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

Yang Chenglong [1], Yuan Xueyan [2,3], Zhang Jie [1], Sun Weihong [2,3], Liu Zhong-jian [2,3], Zou Shuang-quan [2,3*]

[1] Biotechnology Institute, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian, China
[2] Fujian Colleges and Universities Engineering Research Institute of Conservation and Utilization of Natural Bioresources at College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China
[3] 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

Corresponding Author:
Zou Shuang-quan [2,3,*]
No. 15, Shang xia-dian Road, Hong-shan Bridge, Cang-shan District, Fuzhou, Fujian, 350002, China
Email address: zou@fafu.edu.cn

Abstract

Background: With the development of molecular biology technology, research on the molecular mechanism of medicinal plants tends to be popular. Since *Euscaphis konishii* is a medicinal plant with a long history in China, a large number of specific genes in *E. konishii* have been obtained by high-throughput sequencing technology. To quantify target genes accurately, stable and reliable reference genes are the key to improving the accuracy of real-time PCR.

Methods: Based on the RNA-Seq data of different developmental stages, we selected eight candidate reference genes, including glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), α-tubulin (*TUA*), cyclophilin (*CYP*), ubiquitin-conjugating enzyme (*UBC*), ubiquitin *(UBQ)*, malate dehydrogenase (MDH/mMDH) and actin (ACT). GeNorm, NormFinder, BestKeeper and ReFinder were used to analyse the stability of reference genes.

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. This study provides abundant genetic

Commented [RA1]: Unclear meaning, bad English

Commented [RA2]: Unclear final objective

Commented [RA3]: Undescribed RNA data. No references

Commented [RA4]: Not presented or used in this study

Commented [RA5]: No mention of genome paralogy of these genes is ever mentioned, yet this is key to this manuscript

Commented [RA6]: Not clear what these packages/software are or what they are doing

Commented [RA7]: These results are being disputed by reviewers of this manuscript

information for the functional analysis of important characters of *E. konishii* and contributes to the understanding of the biological regulation mechanism of the fruit development stage.

Commented [RA8]: This statement is false as you have not provided data that support these claims

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.

With the rapid development and cost reduction of high-throughput sequencing technology, a multi-group sequencing analysis has been carried out in the field of plants, which enables researchers to discover a large number of specific genes in a relatively short time, laying the foundation for the exploration of plant genetic information and the study of gene function. As we all know, the growth and metabolic regulation of plants are closely related to changes in gene expression. Therefore, the accurate quantification of gene expression is crucial. 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 SA, 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 S, Vandesompele J & Hellemans J, 2010). In previous stages, candidate reference genes were selected as suitable reference genes due to the function of the housekeeping gene. However, lots of housekeeping genes have significantly different levels of expression in different experimental conditions or related species (Dheda K *et al*., 2004.). 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, the gene expression database is 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 T *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 when using EST data (Coker J S & Davies E, 2004). The best reference genes of *Elaeis guineensis* (Xia W *et al*., 2014), *Populus trichocarpa* (Pang QQ *et al*., 2017) and *Populus trichocarpa* (Su XJ *et al*., 2013) were screened out based on the transcript database.

E. 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 WX *et al*., 2018a), triterpene compounds (Cheng JJ *et al.*; Li YC *et al.,*2016), and phenolic acid compounds (Huang Y *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 fruit*, several reference genes in *E. konishii* were selected by Liang. (Liang WX *et al.,* 2019) based on transcriptome data (Liang WX *et al*., 2018b). However, the data being studied

Commented [RA9]: These are all true statements, yet the final objective is not clear. Why is this important? How are you going to use this knowledge? There is a disconnect between what you want to do and why you want to do it

Commented [RA10]: Correct spacing, here and throughout the manuscript

Commented [RA11]: Again, true statements disconnected of their importance. Why is this important?

Commented [RA12]: What Gene Expression Database? Please provide a reference

Commented [RA13]: Unclear statement. Re-write, EST data cannot affect the stability of the transcript

Commented [RA14]: Which transcript database?

Commented [RA15]: What does the "E" stands for? Full name on first citation

from Liang *et al*. lack the transcriptome of the *E. konishii* pericarp. Yuan identified a large number of differentially expressed genes associated with endocarp colouration based on the transcriptome database (Yuan WY *et al*., 2018), but it is unknown which genes are suitable as reference genes related to fruit development.

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. All pericarp samples were separated from the fruit, wrapped in tin foil and then frozen in liquid nitrogen and immediately stored at -80°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 9 samples were 6.78 G. In total, 86,120 unigenes were assembled. We used BLAST software to compare unigene sequences to NR (ftp://ftp.ncbi.gov/blast/db/), Pfam (http://pfam.xfam.org/), GO (http://www.geneontology.org/), KEGG(http://genome.jp/kegg/), Pwiss-Prot (http://uniprot.org/), KOG (ftp://ftpncbi.nih.gov/pub/KOG/), eggNog (http://eggongdb.embl.de/), and COG (ftp://ftp.ncbi.nih.gov/pub/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.

The dataset of the 86,120 unigenes serves as a platform from which to study the regulatory mechanisms of *E. konishii* fruit. Through the analysis of the differential gene expression in the pericarp expression profile at different developmental stages, we found that the coefficient of variation of some genes was less than 0.2; the possible answer is candidate reference genes. We selected eight candidate reference genes, including glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), α-tubulin (*TUA*), cyclophilin (*CYP)*, ubiquitin-conjugating enzyme (*UBC*), ubiquitin (*UBQ*), malate dehydrogenase (*MDH/mMDH*), and actin (*ACT*), and combined them with the RNA-Seq data of the pericarp at different developmental stages. The expression stability was calculated by using GeNorm, NorFinder, BestKeeper and ReFinder.

RNA-Seq database analysis and primer design

Commented [RA16]: It is not clear you solved this problem in this manuscript

Commented [RA17]: First, you must provide SRA Accession numbers for these datasets
Second, you must describe how the assembly was performed, or provide a bibliographic reference that describes the assembly

Commented [RA18]: These results must be described in detail and presented as part of the manuscript

Commented [RA19]: You must provide an NCBI accession number for this dataset

Commented [RA20]: These results must be described in detail and presented with figures as part of the manuscript

According to the screening conditions (Fig.S1), we screened candidate reference genes with relatively stable expression from the RNA-Seq database of *E. konishii* fruit in different development stages. To ensure accuracy and reliability of the data, NCBI-blast was used to verify candidate reference genes. Analysis of the transcriptome and expression of candidate reference genes are shown in Table 1. According to the nucleotide sequence of candidate reference genes and the design principle of qRT-PCR, primers of reference genes were designed using Primer Premier 6.0 software and were synthesized by Jinweizhi Biotechnology Co., LTD (Suzhou, China) and purified by PAGE.

Commented [RA21]: You must provide an analysis of genome paralog for the genes you selected and a graphical representation of the gene models for the selected genes

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°C for further experiment.

Candidate reference genes for RT-PCR amplification

The RT-PCR mixture contained 25 μL of 2 ×EasyTaq®PCR SuperMix, 0.4 μL of forward primer (10 μM), 0.4 μL of reverse primer (10 μM), and 5 μL of diluted cDNA in a final volume of 50 μL. PCR conditions: 40 cycles of 3 min at 94℃, 94℃ at 30 s, 56℃ at 10 s, 72℃ at 30 s, and 72℃ 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 μL of diluted cDNA, 5 μL of 2 × SYBR Green Master, 0.4 μL of forward primer (10 μM), and 0.4 μL of reverse primer (10 μM), with ddH2O added to achieve a final volume of 10 μL. All PCRs were performed using the QuantStudioTM Real-Time PCR System under the following conditions: 40 cycles of 2 min at 95℃, 5 s at 95℃ and 30 s at 60℃. The procedure ended with a melt-curve ramping from 60-95℃. The melting curve was analysed to determine primer specificity.

Statistical analysis

The instrument calculated the Ct 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, we used 9 genes related to the anthocyanin synthesis pathway to verify the expression of candidate reference genes, and the details are shown in Table 2. 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

Reference gene selection and primer design

The selection criteria of the reference genes we used came from the idea from the study of Kimmy. (Kimmy AS, Patrick PE& Joshua RP, 2017). First, we excluded genes with FPKM less than 5 since low expression genes used as reference genes are not suitable for quantifying the expression of the target gene. Then, we screened out 1131 genes with stable expressions based on the coefficient of variation and fold change values. Finally, 8 reference genes were selected from 1131 genes using Blast software to exclude false positive genes on the NCBI website (https://www.ncbi.nlm.nih.gov/). The details of 8 reference genes (*GAPDH, TUA, UBQ, mMDH, ACT, CYP, UBC* and *MDH*) are shown in Table 1. Forward and reverse primers of all candidate reference genes were designed using Prime Premier 6.0 software. All primers were synthesized by Jinweizhi Biotechnology Co., LTD (Suzhou, China). Primer details are shown in Table 3.

Candidate reference genes for RT-PCR amplification

PCR results of the 8 candidate internal reference genes are shown in Fig.S2 with a single band, no primer dimer and nonspecific amplification, which could be used for subsequent qRT-PCR analysis.

Cq values of candidate reference genes

The Cq values for all 8 reference genes are shown in Fig. 1. The Cq values varied from 24.65 (*EkGAPDH2*) to 35.27 (*EkACT7*). Moreover, as shown in Fig. 1, *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

Commented [RA22]: Where is this data?

Commented [RA23]: Again, data not presented

Commented [RA24]: Bad English
As written you make it sound that the false positives excluded are on the NCBI website

Commented [RA25]: This is Materials and Methods

Commented [RA26]: Same here

Commented [RA27]: PCR is fine but you must demonstrate specificity by sequencing the PCR amplified fragments

Commented [RA28]: You have not provided a descriptive legend for any of your figures. This is not acceptable

As shown in Fig. 2, 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

GeNorm, NormFinder, BestKeeper and ReFinder were widely used in the analysis of reference genes. Gene expression stability was determined by the M-value in GeNorm. The gene expression stability increases as the M-value decreases (Wu JY *et al*., 2018; Vandesompele J *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*, *EkmMDH2*, *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. 3-A.

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 CL *et al*., 2004). The analysis results of NormFinder shown in Fig. 3-B 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 MW *et al*., 2004). As the analysis results in Fig. 3-C show, the SD values of *EkACT7, EkmMDH2, 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. 3-D, the stability ranking obtained by ReFinder software is as follows: *EkUBC23, EkCYP38, EkGAPDH2, EkUBQ1, EkMDH, EkTUA3, EkmMDH*, 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

Commented [RA29]: Not clear at all why you are using these three different packages

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. 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.

Commented [RA30]: You need a figure of the pathway and the position of the genes you tested in the pathway Same for the other genes tested

Commented [RA31]: Data not presented

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 D *et al*., 2017), *Sedum alfredii* (Sang J *et al*., 2013) and *Dendrocalamus sinicus* [Guo XJ *et al*., 2018], and so on. In the present study, we selected 3 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 WX *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 SH *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 PK *et al*., 2017) and *E. konishii* by Liang *et al*. (Liang WX *et al*., 2018b). *UBC* has been widely accepted for normalization of gene expression as a reference gene (Yuan W *et al*., 2012). *UBC* was selected as the reference gene in the study of *Prunus pseudocerasus* (2 Zhu YY *et al*., 2015). However, it was unsuitable as a reference gene for *Oryza sativa* (Li QF *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 WC *et al*., 2014) and *Malus domestica* (Kumar G & Singh A K, 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 B& Rapacz M, 2013). However, *GAPDH* has different performances in different species and different experimental conditions. *GAPDH* shows stable expression in *Citrus sinensis* (Wu JX *et al*., 2014)and *Lycopersicon esculentum*(Mascia T *et al.,2010*), but it did not perform well in *S. viridis*(Martins PK *et al*., 2016) and *Panicum virgatum*(Jiang XM *et al*., 2014). In this study, we selected *EkGAPDH2* as the reference gene to quantify the high expression target genes for further study.

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 3 stable reference genes (*EkUBC2*3, *EkCYP38*, and *EkGAPDH*2) from 8 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 developmental stages.

Acknowledgements

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

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 Res.* 64(15): 5245-5250.

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): 1655-8.

Coker J S & 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.

Derveaux S, Vandesompele J & Hellemans J, 2010. How to do successful gene expression analysis using real-time PCR. *Methods.* 50(4): 227-230.

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, 112–119.

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.

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.

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 A Stanton, Patrick P Edger, Joshua R Puzey, 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.

Kozera B, Rapacz M,2013. Reference genes in Real- time PCR. *Journal of Applied Genetics*. 2013, 54(4): 391-406.

Kumar G & Singh A K, 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.

Liang WX, Ni L, Zou XX, Huang W, Zou SQ, 2018. Research progress on chemical constituents of *Euscaphis* and their pharmacological effect. *Chinese Traditional and Herbal Drugs.* (5): 1220-1226.

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.

313 Liang WX, Zou XX, Rebeca CL, Wu LJ, Sun HW, Yuan XY, Wu SQ, Li PF, Ding H, Ni L,
314 Huang W, Zou SQ, 2018. Selection and evaluation of reference genes for qRT-PCR analysis in
315 *Euscaphis konishii* Hayata based on transcriptome data. *Plant Methods.* 14(1): 42.
316 Li QF, Jiang MY, Yu HX, Xin SW, Gu MH, Liu QQ, 2008. Selection of internal reference genes
317 for quantitative RT-PCR analysis of total RNA from endosperm of rice (*Oryza sativa* L.).
318 *Jouranl of YangZhou university (agricultural and life science edition*). 29(2): 61- 66.
319 Li YC, Tian K, Sun LJ, Long H, Li LJ, Wu ZZ, 2016. A new hexacyclic triterpene acid from the
320 roots of *Euscaphis japonica*, and its inhibitory activity on triglyceride accumulation. *Fitoterapia.*
321 109: 261-265.
322 Martins PK, Mafra V, Souza WRD, Riberio AP, Vineky F, Basso MF, Cunhua BADB, Molinari
323 BC, 2016. Selection of reliable reference genes for RT-qPCR analysis during developmental
324 stages and abiotic stress in *setaria viridis. Sci Rep.* 20(6): 28348.
325 Mascia T, Santovito E, Gallitelli D, Fabrizio C, 2010. Evaluation of reference genes for
326 quantitative reverse transcription polymerase chain reaction normalization in infected tomato
327 plants. *Molecular Plant Pathology.* 11(6): 805-816.
328 Ma SH, Niu HW, Liu CJ, Zhang J, Hou CY, Wang DM, 2013. Expression stabilities of candidate
329 reference genes for rt-qpcr under different stress conditions in soybean. PloS ONE. 8(10):
330 e75271.
331 Nolan T, Hands RE & Bustin SA, 2006. Quantification of mRNA using real-time RT-PCR.
332 *Nature Protocols.* 1(3): 1559.
333 Pang QQ, Li ZL, Luo SB, Chen RY, Jin KM, Li ZX, Li DM, Sun BJ, Sun GW, 2017. Selection
334 and evaluation of reference genes for qRT-PCR analysis in Eggplant under high temperatures.
335 *Horticultural Plant Journal.* 44(3): 475-486.
336 Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP, 2004. Determination of stable housekeeping
337 genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool
338 using pair-wise correlations. *Biotechnology Letters.* 26(6):
339 Sang J, Han XJ, Liu MY,Qiao GR, Jiang J, Zhuo RY, 2013. Selection and validation of reference
340 genes for real-time quantitative PCR in hyper accumulating ecotype of *Sedum alfredii* under
341 different heavy metals stresses. *PLoS ONE.* 8(12): e82927.
342 Su XJ, Fan BG, Yuan LC, Cui XN, Lu SF, 2013. Selection and Validation of Reference Genes
343 for Quantitative RT-PCR Analysis of Gene Expression in *Populus trichocarpa. Bulletin of*
344 *Botany.* 48 (5): 507–518.
345 Vandesompele J, Preter K D, Pattyn F, Poppe E, Roy NV, Paepe AD, Speleman F, 2002.
346 Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of
347 multiple internal control genes. *Genome Biology.* 3(7): 34.
348 Xia W, Mason, Xiao Y, Liu Z, Yang YD, 2014. Analysis of multiple transcriptomes of the
349 African oil palm (*Elaeis guineensis*) to identify reference genes for RT- qPCR. *Journal of*
350 *Biotechnology.* 184: 63-73.

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.
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.
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.
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):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.

Supporting information

Figure 1 Cq values of candidate reference genes in 9 samples.

Figure 2 The melting curves of candidate reference genes.

Figure 3 Expression stablility analysis of reference genes bynormfinder (A), genorm (B), BestKeeper (C) and RefFinder (D).

Figure 4 The suitability of reference gene.

Figure S1 The selection of candidate reference genes.

Figure S2 Candidate reference genes for RT-PCR amplification.