

# Comprehensive transcriptome analysis of reference genes for fruit development of *Euscaphis konishii*

Cheng-Long Yang<sup>1,\*</sup>, Xue-Yan Yuan<sup>2,3,\*</sup>, Jie Zhang<sup>1</sup>, Wei-Hong Sun<sup>2,3</sup>, Zhong-Jian Liu<sup>2,3</sup> and Shuang-Quan Zou<sup>2,3</sup>

<sup>1</sup> Biotechnology Institute, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian, China

<sup>2</sup> Fujian Colleges and Universities Engineering Research Institute of Conservation and Utilization of Natural Bioresources at College of Forestry, Fujian Agriculture and Forestry University, Fujian, Fuzhou, China

<sup>3</sup> Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

\* These authors contributed equally to this work.

## ABSTRACT

**Background.** Quantitative real-time reverse transcriptase polymerase chain reaction is the common method to quantify relative gene expression. Normalizing using reliable genes is critical in correctly interpreting expression data from qRT-PCR. *Euscaphis konishii* is a medicinal plant with a long history in China, which has various chemical compounds in fruit. However, there is no report describing the selection of reference genes in fruit development of *Euscaphis konishii*.

**Methods.** We selected eight candidate reference genes based on RNA-seq database analysis, and ranked expression stability using statistical algorithms GeNorm, NormFinder, BestKeeper and ReFinder. Finally, The nine genes related to the anthocyanin synthesis pathway of *Euscaphis konishii* were used to verify the suitability of reference gene.

**Results.** The results showed that the stability of EkUBC23, EkCYP38 and EkGAPDH2 was better, and the low expression reference genes (EkUBC23 and EkCYP38) were favourable for quantifying low expression target genes, while the high expression reference gene (EkGAPDH2) was beneficial for quantifying high expression genes. In this study, we present the suitable reference genes for fruit development of *Euscaphis konishii* based on transcriptome data, our study will contribute to further studies in molecular biology and gene function on *Euscaphis konishii* and other closely related species.

**Subjects** Agricultural Science, Molecular Biology

**Keywords** *Euscaphis konishii*, Fruit development, Reference genes, qRT-PCR

## INTRODUCTION

As an important economic crop, fruit trees play an important role in the agricultural production of various countries. Fruit development research is an important part of fruit science. Screening out stable reference genes in the fruit developmental stage can lay the foundation for the study of functional genes in neighbouring species and accelerate the process of molecular breeding.

Submitted 31 May 2019  
Accepted 27 December 2019  
Published 11 February 2020

Corresponding author  
Shuang-Quan Zou, zou@fafu.edu.cn

Academic editor  
Rodolfo Aramayo

Additional Information and  
Declarations can be found on  
page 12

DOI 10.7717/peerj.8474

© Copyright  
2020 Yang et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

The analysis of gene expression, or more correctly transcript abundance, is essential in all aspects of molecular biology research to understand gene expression patterns in different biological processes (Dussert *et al.*, 2013). So far, the commonly used methods for the detection of gene expression are northern blotting, gene chip and quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR). Among them, qRT-PCR has become one of the most powerful tools for studying gene expression due to its high sensitivity, accuracy and specificity (Bustin, 2002; Nolan, Hands & Bustin, 2006). qRT-PCR is wonderful for fast and accurate gene expression analysis. However, this technology requires suitable reference genes to normalize expression data and control the quantity of cDNA (Derveaux, Vandesompele & Hellemans, 2010). In previous stages, housekeeping genes were selected as stable reference genes. However, lots of housekeeping genes have significantly different levels of expression in different experimental conditions or related species (Dheda *et al.*, 2004). The stability of reference gene is influenced by experimental conditions and plant species. It is necessary to find additional reference genes for most plants. Therefore, according to different experimental conditions, the key is to accurately quantify target genes and screen one or more reference genes with stable expression from multiple candidate reference genes.

Currently, Gene Expression Omnibus was used as a new tool to screen reference genes due to its accuracy and comprehensiveness. GeneChip research of *Arabidopsis thaliana* revealed that the protein phosphatase 2A was favourable for quantifying low expression target genes (Czechowski *et al.*, 2005); Coker and Davies' research showed that the stability of TUA (*Tubulin alpha*), CYP (*Cyclophilin*) and GAPDH (*Glyceraldehyde-3-phosphate dehydrogenase*) is better by using dbEST (Coker & Davies, 2004). The best reference genes of *Elaeis guineensis* (Xia *et al.*, 2014), *Populus trichocarpa* (Pang *et al.*, 2017) and *Populus trichocarpa* (Su *et al.*, 2013) were screened out based on RNA-seq technology.

*Euscaphis konishii* is a member of the family Staphyleaceae, which is used as an ornamental and medicinal source material in China. The fruit of *E. konishii* has various chemical compounds, including flavonoid compounds (Liang *et al.*, 2018a), triterpene compounds (Cheng *et al.*, 2010; Li *et al.*, 2016), and phenolic acid compounds (Huang *et al.*, 2014), which have both anti-inflammatory and anticancer effects according to traditional and modern medical research. Moreover, in order to study the gene regulation mechanisms of medicinal compounds in *E. konishii*, several reference genes in *E. konishii* were selected by Liang *et al.* (2019) based on transcriptome data (Liang *et al.*, 2018b). However, the data being studied from Liang *et al.* lack the transcriptome of the fruit at different developmental stages. Yuan *et al.* (2018) identified a large number of differentially expressed genes associated with endocarp colouration based on the transcriptome database, but it is unknown which genes are suitable as reference genes related to fruit development. Huang *et al.* (2019) revealed that the total triterpenes of *Euscaphis konishii* pericarp showed the immune boosting effect and hepatic protective activity against inflammation, oxidative stress and apoptosis in BCG/LPS-induced liver injury. These findings suggest that pericarp of *E. konishii* might be a promising natural food for immunological hepatic injury. Because of the information on gene expression in *E. konishii* is still unclear, transcriptome technology would play an important role in mining the reference gene. Our study provided

two available internal control genes for qRT-PCR data normalization in *E. konishii*. It will contribute to analyze the expression patterns of compound related gene, and further to reveal the breeding mechanism.

## MATERIALS & METHODS

### Plant material

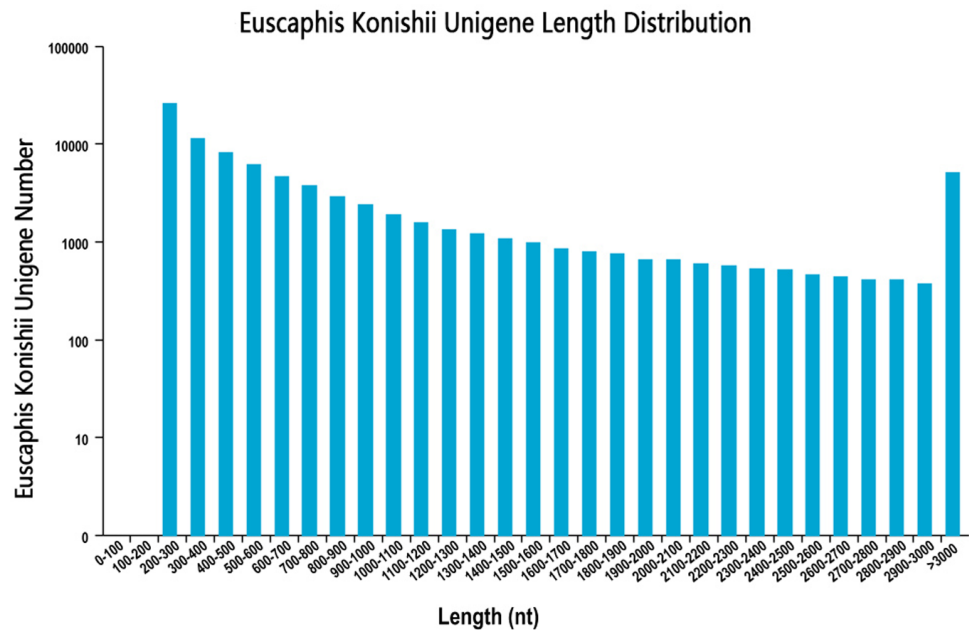
The fruits of *E. konishii* have three developmental stages, including the green stage, turning stage, and red stage. These were collected from Qingliu County, Fujian Province, China, from June to September. The samples were provided by Sanming yisheng agricultural and forestry co., LTD. All pericarp samples were separated from the fruit, wrapped in tin foil and then frozen in liquid nitrogen and immediately stored at  $-80^{\circ}\text{C}$  until they were qualified for further analysis. Three biological replicates for each sample were used for RNA extraction.

### Establishment of the RNA-Seq database

We sequenced the transcriptome of the *E. konishi* i pericarp in three developmental stages. The library produced 67.78 G of clean data, and the average clean data in the nine samples were 6.78 G. In this study, high-quality libraries with mapping rate were higher than 79.74% and Q30 values higher than 92.70%. In total, 86,120 unigenes with a mean length of 893.34 bp and N50 length of 1,642 nt were assembled by using Trinity (*Grabherr et al., 2011*) with `min_kmer_cov` set to 2 by default and all other parameters set default. The length distribution of all unigenes were shown in [Fig. 1](#). The raw data has been submitted to NCBI, Sequence Read Archive (SRA) submission: [SRR9267648](#) to [SRR9267656](#). We used BLAST software to compare unigene sequences to NR (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), Pfam (<http://pfam.xfam.org/>), GO (<http://www.geneontology.org/>), KEGG (<http://genome.jp/kegg/>), Pwiss-Prot (<http://uniprot.org/>), KOG (<http://www.ncbi.nlm.nih.gov/KOG>), eggNog (<http://eggongdb.embl.de/>), and COG (<http://www.ncbi.nlm.nih.gov/COG>). Transcriptome sequences were used as references to analyse the expression profiles of each sample. The sequencing results of each sample were compared to the reference sequence, and the expression amount of each unigene in different samples was obtained.

### RNA-Seq database analysis and primer design

Genes with expression levels lower than 5 FPKM in transcriptomes would make poor qPCR references due to the difficulties in detecting and quantifying their expression. (*Kimmy, Patrick & Joshua, 2017*). After their removal, a total of 75,048 genes in *E. konishii* were evaluated. We calculated the mean expression value, standard deviation, and coefficients of variation based on the raw RNA-seq data. Base on the requirements  $\text{CV} \leq 0.2$  and  $\text{FC} \leq 0.2$ , 1,131 genes ([Table S1](#)) were obtained by removing overabundant genes with low expression levels. Finally, we selected the eight candidate genes according to genome annotation of each, which was assigned base on the best mach of the alignments using Blast to NR, TrEMBL and KEGG databases. The selecting steps of candidate reference genes were shown in [Fig. S1](#). In order to ensure the accuracy of the reference gene predictions, the



**Figure 1** Unigenes length distribution.

Full-size DOI: [10.7717/peerj.8474/fig-1](https://doi.org/10.7717/peerj.8474/fig-1)

**Table 1** Analysis of transcriptome and expression of candidate reference genes.

#ID	<i>Arabidopsis</i> ortholog	Gene	<i>E</i> -value	FPKM Mean	SD	CV	FC		
							FC-1	FC-2	FC-3
c71660.graph_c1	AT1G13440.1	<i>GAPDH2</i>	4e−34	1,418.24	145.33	0.10	0.13	0.08	0.05
c67439.graph_c2	AT5G19770.1	<i>TUA3</i>	3e−68	165.32	18.28	0.11	0.12	0.05	0.08
c67539.graph_c0	AT3G01480.1	<i>CYP38</i>	1e−59	19.10	3.07	0.16	0.08	0.08	0.01
c67010.graph_c0	AT2G16920.1	<i>UBC23</i>	2e−158	18.30	2.05	0.11	0.19	0.13	0.04
c62586.graph_c0	AT3G52590.1	<i>UBQ1</i>	3e−10	230.36	21.98	0.10	0.19	0.00	0.18
c65728.graph_c0	AT3G15020.2	<i>mMDH2</i>	5e−35	187.30	18.88	0.10	0.11	0.20	0.08
c63030.graph_c0	AT3G47520.1	<i>MDH</i>	2e−19	19.87	1.95	0.10	0.16	0.14	0.03
c63658.graph_c0	AT5G09810.1	<i>ACT7</i>	2e−38	407.67	35.20	0.09	0.09	0.17	0.08

**Notes.**

FPKM, Fragments Per Kilobase of transcript per Million fragments mapped; SD, CV, coefficient variation; FC, fold change; *GAPDH2*, Glyceraldehyde-3-phosphate dehydrogenase c-2; *TUA3*, Tubulin alpha-3; *CYP38*, cyclophilin 38; *UBC23*, Ubiquitin-conjugating enzyme 23; *UBQ1*, Ubiquitin extension protein 1; *mMDH2*, Lactate/malate dehydrogenase family protein; *MDH*, malate dehydrogenase; *ACT7*, actin 7.

coding sequences of eight selected genes was used as queries for BLAST orderly through the TAIR database (<http://www.arabidopsis.org/>). The sequences with the highest homology with *Arabidopsis* were shown in Table 1.

Here we obtained the nucleotide sequences of reference genes from our laboratory. The nucleotide sequences of candidate reference genes were appended in Table S2. According to the nucleotide sequence of candidate reference genes and the design principle of qRT-PCR, forward and reverse primers of all candidate reference genes were designed

**Table 2** Genes and primer sets used for qRT-PCR analysis.

#ID	Gene name	Primer sequence	Length	Distance from 3'	Anneal temp
c71660.graph_c1	<i>GAPDH2</i>	F:CCGTGTTCTACTGTTGATGT R:CCTCCTTGATAGCAGCCTTAAT	95	1,244 1,338	62.0 61.9
c67439.graph_c2	<i>TUA3</i>	F:GGGTGGTAGCAAACCCTATTAC R:CCGAAGGTGCAGATGATGAA	103	203 305	62.4 62.2
c67539.graph_c0	<i>CYP38</i>	F:ATCTGTTGGAACCTCCATTC R:AGCCCTGAAGCAAGGTAAAG	114	2,251 2,364	62.0 62.2
c67010.graph_c0	<i>UBC23</i>	F:AGCCACATAATCTCCGTGTAAG R:GCTGACCATGTTTCGAGTAGTT	105	4,161 4,265	62.0 62.0
c62586.graph_c0	<i>UBQ1</i>	F:ACGAGCCAAAGCCATCAA R:GGCCGAACTCTTGCTGATTA	105	1,435 1,539	62.0 62.2
c65728.graph_c0	<i>mMDH2</i>	F:CATCGTAAGTCCCTGCTTTCT R:TGCCAAGTACTGCCCTAATG	104	956 1,059	61.9 62.0
c63030.graph_c0	<i>MDH</i>	F:ATGAAGAAGTCCACGAGCTAAC R:GCCATAGACAGAGTAGCAGAAC	97	995 1,091	62.1 62.0
c63658.graph_c0	<i>ACT7</i>	F:GATCTGGCATCACACCTTCTAC R:CTGAGTCATCTTCTCCCTGTTG	112	394 505	62.3 62.0

**Notes.**

F, forward primer; R, reverse primer.

using Primer Premier 6.0 software and were synthesized by Jinweizhi Biotechnology Co., LTD (Suzhou, China) and purified by PAGE. The primer sequences, amplicon size, anneal temperature were shown in [Table 2](#).

### RNA Extraction and cDNA Synthesis

Total RNA was extracted and purified by using an RNAprep Pure Plant Kit (Polysaccharides and Polyphenolics-Rich, Tiangen, Beijing, China), according to the manufacturer's instructions. Two-hundred nanograms of total RNA of each sample was used as the template. In addition, the cDNA synthesis strand was performed by using a Revert Aid First Strand cDNA Synthesis Kit (Thermo Fisher, Foster City, CA, USA), according to the manufacturer's instructions and was stored at  $-80^{\circ}\text{C}$  for further experiment.

### Candidate reference genes for RT-PCR amplification

The RT-PCR mixture contained 25  $\mu\text{L}$  L of  $2\times$  EasyTaq<sup>®</sup> PCR SuperMix, 0.4  $\mu\text{L}$  L of forward primer (10  $\mu\text{M}$ ), 0.4  $\mu\text{L}$  L of reverse primer (10  $\mu\text{M}$ ), and 5  $\mu\text{L}$  L of diluted cDNA in a final volume of 50  $\mu\text{L}$  L. PCR conditions: 40 cycles of 3 min at  $94^{\circ}\text{C}$ ,  $94^{\circ}\text{C}$  at 30 s,  $56^{\circ}\text{C}$  at 10 s,  $72^{\circ}\text{C}$  at 30 s, and  $72^{\circ}\text{C}$  at 7 min. RNA quality was determined by 1.2% agarose gel electrophoresis.

### Candidate reference genes for qRT-PCR amplification

The qRT-PCR mixture contained 3  $\mu\text{L}$  L of diluted cDNA, 5  $\mu\text{L}$  L of  $2\times$  SYBR Green Master, 0.4  $\mu\text{L}$  L of forward primer (10  $\mu\text{M}$ ), and 0.4  $\mu\text{L}$  L of reverse primer (10  $\mu\text{M}$ ), with ddH<sub>2</sub>O added to achieve a final volume of 10  $\mu\text{L}$  L. All PCRs were performed using the QuantStudio<sup>™</sup> Real-Time PCR System under the following conditions: 40 cycles of 2

**Table 3** Target genes and primers.

Locus	Primer sequence (5'–3')
c50541.graph_c0 (CHS)	F:GGAACCTCGCTGTTCTGGATAG/R:CCTTGTGGCCCTTAACTTCT
c59825.graph_c0 (CHS)	F:GCATGTGTTGTGCGAGTATG/R:CCTTCCCTTCTCCAGAGATTT
c60763.graph_c0 (ANS)	F:CAGCTTGAGTGGGAAGACTATT/R:TACTCGCTTGTGTCCTCTATG
c64532.graph_c1 (F3H)	F:GGTTCAAGATTGGCGTGAAATAG/R:CATCAGCTTCCCCTGTACTC
c69442.graph_c0 (CHI)	F:TCTTGCTGAGGATGATGACTTT/R:TCTCTAGCTGCACTCCATACT
c69862.graph_c1 (UFGT)	F:ACCGCTAATCCCAACTCTTTC/R:GTGGTTCGGTGTGCCTATT
c71357.graph_c2 (FLS)	F:CGACAATCGCTCCATCTTCT/R:ATGGCCTCCTTCTGTATTAAC
c72659.graph_c0 (CHS)	F:TTGGTGACGCCGAAGATAAAA/R:GAGGTCCAGCTACAGTTCTTG
c72737.graph_c0 (CHI)	F:CTCTTGCCAGCAGCATTCTA/R:CAGAGTTTGGCTGCAAGAATATC
c58939.graph_c0(SS3)	F:TGAATGGATGCAGGTGACTGGAAC/R:CCACACTTGCT-GAGTTGCTCTTTC

**Notes.**

F, forward primer; R, reverse primer.

min at 95 °C, 5 s at 95 °C and 30 s at 60 °C. The procedure ended with a melt-curve ramping from 60–95 °C. The melting curve was analysed to determine primer specificity. For further confirmation, PCR products were cloned into pGXT vectors respectively and then sequenced by Biosune Biotech (Fuzhou, China).

### Statistical analysis

The instrument calculated the Cq value of each sample based on the qRT-PCR experiment. We can analyse the stability of candidate reference genes using GeNorm, NormFinder, and BestKeeper software. Finally, ReFinder was used to calculate the comprehensive ranking of the stability of candidate reference genes based on analysis results.

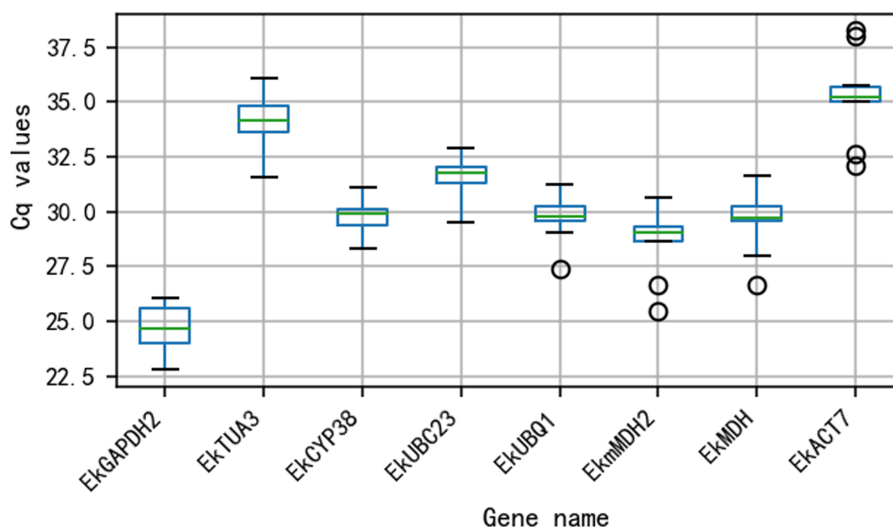
### Validation of the candidate reference genes

Anthocyanin is the key factor affecting fruit colouration as an important plant pigment. In general, fruit colouration is closely related to the content and proportion of anthocyanin. To verify the results of our experiments, the expressions of nine genes related to the anthocyanin synthesis pathway were calculated with the reference genes selected (Fig. S5). Moreover, the EkSS3 gene related to glycometabolism pathway was also used to validate the reliability of reference genes. The details of primer were shown in Table 3. The qRT-PCR experimental method was the same as described above, and the relative expression level was calculated by the  $2^{-\Delta\Delta ct}$  method. Experimental data from three biological replicates were analysed using analysis of variance (ANOVA), followed by Student's *t*-test ( $P < 0.05$ ).

## RESULTS

### Differentially expressed genes between fruit with varied development

Based on the three comparisons of green vs. turning, green vs. red, and turning vs. red, we identified a total of 4,804 differentially expressed genes (DEGs). Among them, 2,175 DEGs between green vs. turning, 3,935 DEGs between green vs. red, and 936 DEGs between turning vs. red were detected. Among the 4804 DEGs selected to predict functions



**Figure 2** Quantification cycle(Cq) value of eight candidate reference genes across all the experimental samples. The green line across the box depicts the median. Whiskers represent the local maximum and minimum values, and points beyond the end of each whisker mark the outliers. The box indicates the 25% and 75% percentiles.

Full-size DOI: 10.7717/peerj.8474/fig-2

by gene ontology (GO) annotation and KEGG pathway analysis. The genes related to the glycometabolism, anthocyanin biosynthesis, flavonoid biosynthesis and carotenoid biosynthesis performed different expression in different stages of fruit ripening.

### Reference gene selection and primer design

According to selecting process (Fig. S1), eight candidate reference genes were chosen to perform the gene normalization studies. The sequences of these candidate reference gene were used to design the qRT-PCR primers. The details of eight reference genes were listed in Table 1. The gene models of eight candidate reference referred to the homologous genes of *Arabidopsis*, and the details were shown in Fig. S2.

### Candidate reference genes for RT-PCR amplification

PCR results of the 8 candidate internal reference genes are shown in Fig. S4 with a single band, no primer dimer and nonspecific amplification, which could be used for subsequent qRT-PCR analysis. We sequenced the PCR amplification products of eight genes and blast the correctness of the PCR amplified fragments (Table S3).

### Cq values of candidate reference genes

The Cq values for all 8 reference genes are shown in Fig. 2. The Cq values varied from 24.65 (*EkGAPDH2*) to 35.27 (*EkACT7*). Moreover, as shown in Fig. 2, *EkmMDH2*, *EkMDH* and *EkACT7* are more variable than other candidate reference genes. The stability of *EkmMDH2*, *EkMDH* and *EkACT7* expression is poor.



### qRT-PCR analysis

As shown in Fig. S3, the melting curve of the 8 candidate reference genes at different developmental stages only had a single main peak, and the amplification curves between the repeated samples had high repeatability.

### Expression stability analysis of reference genes

Evaluating the expression stability of candidate reference genes depended on statistical analysis. Four different statistical software programs (GeNorm, NormFinder, BestKeeper and ReFinder) were commonly used to calculate the variability of the candidate genes expression and determine the one which were the most suitable.

Gene expression stability was determined by the M-value in GeNorm. The gene expression stability increases as the M-value decreases (Wu *et al.*, 2017; Vandesompele *et al.*, 2002). It was considered that candidate reference genes can be used as reference genes when the M-value is less than 1.0; the result of GeNorm revealed that *EkUBQ1* was the most stable gene with the lowest M-value, followed by *EkUBC23*, *EkCYP38* and *EkGAPDH2*. However, *EkACT7*, *EkMDH2*, *EkMDH* and *EkTUA3* were unsuitable as reference genes, with a value greater than 1.0. Expression stability values analysed by GeNorm are shown in Fig. 3A.

NormFinder software ranks all reference gene candidates based on intra- and inter-group variation and combines results into a stability value for each candidate reference gene. A better stable gene expression is indicated by a lower stability value (Andersen, Jensen & Ørntoft, 2004). The analysis results of NormFinder shown in Fig. 3B revealed that *EkGAPDH2* was the most stable gene for all samples.

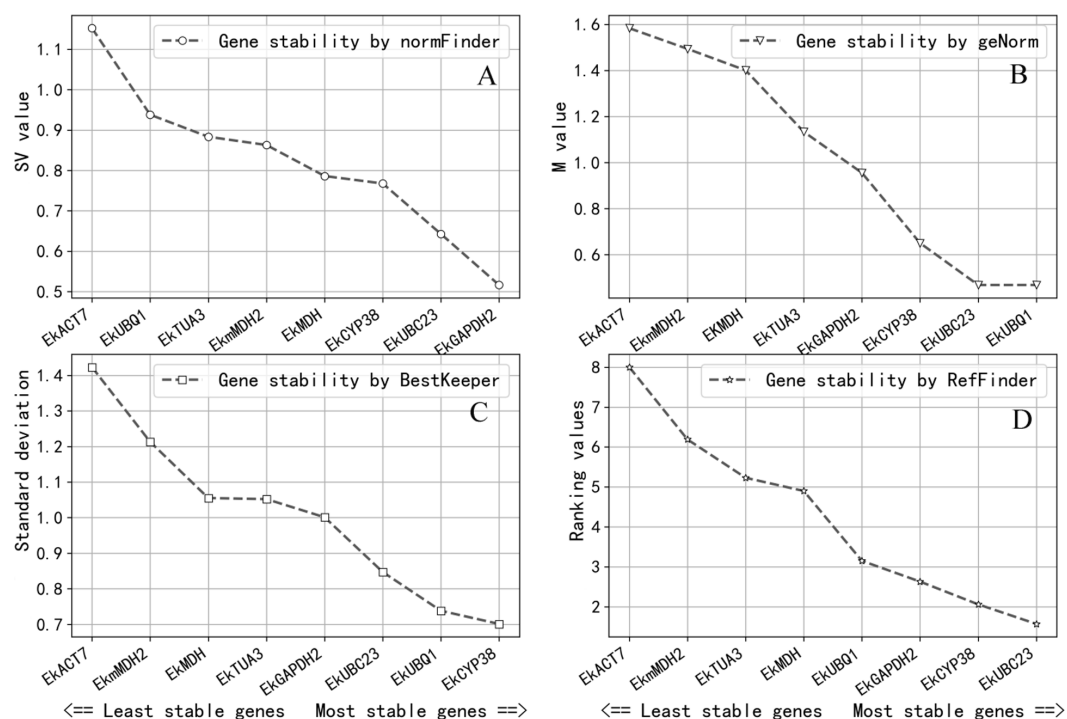
BestKeeper software determines the most stably expressed genes based on the standard deviation (SD). The lower the SD, the greater the reference gene expression stability will be (Pfaffl *et al.*, 2004). As the analysis results in Fig. 3C show, the SD values of *EkACT7*, *EkMDH2*, *EkMDH*, *EkTUA3* and *EkGAPDH2* are greater than 1 and are considered unacceptable as reference genes, according to the selection criteria of BestKeeper software. *EkUBC23*, *EkUBQ1* and *EkCYP38* are stable genes with low SD values.

ReFinder software was used to rank the stability obtained by the analysis of GeNorm, NormFinder and BestKeeper on the comprehensive index. The stability of the reference gene expression is directly related to the size of the index. According to the analysis result of ReFinder shown in Fig. 3D, the stability ranking obtained by ReFinder software is as follows: *EkUBC23*, *EkCYP38*, *EkGAPDH2*, *EkUBQ1*, *EkMDH*, *EkTUA3*, *EkMDH2*, and *EkACT7*. The results showed that the stability of *EkUBC23*, *EkCYP38* and *EkGAPDH2* was better, and the low expression reference genes (*EkUBC23* and *EkCYP38*) were favourable for quantifying their targets, while the high expression reference gene (*EkGAPDH2*) was beneficial for quantifying its target.

### The suitability of the reference gene

According to the results of the four algorithms (GeNorm, NormFinder, BestKeeper and ReFinder), *EkUBC23*, *EkCYP38* and *EkGAPDH2* performed more stably. *EkUBC23* and *EkCYP38* showed similar expression levels during fruit developmental stages.



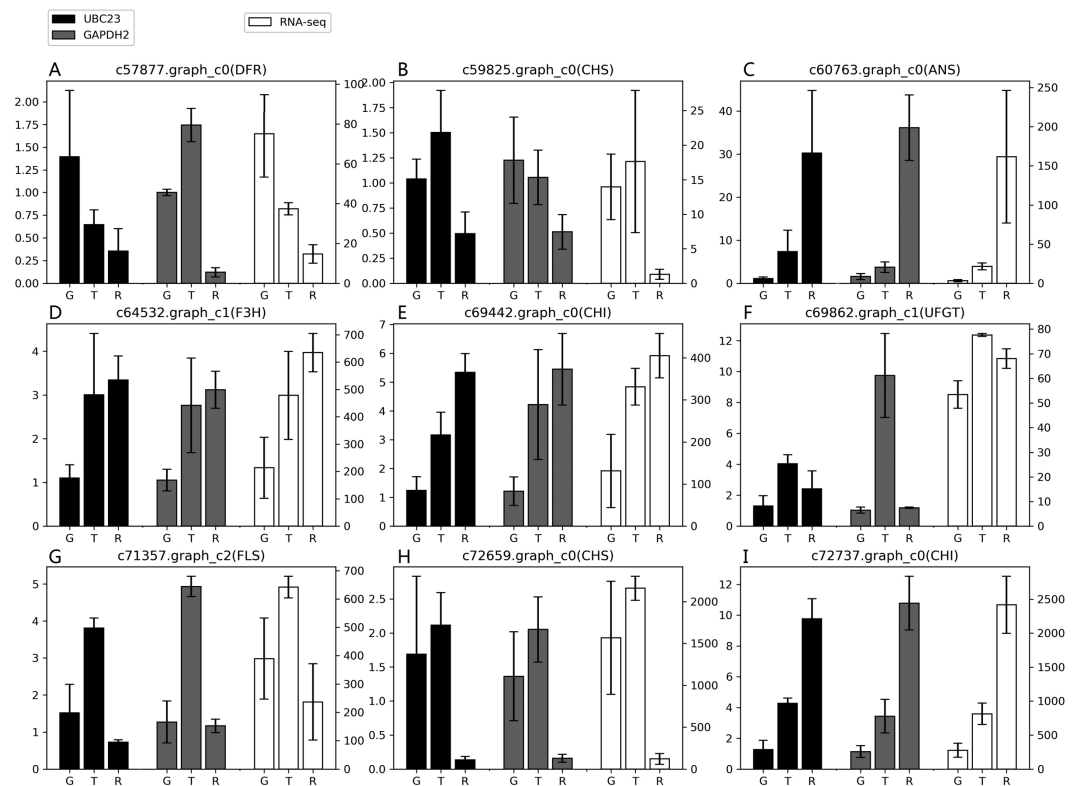


**Figure 3** Expression stability analysis of reference genes by NormFinder (A), geNorm (B), BestKeeper (C) and ReFinder (D). Analysis was performed after pooling Cq data across all three growth conditions. The designation of the analyses were performed on all eight candidate reference genes. Gene were ranked from the least stable (on the left) to the most stable (on the right). (A) Genes were ranked according to their NormFinder SV value. (B) Genes were ranked according to their geNorm M value. (C) Genes were ranked according to their BestKeeper SD value. (D) Genes were ranked according to their ReFinder Ranking value.

Full-size [DOI: 10.7717/peerj.8474/fig-3](https://doi.org/10.7717/peerj.8474/fig-3)

However, *EkUBC23* was used to normalize the expression due to its greater stability. Relative expression levels were normalized using the low expression reference gene (*EkUBC23*) and high expression reference gene (*EkGAPDH2*). The results showed that there were some differences between the expression levels of 9 genes related to the anthocyanin synthesis pathway and transcriptome sequencing (Fig. 4). When *EkUBC23* was used as the reference gene, the results had no significant difference. When *EkGAPDH2* was used for normalizing low expression target genes [*c57877.graph\_c0(DFR)*, *c59825.graph\_c0(CHS)* and *c69862.graph\_c1(UFGT)*], the results were different from the expression of transcriptome. However, *EkGAPDH2* has better accuracy than *EkUBC23* when it was used for normalizing high expression target genes [*c72659.graph\_c0(CHS)* and *c72737.graph\_c0(CHI)*]. Therefore, we suggest that the stability and expression of reference genes should be considered as important selection conditions.

To ensure the reliability of result, We chosen *EkSS3* gene related to glycometabolism pathway as target gene. Relative expression levels were normalized using two most stable reference genes (*EkGAPDH2* and *EkUBC23*). As shown in Fig. S6, when normalized using *EkGAPDH2* and *EkUBC23* as reference gene respectively, the results had no significant



**Figure 4** The suitability of reference gene. (A–I) Genes related to the anthocyanin synthesis pathway were chosen as target genes. Those genes were normalized to *UBC23* and *GAPDH2*, respectively. Error bars represent standard deviation from three independent biological replicates. “G”, “T” and “R” mean green stage, turning stage and red stage in fruit, respectively.

Full-size [DOI: 10.7717/peerj.8474/fig-4](https://doi.org/10.7717/peerj.8474/fig-4)

difference. Compared with the green stage, *EkSS3* expression was clearly downregulated at turning stage and red stage. The *EkGAPDH2* and *EkUBC23* performed well when normalizing target gene in different pathway.

## DISCUSSION

Many studies have shown that it was quicker and more efficient to screen appropriate reference genes based on transcriptome data. In addition, primers designed according to the transcriptome sequencing of the material itself are more stable than those using other materials. Selection of the reference gene based on transcriptome data has been done in many plant species, such as *Corylus* (Yang et al., 2017), *Sedum alfredii* (Sang et al., 2013) and *Dendrocalamus sinicus* (Guo, Chen & Yang, 2018), and so on. In the present study, we selected three stable reference genes (*EkUBC23*, *EkCYP38* and *EkGAPDH2*) based on transcriptome data of *E. konishii* pericarp in three different developmental stages. This result validates and complements the results of reference gene screening by Liang. (Liang et al., 2018b).

The expression of reference genes was related to organ type, developmental stages and external environmental conditions. *ACT* was commonly used as a reference gene. For

example, it was the most stable gene in studies of *Glycine max* (Ma et al., 2013). In contrast, we confirm that *ACT* is the most unstable gene in our study. This consequence is consistent with studies in *Setaria viridis* (Martins et al., 2016) and *E. konishii* by Liang et al. (2018b). *UBC* has been widely accepted for normalization of gene expression as a reference gene (Yuan, Wan & Yang, 2012). *UBC* was selected as the reference gene in the study of *Prunus pseudocerasus* (2 (Zhu et al., 2015). However, it was unsuitable as a reference gene for *Oryza sativa* (Li et al., 2008). In this study, we identified *UBC* as a reference gene to quantify low expression target genes. *CYP* played important roles in *Elaeis guineensis* (Yeap et al., 2014) and *Malus domestica* (Kumar & Singh, 2015) as reference a gene. *CYP* with stability and low expression is also used for quantifying low expression target genes. *GAPDH*, an enzyme in glycolysis, has been widely used as a reference gene in different species (Kozera & Rapacz, 2013). However, *GAPDH* has different performances in different species and different experimental conditions. *GAPDH* shows stable expression in *Citrus sinensis* (Wu et al., 2014) and *Lycopersicon esculentum* (Mascia et al., 2010), but it did not perform well in *S. viridis* (Martins et al., 2016) and *Panicum virgatum* (Jiang et al., 2014). In previous study, our colleague Liang revealed *EkGSTU1* performed better than *EkGAPDH2* in root, leaf, branch and seed samples (Liang et al., 2019). However, in our study, *EkGAPDH2* and *EkUBC23* performed well in different stages of fruit ripening. Our study expanded the scope of reference genes screening for this species. In this study, we selected *EkGAPDH2* as the reference gene to quantify the high expression target genes for further study. We also confirmed *EkGAPDH2* and *EkUBC23* had stable expression in other tissues by semi-quantitative RT-PCR experiment. The RT-PCR experimental method was the same as described above and the details of result were show in Fig. S7. The result showed that *EkGAPDH2* and *EkUBC23* had stable expression in root, branch and leaf, it will facilitate to reference genes screening for whole plant.

The genomic information of *E. konishii* still unclear. The fruit of *Euscaphis konishii* has high medicinal value, which was used as medicinal source material in China. In order to analyze the expression patterns of medicinal compound related gene in fruit of *Euscaphis konishii*, we selected two reliable reference genes for qRT-PCR normalization. The results will contribute to future studies of the gene expression in *E. konishii* and genetic studies related to fruit developmental stages.

## CONCLUSIONS

The development of high-throughput sequencing technology provides an accurate and efficient approach to molecular biology. It plays an important role in molecular breeding and specific gene research. In this study, we construct a screening system for reference genes of *E. konishii* based on an RNA-seq database from different developmental pericarp samples. We selected three stable reference genes (*EkUBC23*, *EkCYP38*, and *EkGAPDH2*) from eight candidate reference genes. Among them, *EkUBC23* and *EkCYP38* with low expression are suitable for quantifying low expression target genes. However, *EkGAPDH2* with a high expression is favourable for quantifying high expression genes. Our study will contribute to future studies of gene expression in *E. konishii* and genetic studies related to fruit. The results also provided reference for neighboring species.

## ACKNOWLEDGEMENTS

We are particularly grateful to Xiaoxing Zou, Zeming Chen, and other members in our research group for their kind suggestions to perfect the experiment.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The authors received no funding for this work.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Cheng-Long Yang and Xue-Yan Yuan conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Jie Zhang conceived and designed the experiments, prepared figures and/or tables, and approved the final draft.
- Wei-Hong Sun performed the experiments, prepared figures and/or tables, and approved the final draft.
- Zhong-Jian Liu and Shuang-Quan Zou conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

### Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

Field experiments were approved by Sanming yisheng agricultural and forestry co. LTD

### Data Availability

The following information was supplied regarding data availability:

The raw data is available at NCBI SRA: [SRR9267648](https://www.ncbi.nlm.nih.gov/sra/SRR9267648) to [SRR9267656](https://www.ncbi.nlm.nih.gov/sra/SRR9267656).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.8474#supplemental-information>.

## REFERENCES

- Andersen CL, Jensen JL, Ørntoft TF. 2004.** 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 Research* **64**(15):5245–5250 DOI [10.1158/0008-5472.CAN-04-0496](https://doi.org/10.1158/0008-5472.CAN-04-0496).
- Bustin SA. 2002.** Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology* **29**(1):23–39.

- Cheng JJ, Zhang LJ, Cheng HL, Chiou CT, Lee IJ, Kuo YH. 2010.** Cytotoxic hexacyclic triterpene acids from *Euscaphis japonica*. *Journal of Natural Products* **73(10)**:1655–1658 DOI [10.1021/np1003593](https://doi.org/10.1021/np1003593).
- Coker JS, Davies E. 2004.** Selection of candidate housekeeping controls in tomato plants using EST data. *Biotechniques* **35(4)**:740–749.
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. 2005.** Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology* **139(1)**:5–17 DOI [10.1104/pp.105.063743](https://doi.org/10.1104/pp.105.063743).
- Derveaux S, Vandesompele J, Hellemans J. 2010.** How to do successful gene expression analysis using real-time PCR. *Methods* **50(4)**:227–230 DOI [10.1016/j.ymeth.2009.11.001](https://doi.org/10.1016/j.ymeth.2009.11.001).
- Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumia A. 2004.** Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* **37**:112–114 DOI [10.2144/04371RR03](https://doi.org/10.2144/04371RR03).
- Dussert S, Guerin C, Andersson M, Joët T, Tanbarger TJ, Pizot M, Sarah G, Omoro A, Durand T, Morcillo F. 2013.** Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiology* **162**:1337–1358 DOI [10.1104/pp.113.220525](https://doi.org/10.1104/pp.113.220525).
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma FD, Birren BW, Nusbaum C, Toh KL, Friedman N, Regev A. 2011.** Full length transcriptome assembly from RNA seq data without a reference genome. *Nature Biotechnology* **29**:644–652 DOI [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883).
- Guo XJ, Chen LN, Yang HQ. 2018.** Reference gene selection for quantitative real-time PCR in studying culm shape development of *Dendrocalamus sinicus*. *Forest Research* **31(02)**:120–125 DOI [10.13275/j.cnki.lykxyj.2018.02.017](https://doi.org/10.13275/j.cnki.lykxyj.2018.02.017).
- Huang W, Ding H, Chen LY, Ni L, Ruan YF, Zou XX, Ye M, Zou SQ. 2019.** Protective effect of the total triterpenes of *Euscaphis konishii* Hayata pericarp on bacillus Calmette-Guérin plus lipopolysaccharide-induced liver injury. *Evidence-Based Complementary and Alternative Medicine* **2019**:1–15 DOI [10.1155/2019/1806021](https://doi.org/10.1155/2019/1806021).
- Huang Y, Xiang DB, Hu QM, Tan Y, Meng YC, Fei G. 2014.** Phenolic acids from fruits of *Euscaphis fukienensis*. *Chinese Traditional and Herbal Drugs* **45(18)**:2611–2613 DOI [10.7501/j.issn.0253-2670.2014.18.007](https://doi.org/10.7501/j.issn.0253-2670.2014.18.007).
- Jiang XM, Zhang XQ, Yan HD, Zhang Y, Yang ST, Huang LK. 2014.** Reference gene selection for real-time quantitative PCR normalization in switchgrass (*Panicum virgatum* L.) root tissue. *Journal of Agricultural Biotechnology* **22(1)**:55–63.
- Kimmy AS, Patrick PE, Joshua RP. 2017.** A whole-transcriptome approach to evaluating reference genes for quantitative gene expression studies: a case study in mimulus. *G3 (Bethesda)* **7(4)**:1085–1095 DOI [10.1534/g3.116.038075](https://doi.org/10.1534/g3.116.038075).
- Kozera B, Rapacz M. 2013.** Reference genes in Real-time PCR. *Journal of Applied Genetics* **2013** **54(4)**:391–406.
- Kumar G, Singh AK. 2015.** Reference gene validation for qRT-PCR based gene expression studies in different developmental stages and under biotic stress in apple. *Scientia Horticulturae* **197**:597–606 DOI [10.1016/j.scienta.2015.10.025](https://doi.org/10.1016/j.scienta.2015.10.025).

- Li QF, Jiang MY, Yu HX, Xin SW, Gu MH, Liu QQ. 2008.** Selection of internal reference genes for quantitative RT-PCR analysis of total RNA from endosperm of rice (*Oryza sativa* L.). *Jouranl of YangZhou University (Agricultural and Life Science Edition)* **29(2)**:61–66.
- Li YC, Tian K, Sun LJ, Long H, Li LJ, Wu ZZ. 2016.** A new hexacyclic triterpene acid from the roots of *Euscaphis japonica*, and its inhibitory activity on triglyceride accumulation. *Fitoterapia* **109**:261–265 DOI [10.1016/j.fitote.2016.01.016](https://doi.org/10.1016/j.fitote.2016.01.016).
- Liang WX, Ni L, Rebeca CL, Zou XX, Sun WH, Wu LJ, Yuan XY, Mao YL, Huang W, Zou SQ. 2019.** Comparative transcriptome among *Euscaphis konishii* Hayata tissues and analysis of genes involved in flavonoid biosynthesis and accumulation. *BMC Genomics* **20**:24 DOI [10.1186/s12864-018-5354-x](https://doi.org/10.1186/s12864-018-5354-x).
- Liang WX, Ni L, Zou XX, Huang W, Zou SQ. 2018a.** Research progress on chemical constituents of *Euscaphis* and their pharmacological effect. *Chinese Traditional and Herbal Drugs* **5**:1220–1226.
- Liang WX, Zou XX, Rebeca CL, Wu LJ, Sun HW, Yuan XY, Wu SQ, Li PF, Ding H, Ni L, Huang W, Zou SQ. 2018b.** Selection and evaluation of reference genes for qRT-PCR analysis in *Euscaphis konishii* hayata based on transcriptome data. *Plant Methods* **14(1)**:Article 42 DOI [10.1186/s13007-018-0311-x](https://doi.org/10.1186/s13007-018-0311-x).
- Ma SH, Niu HW, Liu CJ, Zhang J, Hou CY, Wang DM. 2013.** Expression stabilities of candidate reference genes for rt-qpcr under different stress conditions in soybean. *PLOS ONE* **8(10)**:e75271 DOI [10.1371/journal.pone.0075271](https://doi.org/10.1371/journal.pone.0075271).
- Martins PK, Mafra V, Souza WRD, Riberio AP, Vineky F, Basso MF, Cunhua BADB, Molinari BC. 2016.** Selection of reliable reference genes for RT-qPCR analysis during developmental stages and abiotic stress in *setaria viridis*. *Scientific Reports* **20(6)**:28348.
- Mascia T, Santovito E, Gallitelli D, Fabrizio C. 2010.** Evaluation of reference genes for quantitative reverse transcription polymerase chain reaction normalization in infected tomato plants. *Molecular Plant Pathology* **11(6)**:805–816.
- Nolan T, Hands RE, Bustin SA. 2006.** Quantification of mRNA using real-time RT-PCR. *Nature Protocols* **1(3)**:1559–1582 DOI [10.1038/nprot.2006.236](https://doi.org/10.1038/nprot.2006.236).
- Pang QQ, Li ZL, Luo SB, Chen RY, Jin KM, Li ZX, Li DM, Sun BJ, Sun GW. 2017.** Selection and evaluation of reference genes for qRT-PCR analysis in Eggplant under high temperatures. *Horticultural Plant Journal* **44(3)**:475–486.
- Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. 2004.** Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: bestKeeper—excel-based tool using pair-wise correlations. *Biotechnology Letters* **26(6)**:509–515 DOI [10.1023/B:BILE.0000019559.84305.47](https://doi.org/10.1023/B:BILE.0000019559.84305.47).
- Sang J, Han XJ, Liu MY, Qiao GR, Jiang J, Zhuo RY. 2013.** Selection and validation of reference genes for real-time quantitative PCR in hyper accumulating ecotype of *Sedum alfredii* under different heavy metals stresses. *PLOS ONE* **8(12)**:e82927 DOI [10.1371/journal.pone.0082927](https://doi.org/10.1371/journal.pone.0082927).



- Su XJ, Fan BG, Yuan LC, Cui XN, Lu SF. 2013.** Selection and validation of reference genes for quantitative RT-PCR analysis of gene expression in *Populus trichocarpa*. *Bulletin of Botany* **48**(5):507–518.
- Vandesompele J, Preter KD, Pattyn F, Poppe E, Roy NV, Paepe AD, Speleman F. 2002.** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**(7):Article reports0034 DOI [10.1186/gb-2002-3-7-reports0034](https://doi.org/10.1186/gb-2002-3-7-reports0034).
- Wu JY, He B, Du YJ, Li WC, Wei YZ. 2017.** Analysis method of systematically evaluating stability of reference genes using geNorm, NormFinder and BestKeeper. *XianDai NongYe KeJi* **5**:278–281.
- Wu JX, Su SY, Fu LL, Zhang YJ, Chai LJ, Yi HL. 2014.** Selection of reliable reference genes for gene expression studies using quantitative real-time PCR in navel orange fruit development and *pummelo* floral organs. *Scientia Horticulturae* **176**(2):180–188 DOI [10.1016/j.scienta.2014.06.040](https://doi.org/10.1016/j.scienta.2014.06.040).
- Xia W, Mason , Xiao Y, Liu Z, Yang YD. 2014.** Analysis of multiple transcriptomes of the African oil palm (*Elaeis guineensis*) to identify reference genes for RT- qPCR. *Journal of Biotechnology* **184**:63–73 DOI [10.1016/j.jbiotec.2014.05.008](https://doi.org/10.1016/j.jbiotec.2014.05.008).
- Yang D, Li Q, Wang GX, Ma QH, Zhu LQ. 2017.** Reference genes selection and system establishment for real-time qPCR analysis in *Ping'ou Hybrid Hazelnut*. *Scientia Agricultura Sinica* **50**(12):2399–2410.
- Yeap WC, Loo JM, Wong YC, Kulaveerasingam H. 2014.** Evaluation of suitable reference genes for qRT-PCR gene expression normalization in reproductive, vegetative tissues and during fruit development in *oil palm*. *Plant Cell Tissue & Organ Culture* **116**(1):55–66 DOI [10.1007/s11240-013-0382-3](https://doi.org/10.1007/s11240-013-0382-3).
- Yuan W, Wan HJ, Yang YJ. 2012.** Characterization and selection of reference genes for real-time quantitative RT-PCR of plants. *Bulletin of Botany* **47**(4):427–436.
- Yuan XY, Sun WH, Zou XX, Liu BB, Huang W, Chen ZM, Li YL, Qiu MY, Liu ZJ, Mao YL, Zou SQ. 2018.** Sequencing of *Euscaphis konishii* endocarp transcriptome points to molecular mechanisms of endocarp coloration. *International Journal of Molecular Sciences* **10**(19):Article 3209.
- Zhu YY, Wang Y, Zhang H, Li YQ, Guo WD. 2015.** Selection and characterization of reliable reference genes in Chinese Cherry (*Prunus pseudocerasus*) using quantitative real-time PCR (qRT-PCR). *Journal of Agricultural Biotechnology* **23**(5):690–700.