worst reference genes were used to validate the expression of the target genes (*ADS* and *CYP*) in the preflowering stage of the leaf. The target genes, *ADS* and *CYP,* when normalized by reference genes *PP2A*/*TAP42* and *PP2A*, indicated a similar pattern of overexpression of *ADS* in both transgenic lines (Trans. 1 and Trans. 2), whereas the gene *CYP* which was overexpressed only in Trans. 2 showed higher expression in comparison to Trans. 1 as per the expectation. On the other hand, on normalization with *TUB* as reference gene (worst), *ADS* and *CYP* showed several folds enhanced expression. These transgenes, when normalized by the best reference gene in the root of *A. annua* L. the expression, was lower in comparison to leaf samples as per the expectation, whereas on normalization with worse reference gene, expression of these transgenes showed higher expression than leaf samples, thus validating our normalization results. To setup a standardized RT-qPCR protocol with well-characterized reference genes in *A. annua* L. will be beneficial in this area of research. In summary, easily available algorithms, namely, geNorm, NormFinder, BestKeeper, and delta Ct method were used to analyze the expression profile of six candidate reference genes in various tissues at different developmental stages of wild-type and transgenic plants of *A. annua* L. In each experimental condition, different sets of reference genes were required, thus reestablish that in different tissues, experimental conditions,

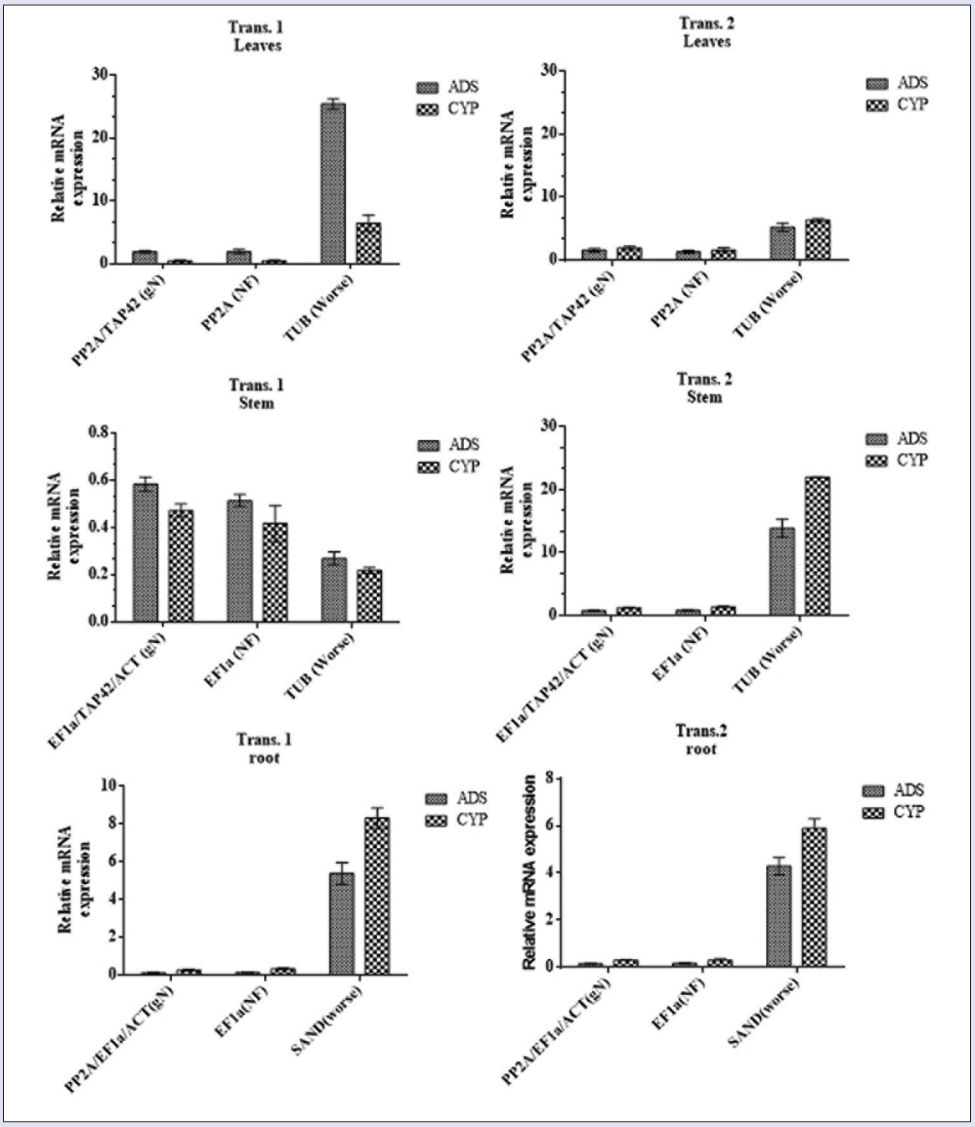


Figure 5: The relative expression level of amorpha-4,11-diene synthase and cytochrome P450 monooxygenase gene in different tissues normalized by the reference gene as recommended by geNorm and NormFinder at the preflowering stage. gN: geNorm; NF: NormFinder

developmental stage, etc., reference genes respond differently. These results, hence, implying that no single reference gene can be used as a universal reference gene for all the tissues, experimental conditions. Our study constitutes to setup a standardized RT-qPCR protocol for the functional genomics research in *A. annua* L.

CONCLUSION

Artemisinin, the sole source of the antimalarial drug, makes *A. annua* L. plants an economically important crop. Hence, functional genomics and gene-expression studies would continue to constitute an important part of the basic research on this plant. This work establishes the first thorough study for the validation of an optimal number of reference genes required to quantify the transcript levels in various tissues and developmental stages in transgenic and wild-type plants of *A. annua* L. The six reference genes tested in our study showed that different organs and developmental stages affect the reference gene expression in *A. annua* L. plants, but in transgenic and wild-type plants, the best reference gene selected by geNorm and NormFinder remains the same. Furthermore, the stability of the reference genes, irrespective of a different number of overexpressed genes in Trans. 1 (*ADS* and *HMGR*) and Trans. 2 (*CYP*, *ADS*, and *HMGR*), have no significant effect. In summary, selection of *A. annua* L. reference genes for gene expression studies in RT-qPCR will facilitate more precise and reliable normalization of expression data, in this pharmaceutically and economically important plant.

Acknowledgements

The authors would like to thank Dr. Gaurav Dogra, application engineer from Rosche Life Sciences for providing technical support to perform RT-qPCR.

Financial support and sponsorship

This work was financially supported by the University Grant Commission (UGC), New Delhi, under SAP-DRS II Program, UGC-BSR fellowship and the facilities provided by the Department of Biotechnology, Jamia Hamdard, New Delhi - 110 062. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Brown GD. The biosynthesis of artemisinin (Qinghaosu) and the phytochemistry of *Artemisia annua* L. (Qinghao). Molecules 2010;15:7603-98.
2. Romero MR, Serrano MA, Vallejo M, Efferth T, Alvarez M, Marin JJ. Antiviral effect of artemisinin from *Artemisia annua* against a model member of the flaviviridae family, the Bovine Viral Diarrhoea Virus (BVDV). Planta Med 2006;72:1169-74.
3. Efferth T. Molecular pharmacology and pharmacogenomics of artemisinin and its derivatives in cancer cells. Curr Drug Targets 2006;7:407-21.
4. Utzinger J, Xiao SH, Tanner M, Keiser J. Artemisinins for schistosomiasis and beyond. Curr Opin Investig Drugs 2007;8:105-16.
5. Ranjbar M, Naghavi MR, Alizadeh H, Soltanloo H. Expression of artemisinin biosynthesis genes in eight *Artemisia* species at three developmental stages. Ind Crop Prod 2015;76:836-43.
6. Towler MJ, Weathers PJ. Evidence of artemisinin production from IPP stemming from both the mevalonate and the nonmevalonate pathways. Plant Cell Rep 2007;26:2129-36.
7. Ram M, Khan M, Jha P, Khan S, Kiran U, Ahmad M, *et al.* HMG-CoA reductase limits artemisinin biosynthesis and accumulation in *Artemisia annua* L. plants. Acta Physiol Plant 2010;32:859-66.
8. Lata C, Sahu PP, Prasad M. Comparative transcriptome analysis of differentially expressed genes in foxtail millet (*Setaria italica* L.) during dehydration stress. Biochem Biophys Res Commun 2010;393:720-7.
9. Huggett J, Dheda K, Bustin S, Zumla A. Real-time RT-PCR normalisation; strategies and considerations. Genes Immun 2005;6:279-84.
10. Ditt RF, Kerr KF, de Figueiredo P, Delrow J, Comai L, Nester EW. The arabidopsis thaliana transcriptome in response to *Agrobacterium tumefaciens*. Mol Plant Microbe Interact 2006;19:665-81.
11. Li QF, Sun SS, Yuan DY, Yu HX, Gu MH, Liu QQ. Validation of candidate reference genes for the accurate normalization of real-time quantitative RT-PCR data in rice during seed development. Plant Mol Biol Rep 2010;28:49.
12. Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A. Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. Planta 2008;227:1343-9.
13. Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG. Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. BMC Mol Biol 2009;10:1.
14. Tian C, Jiang Q, Wang F, Wang GL, Xu ZS, Xiong AS. Selection of suitable reference genes for qPCR normalization under abiotic stresses and hormone stimuli in carrot leaves. PLoS One 2015;10:e0117569.
15. Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, 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 2010;10:49.
16. Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. Mol Genet Genomics 2010;283:233-41.
17. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:RESEARCH0034.
18. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 2004;64:5245-50.
19. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: Best keeper-excel-based tool using pair-wise correlations. Biotechnol Lett 2004;26:509-15.
20. Silver N, Best S, Jiang J, Thein SL. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol Biol 2006;7:33.
21. Wu ZJ, Tian C, Jiang Q, Li XH, Zhuang J. Selection of suitable reference genes for qRT-PCR normalization during leaf development and hormonal stimuli in tea plant (*Camellia sinensis*). Sci Rep 2016;6:19748.
22. Petriccione M, Mastrobuoni F, Zampella L, Scortichini M. Reference gene selection for normalization of RT-qPCR gene expression data from *Actinidia deliciosa* leaves infected with *Pseudomonas syringae* pv. Actinidiae. Sci Rep 2015;5:16961.
23. Lin YL, Lai ZX. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. Plant Sci 2010;178:359-65.
24. Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. BMC Mol Biol 2009;10:11.
25. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, *et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clin Chem 2009;55:611-22.
26. Ditt RF, Nester EW, Comai L. Plant gene expression response to *Agrobacterium tumefaciens*. Proc Natl Acad Sci U S A 2001;98:10954-9.
27. Chen J, Huang Z, Huang H, Wei S, Liu Y, Jiang C, *et al.* Selection of relatively exact reference genes for gene expression studies in goosegrass (*Eleusine indica*) under herbicide stress. Sci Rep 2017;7:46494.
28. Kuijk EW, du Puy L, van Tol HT, Haagsman HP, Colenbrander B, Roelen BA. Validation of reference genes for quantitative RT-PCR studies in porcine oocytes and preimplantation embryos. BMC Dev Biol 2007;7:58.
29. Silveira ED, Alves-Ferreira M, Guimarães LA, da Silva FR, Carneiro VT. Selection of reference genes for quantitative real-time PCR expression studies in the apomictic and sexual grass *Brachiaria brizantha*. BMC Plant Biol 2009;9:84.
30. Cruz F, Kalaoun S, Nobile P, Colombo C, Almeida J, Barros LM, *et al.* Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. Mol Breed 2009;23:607-16.
31. Alam P, Kamaluddin M, Khan MA, Mohammad A, Khan R, Abdin MZ. Enhanced artemisinin accumulation and metabolic profiling of transgenic *Artemisia annua* L. plants over-expressing by rate-limiting enzymes from isoprenoid pathway. J Plant Interact 2014;9:655-65.
32. Woerdenbag HJ, Pras N, Chan NG, Bang BT, Bos R, van Uden W, *et al.* Artemisinin, related sesquiterpenes, and essential oil in *Artemisia annua* during a vegetation period in Vietnam. Planta Med 1994;60:272-5.
33. Liersch R, Soicke H, Stehr C, Tüllner HU. Formation of artemisinin in *Artemisia annua* during one vegetation period*,1. Planta Med 1986;52:387-90.
34. Jha P, Ram M, Khan M, Kiran U, Abdin M. Impact of organic manure and chemical fertilizers on artemisinin content and yield in *Artemisia annua* L. Ind Crop Prod 2011;33(2):296-301.
35. Xiang L, Yan Z, Wang G, Liu W, Tang K, Liao Z. Relative expression of genes involved in artemisinin biosynthesis and artemisinin accumulation in different tissues of *Artemisia annua*. Zhongguo Zhong Yao Za Zhi 2012;37:1169-73.
36. Huis R, Hawkins S, Neutelings G. Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). BMC Plant Biol 2010;10:71.
37. Hellemans J, Vandesompele J. Selection of reliable reference genes for RT-qPCR analysis. Quantitative Real-Time PCR. New York: Springer; 2014. p. 19-26.
38. Koramutla MK, Aminedi R, Bhattacharya R. Comprehensive evaluation of candidate reference genes for qRT-PCR studies of gene expression in mustard aphid, *Lipaphis erysimi* (Kalt). Sci Rep 2016;6:25883.
39. 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 2010;11:15.