



OPEN

Evaluation of *Angelica decursiva* reference genes under various stimuli for RT-qPCR data normalization

Yuedong He^{1,2,4}, Yuan Zhong^{1,4}, Zhenzhen Bao¹, Weiqi Wang¹, Xiaoqing Xu¹, Yanan Gai³✉ & Jie Wu¹✉

Angelica decursiva is one of the leading traditional Chinese medicinal plants producing coumarins. Notably, several studies have focused on the biosynthesis and not the RT-qPCR (quantitative real-time reverse transcription polymerase chain reaction) study of coumarins. This RT-qPCR technique has been extensively used to investigate gene expression levels in plants and the selection of reference genes which plays a crucial role in standardizing the data from the RT-qPCR analysis. In our study, 11 candidate reference genes were selected from the existing transcriptome data of *Angelica decursiva*. Here, four different types of statistical algorithms (geNorm, NormFinder, BestKeeper, and Delta Ct) were used to calculate and evaluate the stability of gene expression under different external treatments. Subsequently, RefFinder analysis was used to determine the geometric average of each candidate gene ranking, and to perform comprehensive index ranking. The obtained results showed that among all the 11 candidate reference genes, SAND family protein (*SAND*), protein phosphatase 2A gene (*PP2A*), and polypyrimidine tract-binding protein (*PTBP*) were the most stable reference genes, where Nuclear cap binding protein 2 (*NCBP2*), TIP41-like protein (*TIP41*), and Beta-6-tubulin (*TUBA*) were the least stable genes. To the best of our knowledge, this work is the first to evaluate the stability of reference genes in the *Angelica decursiva* which has provided an important foundation on the use of RT-qPCR for an accurate and far-reaching gene expression analysis in this medicinal plant.

Angelica decursiva (Miq.) Franch. et Sav. is an important traditional medicinal plant from the genus *Angelica*, family Umbelliferae. The major bioactive components of this herbal drug are coumarins and this drug has been listed as one of the main special coumarin sources by the Chinese Pharmacopoeia. Coumarin has a core structure (2H-1-benzopyran-2-one core), that widely exists in higher plant species like the Rutaceae, Leguminosae, and Umbelliferae. Besides, they are slightly distributed in both animals and microorganisms¹. Recent pharmacological studies have proved that the 70% ethanol extract of *A. decursiva* has an anti-inflammatory activity^{2,3}, antioxidant^{4,5}, inhibitory effect on C6 rat glioma cells⁶ and inhibits the proliferation of osteogenic sarcoma cells⁷. Hence, certifying the supply of coumarin compounds from the drug has been a long-term practice⁸. Conversely, there is a huge shortage of coumarin raw materials, which has been attributed to low abundance, season, or region-dependent sourcing. Also, because of its complicated structure and multiple chiral centers⁹, there lacks a feasible method that has the capacity of mass chemical synthesis. Therefore, genetic and metabolic engineering strategies used in the production of coumarin compounds have become the main developing research direction. Subsequently, the study of functional genes that are present in their biosynthetic pathways is a study of interest.

In recent years, the rapid development of next-generation sequencing (NGS) technologies such as Roche / 454 and Illumina HiSeq platforms, has made it possible to study the distribution of mRNA and their expression levels in different tissues and cells. Currently, due to its high sensitivity, accurate quantification, and strong specificity, the RT-qPCR technology has become one of the most powerful tools for studying gene transcription levels. Besides, it quantifies the relative abundance and improves the quantitative accuracy of target genes^{10–12}. Nonetheless, the accuracy of this technique is affected by several factors¹³, like the amplification efficiency of the

¹School of Pharmacy, Jiangsu Health Vocational College, Nanjing, Jiangsu, China. ²College of Horticulture, Hunan Agricultural University, Changsha, Hunan, China. ³Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, Jiangsu, China. ⁴These authors contributed equally: Yuedong He and Yuan Zhong. ✉email: gaiyanan@cncbg.net; wujie1760@126.com

Gene abbreviation	Gene name	Arabidopsis homolog locus	Primer sequences (5'-3') (forward/reverse)	length	T _m (°C)	PCR efficiency	R ²
<i>CYP2</i>	Cyclophilin 2	AT4G33060	CGCCACTTTTTTTGTTCTCT/ TTGCGGATTATATCCGACA	108	82.9	1.711	0.971
<i>EXP-1</i>	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	AT2G32170	CCAAGTAGGAGCTGGGATG/ CACACCACCGTCTTTAGAA	109	81.7	1.789	0.958
<i>GAPDH</i>	Glyceraldehyde 3-phosphate dehydrogenase	AT1G42970	TGGTCCCTTAACGATACCA/ CTTACGTTGTTGGCGTGAAT	138	82.1	1.713	0.999
<i>NCBP2</i>	Nuclear cap binding protein 2	AT5G44200	TAGATGGCATGACAGGTGGA/ CAGGTACGCGATGAATATCG	145	85.6	1.333	0.993
<i>PP2A</i>	protein phosphatase 2A subunit A3	AT1G13320	GGTTGCCAATCAGCTCTATG/ AACTTGGTGACTTTGCCAGC	141	85	1.880	1.000
<i>PTBP</i>	Polypyrimidine tract-binding protein	AT3G01150	AATGTAAAGGCCCTTCAGCGA/ GGAGCAGCATGAGGATTCTG	110	85.2	1.847	1.000
<i>SAND</i>	SAND family protein	AT2G28390	GAGCCTCATGAATCCCTCAG/ CCCAAGCAAAGGTGTCATAT	126	82.1	1.845	0.972
<i>TIP41</i>	TIP41-like family protein	AT4G34270	CACCTGCATCAAAAGAGCCT/ GAAGAAACCAACAGCTTGGC	126	82.7	1.838	1.000
<i>UBQ10</i>	Polyubiquitin 10	AT4G05320	TTAAGCAGGGATCAAAACCC/ TTTGCTGGAAAGCAGCTAGA	124	84.6	1.721	0.888
<i>ACT</i>	Actin 2	AT3G18780	TGCTGGTCTGATCTCACTG/ GGTTTCAAGCTCTTGCTCGT	148	85.2	1.850	1.000
<i>TUBA</i>	Beta-6-tubulin	AT5G12250	CACCCAGCTTTGGTGATCTC/ AAGCCTAGGGAAGGGGATTA	134	86	1.685	0.999

Table 1. Details of primers used in this study.

target gene, the RNA yield, quality, and the reverse transcription efficiency that is present in different samples. Therefore, to overcome experimental errors and assure accuracy of experimental results, a certain number of internal reference genes, also called housekeeping genes, are often selected for calibration and standardization^{14,15}. These genes are usually used as a reference for either tissues or cells when they are stably expressed in a diversity of experimental conditions, and when changes are detected in the expression level of a target gene. Among the most commonly used internal reference genes for RT-qPCR are Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), Elongation factor-1 α (*EF-1 α*), Tubulin (*TUB*), β -Actin (*ACT*), 18S ribosomal RNA protein (*18S*), Polyubiquitin 10 (*UBQ10*), etc. These housekeeping genes play a significant role in the cell structure maintenance and during primary metabolic activities. Moreover, more internal reference genes have been identified from the gene expression chip data. They include protein phosphatase 2A (*PP2A*), SAND family protein (*SAND*), Tap42-interacting protein of 41 kDa (*TIP41*), etc. Consequently, the reference gene screening and evaluation have been accomplished for most species, such as rice¹⁶, *Lycoris aurea*¹⁷, radish¹⁸, potato¹⁹, *Peucedanum praeruptorum*²⁰, etc. However, no study has demonstrated a systematic selection of reference genes in *A. decursiva* under external challenges (abiotic stress and hormone treatments). The study of the coumarin biosynthesis pathway is one of our primary research interests relates. Here, previous studies have characterized some candidate genes that associate with *A. decursiva*⁸. However, it is imperative to find suitable reference genes, that enhance the detection of differential gene expression levels under various experimental conditions in *A. decursiva* using the RT-qPCR application.

In our study, 11 candidate reference genes (*SAND*, *PP2A*, *PTBP*, *ACT*, *CYP2*, *EXP-1*, *GAPDH*, *TUBA*, *NCBP2*, *UBQ10*, and *TIP41*) were selected about the transcriptome sequencing datasets of *A. decursiva*. Besides, 7 different forms of treatments such as cold (4 °C), drought (20% PEG 6000), methyl jasmonate (25 mM MeJA), salt (600 mM NaCl), oxidative (50 mM H₂O₂), ultraviolet (UV) induction, and metal (500 mM CuSO₄) were set. Next, the application of four statistical algorithms (geNorm, NormFinder, BestKeeper, and Delta Ct) to evaluate their expression stability for normalization, and comprehensive stability ranking was also performed by RefFinder. In conclusion, our work provides a basis for further studies on gene expression profiling and the regulation mechanisms of coumarin biosynthesis in *A. decursiva* under diverse experimental conditions.

Results

Selection of candidate reference genes, evaluation of amplification specificity and PCR efficiency. Here, 11 candidate reference genes were selected based on the transcriptome data of *A. decursiva* (unpublished). Table 1 lists all the candidate reference gene names and abbreviations, homologous genes from Arabidopsis, primer sequences, amplification length, annealing temperature (°C), PCR efficiency (E), and correlation coefficient (R²). Next, conventional PCR and RT-qPCR were used to verify the primer-specific amplification of all candidate reference genes. As illustrated in Fig. S1, based on agarose gel electrophoresis and melting curve analysis which showed a peak, the 11 primer pairs were amplified by a single specific amplicon. Here, the amplification efficiency range of *CYP2* and *PP2A* was 1.711 and 1.880 respectively whereas the correlation coefficient range of *UBQ10* was 0.888, and that of *PP2A*, *PTBP*, *TIP41*, and *ACT* was 1.000.

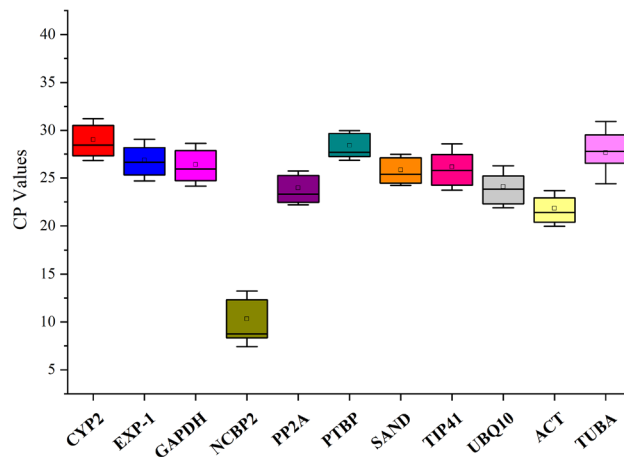


Figure 1. The RT-qPCR Cp values for 11 candidate reference genes in all samples. The expression data is shown as the Cp value of each reference gene in the samples of *A. decursiva*. Boxes indicate the 25th/75th percentiles, the lines represent the median, squares represent the means and whiskers represent the maximum and minimum values.

Expression profile of candidate reference genes. Here, the cycle threshold values (Cp) showed the number of cycles when the generated fluorescent signal reached a level that could be detected. Therefore, in this study, the expression profile of the candidate reference genes were directly determined using the obtained Cp values. As shown in Fig. 1, the average Cp values of these 11 reference genes are distributed between 10 and 30, and a majority of them are distributed between 23 and 27. Among all the candidate reference genes, *NCBP2* had the lowest average Cp values, whereas *CYP2*, *PTBP*, and *TUBA* had the highest average Cp values. Moreover, each reference gene had a different coefficient of variation under different conditions. Also, it was observed that *SAND* and *PTBP* had the lowest variability, whereas *TUBA*, which had the highest Cp value, had the highest variability. This value ranged between 25 and 32. Therefore, it may not be qualified as a reference gene. Through the Cp value combined with box-plot not only display the expression profile of the reference gene, but also give us a glimpse in their stability (Table S2). However, considering the complexity of their surroundings, we need to check the proper use of the references in every particular experimental condition. Thus, more statistical tools and further analyses are required.

Expression stability of candidate reference genes. Based on the relative expression levels of the 11 candidate reference genes, four algorithms such as BestKeeper, geNorm, NormFinder, and Delta Ct were used to examine their expression stability. Subsequently, the RefFinder tool was employed to sequence the expression stability of all these candidate reference genes and select the most suitable ones.

geNorm analysis: The geNorm analysis evaluated the stability of all the 11 candidate reference genes using the M value (reference expression stability measure). These M values were calculated from the average variation of the gene relative against other candidate reference genes, and the lower M values indicated a higher gene expression stability²¹. As illustrated in Fig. 2, all the candidate reference genes had different levels of stability under different treatments. Here, the M value of *PP2A* and *SAND* was the lowest in most treatments and was deliberated as the most stable reference genes. Besides, *ACT* showed good stability under H₂O₂, MeJA, NaCl, and UV treatments, whereas *CYP2* was one of the most stable reference genes in the CuSO₄ and UV treatment groups. On the other hand, the stability of *NCBP2* seemed to be less satisfactory than other candidate reference genes. Subsequently, the geNorm algorithm can also determine the optimal number of normalized reference genes²¹ by calculating pairwise variations ($V_{n/n+1}$). Normally, the ideal paired variation (V) score is less than 0.15, which means that the addition of any other gene will not have a substantial impact on standardization. In our study, in the NaCl, Cold, UV, H₂O₂, CuSO₄, and WT subsets, their pairwise variation values ($V_{2/3}$) were all less than 0.15. This showed that the addition of a third reference gene lacked significant effects on the normalization of the target gene, thus, the number of the optimal reference genes that were determined was two. In contrast, as shown in Fig. 3 and Table S3, the pairwise variation values ($V_{3/4}$) of the PEG and MeJA subsets, was also less than 0.15. This indicated that the number of reference genes that were most suitable for these two treatments was three.

NormFinder analysis: To normalize raw data and measure the expression stability of candidate reference genes through intra- and inter-group variation, the NormFinder analysis uses the $2^{-\Delta Ct}$ method. Like the geNorm analysis, its lower values show higher stability. Table 2 shows the ranking of all candidate genes as calculated using the NormFinder algorithm. Among the PEG, Cold, and 'total' treatment subsets, it was observed that *SAND* was the most stable candidate gene, and also ranked higher in other subsets. After MeJA treatment (0.078) on all sample subsets, *TUBA* was deliberated as the most stable candidate gene. Besides, among the other five groups (NaCl, UV, H₂O₂, CuSO₄, and WT), *PTBP* (0.670), *TIP41* (0.307), *GAPDH* (0.495), *CYP2* (0.483), and *UBQ10* (0.451) were observed to be the most stable genes, respectively. Similar to the geNorm analysis, our study found that the *NCBP2* was the most unstable candidate gene.

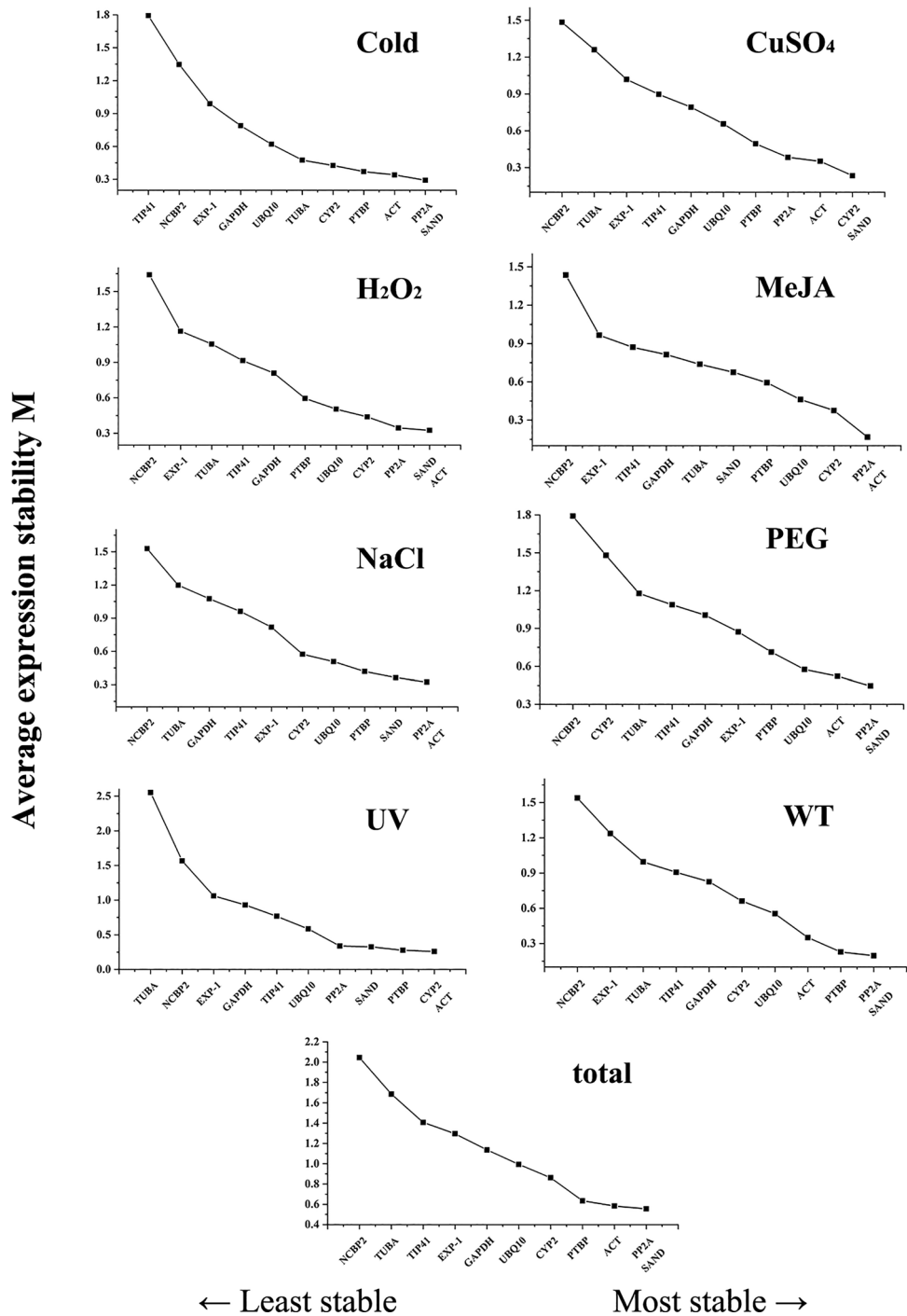


Figure 2. Average expression stability values (M) of the 11 candidate reference genes using geNorm software. Expression stability was evaluated in samples from *A. decursiva* was submitted to cold stress, drought stress (20% PEG), Methyl jasmonate (MeJA) stress, salt stress (0.5 M NaCl), oxidative stress (H₂O₂), ultraviolet (UV) induction, Metal stress (CuSO₄), untreated (WT) and 'total' (all treated samples). The least stable genes are on the left with higher M-value and the most stable genes on the right with lower M-value.

BestKeeper analysis: Here, the raw data of the Ct value was directly calculated and the more stable candidate reference genes was evaluated by calculating the standard deviation (SD) and the Ct value²². Notably, lower SD and Ct values indicated higher gene expression stability, particularly when the SD was greater than 1 which indicated that the reference gene was unstable²³ and could not be utilized for normalization. Table 3 shows that in the NaCl, MeJA, UV, WT, and 'total' treatment subsets, the *PTBP* gene was considered the most stable. However, the MeJA and 'total' treatment subsets were postulated to be expression-insensitive since their SD values were exceed 1 and could not be used for normalization. For this reason, such values should be excluded. Besides,

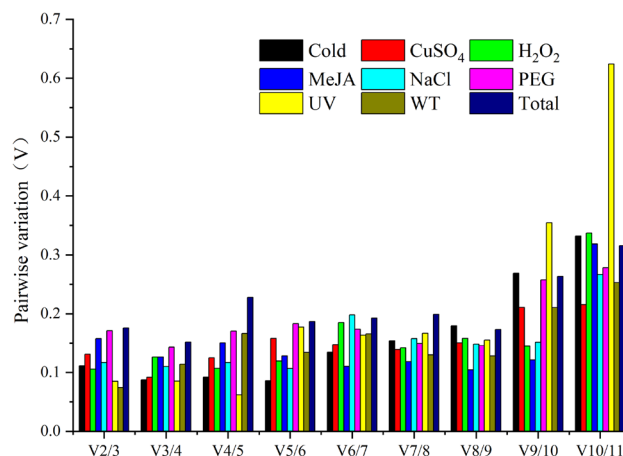


Figure 3. Determination of the optimal numbers of reference genes for normalization by pairwise variation ($V_{n/n+1}$). Different treatments are marked as square frame with different colors. The cut-off value is 0.15 and used to determine the optimal number of candidate reference genes for qRT-PCR normalization.

Rank	NaCl	PEG	Cold	MeJA	UV	H ₂ O ₂	CuSO ₄	WT	Total
1	PTBP (0.670)	SAND (0.319)	SAND (0.146)	TUBA (0.078)	TIP41 (0.307)	GAPDH (0.495)	CYP2 (0.483)	UBQ10 (0.451)	SAND (0.568)
	PP2A (0.672)	PP2A (0.399)	ACT (0.313)	PTBP (0.452)	PTBP (0.606)	PTBP (0.528)	SAND (0.500)	SAND (0.656)	PP2A (0.624)
3	ACT (0.676)	UBQ10 (0.560)	TUBA (0.407)	SAND (0.472)	CYP2 (0.615)	TIP41 (0.594)	UBQ10 (0.535)	GAPDH (0.672)	PTBP (0.696)
	SAND (0.714)	ACT (0.616)	PP2A (0.418)	GAPDH (0.502)	PP2A (0.744)	PP2A (0.706)	GAPDH (0.571)	TIP41 (0.692)	ACT (0.772)
5	TIP41 (0.719)	EXP-1 (0.739)	PTBP (0.493)	UBQ10 (0.502)	SAND (0.776)	SAND (0.766)	ACT (0.597)	PTBP (0.699)	GAPDH (0.918)
	UBQ10 (0.730)	PTBP (0.869)	CYP2 (0.667)	PP2A (0.641)	ACT (0.851)	CYP2 (0.855)	PP2A (0.604)	PP2A (0.728)	UBQ10 (1.198)
7	GAPDH (0.999)	GAPDH (0.957)	GAPDH (0.998)	ACT (0.675)	GAPDH (0.954)	UBQ10 (0.960)	TIP41 (0.773)	CYP2 (0.815)	CYP2 (1.275)
	CYP2 (1.014)	TIP41 (1.086)	UBQ10 (1.032)	TIP41 (0.699)	UBQ10 (1.176)	ACT (1.044)	PTBP (1.075)	ACT (1.011)	EXP-1 (1.373)
9	EXP-1 (1.075)	TUBA (1.398)	EXP-1 (1.505)	CYP2 (1.085)	EXP-1 (1.210)	TUBA (1.284)	EXP-1 (1.222)	TUBA (1.089)	TIP41 (1.422)
	TUBA (1.528)	CYP2 (2.677)	NCBP2 (2.711)	EXP-1 (1.324)	NCBP2 (3.510)	EXP-1 (1.291)	TUBA (2.200)	EXP-1 (2.013)	TUBA (2.635)
11	NCBP2 (2.918)	NCBP2 (3.037)	TIP41 (3.636)	NCBP2 (3.498)	TUBA (6.852)	NCBP2 (3.698)	NCBP2 (2.350)	NCBP2 (2.764)	NCBP2 (3.449)

Table 2. Stability rank of 11 candidate genes by NormFinder.

under H₂O₂, CuSO₄, Cold, and PEG treatments, the most stable genes were observed to be CYP2, SAND, TUBA, and PP2A. Generally, as illustrated in Table 3, it was observed that the NCBP2 gene was still the most unstable under all treatments.

Delta Ct analysis: This method evaluated gene expression stability by calculating the mean standard deviation (SD) of each gene. Here, the smaller the value, the higher the stability²⁴. As shown in Table S4, the results of this analysis are consistent with the geNorm analysis. The only difference is that in the Cold and H₂O₂ subsets, TUBA and PP2A candidate genes are the most stable respectively. According to the geNorm analysis, the two most stable candidate genes are SAND and ACT. Hence, SAND and PP2A are the most qualified reference genes.

Reffinder analysis: As shown in Fig. 4, we further calculated the geometric mean of the ranking of each candidate gene using the Reffinder algorithm (<http://150.216.56.64/referencegene.php#>). This was based on the results obtained from the three statistical algorithms such as geNorm, NormFinder, and BestKeeper. Tables S5 and S6, show the comprehensive index ranking, whereby, the smaller the index, the more stable the gene expression¹⁹.

Rank	NaCl	PEG	Cold	MeJA	UV	H ₂ O ₂	CuSO ₄	WT	Total
1	<i>PTBP</i>	<i>PP2A</i>	<i>TUBA</i>	<i>PTBP</i>	<i>PTBP</i>	<i>CYP2</i>	<i>SAND</i>	<i>PTBP</i>	<i>PTBP</i>
	0.97 ± 0.29	2.73 ± 0.65	2.50 ± 0.68	4.22 ± 1.23	0.54 ± 0.15	3.45 ± 1.07	1.04 ± 0.26	1.10 ± 0.30	4.63 ± 1.32
2	<i>PP2A</i>	<i>PTBP</i>	<i>CYP2</i>	<i>SAND</i>	<i>SAND</i>	<i>SAND</i>	<i>CYP2</i>	<i>SAND</i>	<i>SAND</i>
	1.76 ± 0.46	2.89 ± 0.82	2.76 ± 0.80	4.92 ± 1.30	0.87 ± 0.22	3.70 ± 1.03	1.13 ± 0.31	1.17 ± 0.28	5.40 ± 1.40
3	<i>SAND</i>	<i>SAND</i>	<i>PTBP</i>	<i>TUBA</i>	<i>PP2A</i>	<i>TIP41</i>	<i>UBQ10</i>	<i>ACT</i>	<i>EXP-1</i>
	2.23 ± 0.63	3.17 ± 0.81	3.44 ± 0.94	6.08 ± 1.70	0.89 ± 0.20	3.98 ± 1.13	1.83 ± 0.42	1.19 ± 0.23	6.33 ± 1.70
4	<i>ACT</i>	<i>EXP-1</i>	<i>GAPDH</i>	<i>TIP41</i>	<i>CYP2</i>	<i>PP2A</i>	<i>PP2A</i>	<i>PP2A</i>	<i>PP2A</i>
	2.32 ± 0.56	3.88 ± 1.04	3.92 ± 0.98	6.30 ± 1.69	1.21 ± 0.33	4.27 ± 1.11	2.03 ± 0.47	1.34 ± 0.30	6.34 ± 1.52
5	<i>CYP2</i>	<i>UBQ10</i>	<i>ACT</i>	<i>UBQ10</i>	<i>ACT</i>	<i>PTBP</i>	<i>ACT</i>	<i>CYP2</i>	<i>CYP2</i>
	2.48 ± 0.77	4.52 ± 1.05	4.07 ± 0.86	6.35 ± 1.65	1.49 ± 0.32	4.27 ± 1.30	2.51 ± 0.51	2.73 ± 0.73	6.35 ± 1.84
6	<i>UBQ10</i>	<i>GAPDH</i>	<i>SAND</i>	<i>GAPDH</i>	<i>TIP41</i>	<i>GAPDH</i>	<i>PTBP</i>	<i>UBQ10</i>	<i>GAPDH</i>
	2.91 ± 0.74	4.52 ± 1.19	4.31 ± 1.08	6.39 ± 1.71	3.90 ± 0.98	4.44 ± 1.29	2.54 ± 0.71	3.20 ± 0.68	6.88 ± 1.82
7	<i>EXP-1</i>	<i>ACT</i>	<i>PP2A</i>	<i>CYP2</i>	<i>UBQ10</i>	<i>ACT</i>	<i>TIP41</i>	<i>TUBA</i>	<i>TIP41</i>
	3.16 ± 0.88	4.81 ± 1.05	5.04 ± 1.17	6.46 ± 1.94	3.94 ± 0.87	4.87 ± 1.18	3.09 ± 0.74	3.63 ± 0.92	7.27 ± 1.90
8	<i>TIP41</i>	<i>TIP41</i>	<i>EXP-1</i>	<i>PP2A</i>	<i>GAPDH</i>	<i>UBQ10</i>	<i>GAPDH</i>	<i>TIP41</i>	<i>ACT</i>
	3.96 ± 1.12	5.19 ± 1.35	5.15 ± 1.39	6.80 ± 1.70	4.52 ± 1.15	5.05 ± 1.34	3.30 ± 0.83	4.05 ± 0.99	7.30 ± 1.59
9	<i>TUBA</i>	<i>TUBA</i>	<i>UBQ10</i>	<i>ACT</i>	<i>EXP-1</i>	<i>TUBA</i>	<i>EXP-1</i>	<i>GAPDH</i>	<i>UBQ10</i>
	4.32 ± 1.28	5.87 ± 1.64	5.91 ± 1.47	7.24 ± 1.62	5.09 ± 1.36	6.37 ± 1.93	4.10 ± 1.03	4.15 ± 1.01	7.30 ± 1.76
10	<i>GAPDH</i>	<i>CYP2</i>	<i>TIP41</i>	<i>EXP-1</i>	<i>TUBA</i>	<i>EXP-1</i>	<i>TUBA</i>	<i>EXP-1</i>	<i>TUBA</i>
	4.36 ± 1.27	6.49 ± 1.93	9.03 ± 2.34	7.96 ± 2.20	15.44 ± 3.98	6.71 ± 1.93	5.34 ± 1.44	7.75 ± 1.94	7.30 ± 2.02
11	<i>NCBP2</i>	<i>NCBP2</i>	<i>NCBP2</i>	<i>NCBP2</i>	<i>NCBP2</i>	<i>NCBP2</i>	<i>NCBP2</i>	<i>NCBP2</i>	<i>NCBP2</i>
	23.24 ± 2.56	21.11 ± 2.17	17.41 ± 1.52	21.69 ± 2.06	27.51 ± 3.12	24.04 ± 4.27	19.57 ± 1.98	22.34 ± 2.55	8.19 ± 0.82

Table 3. Rank of 11 candidate genes by BestKeeper.

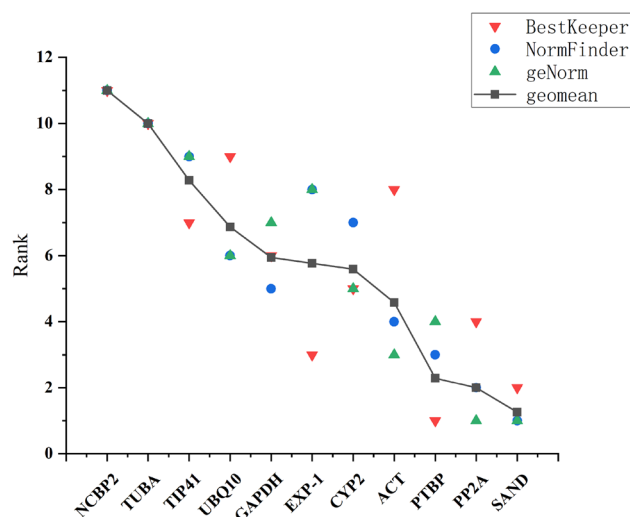


Figure 4. Comprehensive ranking of BestKeeper, NormFinder, and geNorm. Using BestKeeper, NormFinder, and geNorm algorithms, the 11 candidate reference genes were ranked comprehensively, and the geometric mean of each reference gene was calculated.

This study showed that *SAND* and *PP2A* ranked the highest in most subsets, whereas *NCBP2* and *TUBA* ranked the lowest, making them the most unstable reference genes. In contrast, in the MeJA and Cold subsets, *TUBA* seemed to be a relatively stable reference gene. Despite the different assessment methods, this resulting difference is reasonable and acceptable. In summary, the stability of these 11 candidate reference genes from the highest to the lowest is: *SAND*, *PP2A*, *PTBP*, *ACT*, *CYP2*, *EXP-1*, *GAPDH*, *UBQ10*, *TIP41*, *TUBA*, and *NCBP2*. These results were similar to those obtained from the geNorm and NormFinder analysis, but slightly different from those of the BestKeeper analysis.

Discussion

Currently, a majority of medicinal plants enhance the bioactivity of their medicinal components by inducing and regulating functional genes present in their biosynthetic pathway²⁵. Several reports have shown that in plants, there is a significant correlation between the synthesis and accumulation of secondary metabolites and the expression levels found in the functional pathways of their biosynthetic pathways^{26,27}. Coumarins are one of the main active components extracted from *A. decursiva*, which is one of the main sources of coumarin found in China and is listed as a special coumarin resource by the Chinese Pharmacopoeia. However, there exist no previous reports on RT-qPCR studies, which confines the research on the biosynthetic pathway of coumarin compounds from *A. decursiva*.

Here, we screened a total of 11 potential reference genes from the transcriptome data of *A. decursiva* (unpublished). First, as shown in Table 1 and S1, their primer specificity and PCR experimental conditions were confirmed, whereby the single peak of the dissolution curve exposed the specificity of these 11 pairs of primers (Fig. S1). As illustrated in Fig. 1 and Table S2, the average Ct-values for the 11 candidate reference genes is between 8.04 and 32.65, and the majority of these values were found to be between 23 and 27, which means that most of the 11 candidate reference genes were likely to afford accurate normalization²⁰. The *SAND* and *PTBP* genes have the lowest variability, which suggests that they could have stable expression levels under different treatments. Conversely, considering the complexity of the ambient environment, the expression stability of these reference genes under different conditions must be investigated further²⁸. Therefore, there is a need to use more statistical tools and conduct further analyses.

To ensure the accuracy and reliability of experimental data, four commonly used statistical algorithms (geNorm, NormFinder, BestKeeper, and Delta Ct) were combined for reference gene selection. Subsequently, RefFinder analysis was used to calculate the geometric ranking mean of each candidate reference gene, and then a comprehensive index ranking was done. Here, the results from this analysis showed that the ranking results acquired using different statistical algorithms were not similar, and the candidate reference genes also showed different levels of stability under different treatments. For example, Table 2 shows that *TUBA* is the best reference gene in the MeJA subset, and this is consistent with the results obtained when *Atropa belladonna*²⁹ was exposed to different hormone treatments which showed that it had the most stable expression. On the other hand, when this gene was exposed to other treatments, its stability was unsatisfactory. This is most likely because some genes belonging to the tubulin family and that express isomers in specific developmental stages or tissues are regulated by different developmental processes³⁰. Similar to the results from tomato³¹ and *Arabidopsis thaliana*³², *UBQ10* is the most stable gene in the WT group (Table 2). Also, the *PTBP* gene performed well in the UV subset, unlike in the Cold, PEG, and CuSO_4 which is illustrated on Table 2 and S5). This result contrasted with the most stable performance of the *PTBP* gene in Cold, PEG, and CuSO_4 subsets in *Peucedanum praeruptorum*²⁰. Table 3 shows that under cold stress (4 °C) and the 'total' group, the *TUBA* and *PTBP* genes were the most stable reference genes, respectively, but this was different from the results obtained from geNorm and NormFinder analyses which showed that *SAND* was the gene with the best stability. Since various calculation principles are inconsistent, even under the same conditions, the ranking of these genes could be different. Therefore, the single software analysis has certain disadvantages when evaluating the stability of the reference gene. Our study recommends that a comprehensive evaluation and analysis need to be used to ensure the reliability and accuracy of the results. Besides, it is necessary to flexibly select the most appropriate reference gene for different experimental treatments.

Also, bearing in mind that a single reference gene could cause errors in the expression level of the target gene^{33,34}, the geNorm statistical algorithm calculates the pairwise variation ($V_{n/n+1}$) to attain the optimal number of reference genes³⁵. Notably, the ideal pairwise variation value must be lower than the critical value of 0.15^{33,36}. For instance, as illustrated in Fig. 3 and Table S3, the pairwise variation $V_{2/3}$ in the Cold, NaCl, CuSO_4 , H_2O_2 , UV, WT, and the 'total' treatment subsets, is less than 0.15. This demonstrated that only two genes are required to complete the gene expression normalization under these conditions. Consequently, it was observed that the *PP2A* and *ACT* are relatively stable reference genes under the treatment of MeJA, their pairing value $V_{3/4}$ was < 0.15, and their optimal number of reference genes was three.

Taken together, the analysis results from the three algorithms (geNorm, NormFinder, and Delta Ct) seemed to be consistent, whereas those of BestKeeper are different. Figure 4 and Table S5, confirm that under most external conditions, the stability of *SAND* and *PP2A* genes are significantly better than that of other reference genes, whereas *NCBP2* and *TUBA* are nearly always defined as the most unstable reference genes. Of note, the *SAND* and *PP2A* are new reference genes that were identified through screening the data from the *Arabidopsis* gene chip. To date, many studies choose the *SAND* and *PP2A* genes for their reference gene standardization studies. For instance, studies that used *Arabidopsis thaliana*³², tomato³⁷, *Petunia hybrida*³⁸, *Peucedanum praeruptorum*²⁰, etc. Conventionally, reference genes, such as *TUBA*, *GAPDH*, *UBQ10*, and *ACT* play a housekeeping role in the maintenance of cell structure or primary metabolic activities. Nevertheless, increasing research indicates that the expression levels in most of these traditional reference genes can somehow vary greatly and in many species, they are unsuitable for gene normalization under specific conditions^{39,40}. Therefore, given a subset of experimental conditions, the housekeeping genes used as reference genes in each different species should be handled carefully^{41,42}.

Materials and methods

Samples preparation and treatments. Here, one-year-old plants of *A. decursiva* were collected from Ningguo City, Anhui Province, China (longitude: 118.95E, latitude: 30.62 N), and was identified as *Angelica decursiva* (Miq.) Franch. et Sav. by Zhang Ning, school of pharmacy, Jiangsu Health Vocational College, and deposited in the herbarium of the medicinal botanical garden (ID: JSJK-AD-021). The planting field of *A. decursiva* was a private land and the landowner has allowed us to collect plant materials for further study. The collec-

tion of plant material is in full compliance with relevant institutions, national and international guidelines and legislation. Then, these accessions were transplanted into plastic pots that contained a mixture of vermiculite, perlite, and peat moss (1:1:1 v/v). Next, the plants were grown in a greenhouse at a temperature of 25 °C, a long photoperiod of 16 h light and 8 h darkness, 40–65% relative humidity, and 3000 lx light intensity until treated. For drought treatments, a 200 mL solution of 25% PEG 6000 (w / v, polyethylene glycol, Sangon, China) was used to treat the plants for one week. For salt stress treatment, an approximate amount of 200 mL (600 mM) of NaCl was used to water the plants for 7 days. For cold stimulation treatments, the plants were incubated at 4 °C for 48 h. To study hormone therapy, 25 mM MeJA was applied for 6 h. To assess heavy metal stress, 500 mM of Copper sulfate (CuSO₄) was applied to the plants for 24 h. In the case of oxidative stress, 50 mM H₂O₂ was used for 24 h. To induce ultraviolet (UV) light, these plants were irrigated using distilled water (100 mL) and then exposed to ultraviolet light for 24 h. Notably, all the above treatments had three biological replicates. The remaining plants that received no treatment served as controls. Lastly, all samples from each treatment were washed with MINIQ- filtered water, quickly frozen in liquid nitrogen, and stored at –80 °C.

Total RNA isolation and cDNA synthesis. About 100 mg of the different frozen tissue samples were used to extract total RNA using the Spectrum Plant Total RNA kit (Sigma, USA). Next, the quality and purity of the extracted total RNA was determined using the NanoDrop spectrophotometer 2000 (Thermo Scientific, USA), and its integrity confirmed on a 1.5% agarose gel. Here, only RNA with A260/280 ratios between 1.8 and 2.2 and an A260/230 above 2.0 were used for cDNA synthesis. Subsequently, RNA samples were pre-treated with RNase-free DNase I (Takara Biotechnology, Dalian, China) to remove contaminating traces of genomic DNA and then used for reverse transcription. Lastly, following the instructions from the HiScript Q RT SuperMix for qPCR (Vazyme, China) was used for first-strand cDNA synthesis with oligo (dT) as the primer.

Primer design and RT-qPCR conditions. Here, a total of 11 candidate reference genes were selected from the transcriptome data of *A. decursiva* (unpublished). Next, the built-in program in the Bioedit Sequence Alignment Editor (v7.0.9) was used to screen and select potential single genes using local blast (TBLASTN). Subsequently, the corresponding homologs of these reference genes were selected from the database of The Arabidopsis Information Resource (TAIR) (<http://www.arabidopsis.org>), and only single genes with lower E-values and higher scores were selected for subsequent analysis. Table S1, shows the reference gene IDs, homologous loci, gene sequences, and the different expression levels (FPKM, Fragments Per Kilobase per Million) of all the 11 candidate reference genes.

To avoid amplification of any contaminating genomic DNA, primers were designed to cross at least one intron/exon border that contained both donor and acceptor sites, and then exon analysis was performed using the AlignX program in the vector NTI advance 11.5 package. Subsequently, the Primer3Plus (<http://primer3plus.com/cgi-bin/dev/primer3plus.cgi>) was used to design primers with the following characteristics: amplicon length was 100 to 150 bp, GC content was between 40 and 60%, primer length was 18 to 24 bp, temperature difference of each primer pair was less than 1 °C, and the melting temperature (T_m) was between 55 and 60 °C. Consequently, all the primer pairs were tested using conventional PCR to determine the combination of forward and reverse primers that performed optimally and the resulting products were examined using 1.0% agarose gel electrophoresis. Besides, from a series of standard curves of five different cDNA dilutions, the amplification efficiency (E) and correlation coefficient (R²) were calculated. Table 1 lists all the gene-specific primer pairs that were designed and used in the RT-qPCR analysis.

Stability evaluation of candidate reference genes. To obtain the RT-qPCR data, three biological and technical replicates were done for each sample, and all data presented as mean ± standard error of the mean (SEM). Consequently, statistical analyses were performed using the Student's *t*-test. Next, representative graphs were generated using OriginPro 9.1 (OriginLab Corporation, Northampton, MA, USA). Subsequently, data analysis was performed using GeNorm (ver.3.5), NormFinder (ver.0.953), BestKeeper (ver.1.0), and the Delta Ct method following the instructions from the manufacturer. Lastly, a comprehensive stability ranking analysis was performed using RefFinder (<http://150.216.56.64/referencegene.php>).

Conclusions

Selecting appropriate reference genes is a significant prerequisite for quantifying gene expression using RT-qPCR. This is a conducive research on the key enzyme genes that are involved in the biosynthesis of coumarins and other interesting secondary metabolites found in *A. decursiva*. So far, this is the first systematic screening experiment of the most suitable reference genes for *A. decursiva* under various external treatments, like cold stress (4 °C), drought stress (20% PEG), Methyl jasmonate (MeJA), salt stress (0.5 M NaCl), oxidative stress (H₂O₂), ultraviolet (UV) induction, metal stress (CuSO₄), untreated (WT), and 'total' (all treated samples). Our results have exposed that the 11 candidate genes have different stability in *A. decursiva* when exposed to different experimental treatments. From the overall stability ranking, *SAND* is the most stable candidate reference gene, followed by *PP2A* and then *PTBP*. On the other hand, the *NCBP2* gene has the lowest stability making it unsuitable for further research. In summary, the reference genes evaluated in this study can be helpful for accurate normalization of the RT-qPCR data and any other future work on the gene expression of coumarin synthesis present in *A. decursiva*.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 12 November 2020; Accepted: 17 August 2021

Published online: 23 September 2021

References

- Ojala, T. *et al.* Antimicrobial activity of some coumarin containing herbal plants growing in Finland. *J. Ethnopharmacol.* **73**, 299–305 (2000).
- Lim, H. J. *et al.* Inhibition of airway inflammation by the roots of *Angelica decursiva* and its constituent, columbianadin. *J. Ethnopharmacol.* **155**, 1353–1361 (2014).
- Islam, M. N. *et al.* Mechanism of anti-inflammatory activity of umbelliferone 6-carboxylic acid isolated from *Angelica decursiva*. *J. Ethnopharmacol.* **144**, 175–181 (2012).
- Zhao, D. *et al.* In vitro antioxidant and anti-inflammatory activities of *Angelica decursiva*. *Arch. Pharm. Res.* **35**, 179–192 (2012).
- Ali, M. Y. *et al.* Coumarins from *Angelica decursiva* inhibit α -glucosidase activity and protein tyrosine phosphatase 1B. *Chem. Biol. Interact.* **252**, 93–103 (2016).
- Cho, S. H. *et al.* Induction of apoptosis by *Angelica decursiva* extract is associated with the activation of caspase in glioma cells. *J. Korean Soc. Appl. Bi.* **52**, 241–246 (2009).
- Lee, S. W. *et al.* The effects of *Angelica decursiva* extract in the inhibition of cell proliferation and in the induction of apoptosis in osteogenic sarcoma cells. *J. Med. Plants Res.* **3**, 241–245 (2009).
- Zhao, Y. C. *et al.* Elucidation of the biosynthesis pathway and heterologous construction of a sustainable route for producing umbelliferone. *J. Biol. Eng.* **13**, 44 (2019).
- Feher, M. & Schmidt, J. M. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Cheminform.* **34**, 218 (2003).
- Dekkers, B. J. *et al.* Identification of reference genes for RT-qPCR expression analysis in Arabidopsis and tomato seeds. *Plant Cell Physiol.* **53**, 28–37 (2012).
- Maltseva, D. V. *et al.* High-throughput identification of reference genes for research and clinical RT-qPCR analysis of breast cancer samples. *J. Clin. Bioinform.* **3**, 13 (2013).
- Sun, H. P., Jiang, X. F., Sun, M. L., Cong, H. Q. & Qiao, F. Evaluation of reference genes for normalizing RT-qPCR in leaves and suspension cells of *Cephalotaxus hainanensis* under various stimuli. *Plant Methods* **15**, 31 (2019).
- 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).
- Suzuki, T., Higgins, P. J. & Crawford, D. Control selection for RNA quantitation. *Biotechniques* **29**, 332–337 (2000).
- Bustin, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* **29**, 23–39. <https://doi.org/10.1677/jme.0.0290023> (2002).
- Pabuayon, I. M. *et al.* Reference genes for accurate gene expression analyses across different tissues, developmental stages and genotypes in rice for drought tolerance. *Rice* **9**, 32 (2016).
- Ma, R., Xu, S., Zhao, Y. C., 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).
- Duan, M. 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).
- Tang, X., Zhang, N., Si, H. & Calderón-Urrea, A. Selection and validation of reference genes for RT-qPCR analysis in potato under abiotic stress. *Plant Methods* **13**, 85 (2017).
- Zhao, Y. C. *et al.* Selection of reference genes for gene expression normalization in *Peucedanum praeruptorum* dunn under abiotic stresses, hormone treatments and different tissues. *PLoS ONE* **11**, e0152356 (2016).
- Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 1–11. <https://doi.org/10.1186/gb-2002-3-7-research0034> (2002).
- 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).
- Xiao, X. *et al.* Validation of suitable reference genes for gene expression analysis in the halophyte *Salicornia europaea* by real-time quantitative PCR. *Front. Plant Sci.* **5**, 788 (2014).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408. <https://doi.org/10.1006/meth.2001> (2001).
- Kumar, P. *et al.* Expression analysis of biosynthetic pathway genes vis-à-vis podophyllotoxin content in *Podophyllum hexandrum* Royle. *Protoplasma* **252**, 1253–1262 (2015).
- Olofsson, L., Engström, A., Lundgren, A. & Brodelius, P. E. Relative expression of genes of terpene metabolism in different tissues of *Artemisia annua* L. *BMC Plant Biol.* **11**, 45 (2011).
- Lau, W. & Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **34**, 1224–1228 (2015).
- Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).
- Li, J. D. *et al.* Reference gene selection for gene expression studies using quantitative real-time PCR normalization in *Atropa belladonna*. *Plant Mol. Biol. Rep.* **32**, 1002–1014 (2014).
- Montoliu, L., Rigau, J. & Puigdomènech, P. A tandem of α -tubulin genes preferentially expressed in radicular tissues from *Zea mays*. *Plant Mol. Biol.* **14**, 1–15 (1990).
- Løvdaal, 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).
- 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. <https://doi.org/10.2307/4281837> (2005).
- Gu, C. *et al.* Reference gene selection for quantitative real-time PCR in Chrysanthemum subjected to biotic and abiotic stress. *Mol. Biotechnol.* **49**, 192–197 (2011).
- Veazey, K. J. & Golding, M. C. Selection of stable reference genes for quantitative rtPCR comparisons of mouse embryonic and extra-embryonic stem cells. *PLoS ONE* **6**, e27592 (2011).
- Sgamma, T., Pape, J., Massiah, A. & Jackson, S. Selection of reference genes for diurnal and developmental time-course real-time PCR expression analyses in lettuce. *Plant Methods* **12**, 21. <https://doi.org/10.1186/s13007-016-0121-y> (2016).
- Huang, Y. X. *et al.* Stable internal reference genes for normalizing real-time quantitative PCR in *Baphicacanthus cusia* under hormonal Stimuli and UV irradiation, and in different plant organs. *Front. Plant Sci.* **8**, 668 (2017).
- Expósito-Rodríguez, M., Borges, A. A., Borges-Pérez, A. & Pérez, J. A. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biol.* **8**, 131–142 (2008).
- Mallona, I., Lischewski, S., Weiss, J., Hause, B. & Egea-Cortines, M. Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. *BMC Plant Biol.* **10**, 4 (2010).
- Die, J. V., Román, B., Nadal, S. & González-Verdejo, C. I. Evaluation of candidate reference genes for expression studies in *Pisum sativum* under different experimental conditions. *Planta* **232**, 145–153. <https://doi.org/10.2307/23391626> (2010).

40. Zhu, J. *et al.* Reference gene selection for quantitative real-time PCR normalization in *Caragana intermedia* under different abiotic stress conditions. *PLoS ONE* **8**, e53196 (2013).
41. Ma, S. *et al.* Expression stabilities of candidate reference genes for RT-qPCR under different stress conditions in soybean. *PLoS ONE* **8**, e75271 (2013).
42. Lin, Y. *et al.* Validation of potential reference genes for qPCR in Maize across abiotic stresses, hormone treatments, and tissue types. *PLoS ONE* **9**, e95445 (2014).

Acknowledgements

This project is funded by Practical Training Program for Natural Science Foundation of the Jiangsu Higher Education Institutions of China (16KJB150042, 19KJD430005). This Project was also funded by Qinglan project of excellent teaching team in Jiangsu and teaching and research project of Jiangsu Health Vocational College (JKKYTD201701, JKA201706, JKA201812, JKB201839, JKA201902, JKB201911) and the Jiangsu "Six one" Project of Health top talent program (LGY2018089).

Author contributions

Y.H., Y.G., and J.W. conceived and designed the experiments. Y.H. and Y.Z. performed the experiments. Y.Z., Z.B., and W.W. analyzing the data. Z.B., W.W., and X.X. contributed reagents, materials, and analysis tools. Y.H., Y.G., and J.W. wrote and revised the paper. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98434-6>.

Correspondence and requests for materials should be addressed to Y.G. or J.W.

Reprints and permissions information is available at www.nature.com/reprints.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021