Transcriptome-based screening and the optimal reference genes for real-time quantitative PCR in *Rehmannia chingii* and *R. henryi*

X. ZUO, F.-Q. WANG*, X.-R. LI, and M.-M. LI

College of Agronomy, Henan Agricultural University, Zhengzhou 450002, P.R. China

Abstract

Real time quantitative PCR (qPCR) is a powerful tool for studying the expression of specific genes. The accuracy and reliability of qPCR analysis data require the selection of reference genes with stable expression. However, the reference genes that can be used for qPCR of *Rehmannia chingii* and *R. henryi* have not yet been identified. In this study, based on the transcriptome data of *R. chingii* and *R. henryi*, we initially selected genes with relatively stable expression in different samples. We screened six candidate reference genes in *R. chingii* and *R. henryi* and calculated their expression abundance by real time qPCR. Their expression stability was evaluated by three algorithms *geNorm*, *NormFinder*, and *BestKeeper*. Although the results obtained by different algorithms were not completely consistent, *R. chingii type 2A phosphatase activator TIP41* and *R. chingii 18S ribosomal RNA* had the highest expression stability in six different samples of *R. chingii*, and *R. henryi 18S ribosomal RNA* and *R. henryi actin* showed the most stable expression in different samples of *R. henryi*. In addition, based on transcriptome data, four genes were screened in *R. chingii* and *R. henryi*, and the expression stability of the selected reference genes was further verified. This study laid the foundation for further analysis and verification of the functions of important genes in *R. chingii* and *R. henryi*.

Additional key words: *BestKeeper*, gene expression stability, *geNorm*, *NormFinder*.

Introduction

Real time quantitative PCR (qPCR) is an important method for studying the transcription and expression patterns of target genes. It has the advantages of accurate quantification, high sensitivity, high throughput, and good reproducibility (Bustin 2000, 2002, Bustin *et al*. 2005). The premise of qPCR for accurate correction and standardization of target gene expression is to screen out the reference gene (Huggett *et al*. 2005, Kozera and Rapacz 2013), and then to calculate the relative expression of a target gene by $2^{-\triangle\triangle Cq}$ method. This method is of the great significance for the study of gene expression regulation and biological functions. However, the stability of the reference gene directly affects the quantitative results of the target genes, and the relative stability of the reference gene in different species, tissues at different developmental stages, and under different experimental conditions is not constant and universal (Gutierrez *et al*. 2008). Therefore, screening and determining the most suitable reference gene with relatively stable expression characteristics according

to the research materials are the prerequisites to ensure the accuracy and reliability of qPCR results.

Diploid (n=28) *Rehmannia chingii* and *R. henryi* are two wild relatives of tetraploid (n=56) *R. glutinosa*. Zheng (2014) and Liu (2016) isolated eight new chemical components and nine iridoid compounds from the shoots and the whole plant of *R.chingii*. In addition, Zhou (2019a,b) isolated thirteen new iridoid compounds and an ionone glycoside from *R. henryi*. Among them, iridoid compounds have important biological activities in anti-inflammatory, anti-tumour and neuroprotection. *R. chingii* and *R. henryi* have obvious differences in appearance, especially the shape of flowers. Among them, the *R. chingii* flower is large and gorgeous, the corolla is fuchsia, there are yellow oil purple spots in the corolla tube, while the corolla of *R. henryi* is pale yellow-white, which is often used as one of the important characteristics to distinguish *R. henryi*. Different flower colours can attract insects to spread pollen, reflect light, and produce certain cell protection effects. Moreover, the material basis for the formation of different flower colours is mainly flavonoid

compounds such as anthocyanins and carotenoids, which have great development and application value. However, in *R. chingii* and *R. henryi*, the biomass of the leaves is the largest, and the flowers are also one of the important tissues in the aerial part. It can be seen that further research on the flowers and leaves of *R. chingii* and *R. henryi* is of great significance.

The analysis of gene expression is the basis of our further study of genes function and metabolic regulation network at the molecular level. Real time qPCR can further verify the accuracy and reliability of transcriptome results by measuring gene expression. However, there are no reports about the selection and determination of reference genes in *R. chingii* and *R. henryi*, which will limit further research. The use of transcriptome data has the advantages of high efficiency and ideal results for screening reference genes. Based on the transcriptome results, we tried to select genes with relatively high and stable fragments per kilobase of transcript per million fragments mapped (FPKM) values from different samples of *R. chingii* and *R. henryi*, combined with gene annotation and identified common reference genes in plants as candidates. They are *Actin*, 18S ribosomal RNA (*18S*), glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), elongation factor 1-alpha (*EF-1α*), ubiquitin (*UBQ*), and *TIP41*. Among them, actin supports the cytoskeleton; 18S participates in the translation of amino acids; GAPDH is a key enzyme in the glycolysis process; EF-1α participates in RNA transcription extension; UBQ participates in the plant's response to stress and has a proteolytic effect. Quatitative PCR analysis and statistical software *geNorm* (Vandesompele *et al.* 2002), *NormFinder* (Andersen *et al*. 2004), and *BestKeeper* (Pfaffl *et al.* 2004) were used to comprehensively evaluate the expression stability of these candidate genes in the corolla (five different developmental stages) and leaves of *R. chingii* and *R. henryi*. It is hoped that the most suitable reference genes will be screened out in *R. chingii* and *R. henryi*, respectively, which will provide a reference for further research on genes relative expression and function analysis in *R. chingii* and *R. henryi*.

Materials and methods

Collection of plant samples: Corollas of *Rehmannia chingii* Li. and *Rehmannia henryi* N.E. Brown in five stages including small flower bud stage (C1, bud length less than 1 cm), middle flower bud (C2, bud length between 2 and 3 cm), big flower bud (C3, bud length between 4 and 5 cm), opening flower (C4), and fully opened flower (C5) and leaves were used as experimental materials (Fig. 1). A single sample was taken by mixed sampling and three biological replicates were performed. Samples were quickly frozen in liquid nitrogen and stored in a refrigerator at -80 °C for RNA extraction.

Candidate reference genes selection and primers design: According to the gene annotation information obtained by transcriptome sequencing, combined with the FPKM values of the genes in different samples, we screened the common reference genes in six samples of *R. chingii* and *R. henryi* as candidates, namely *18S*, *GAPDH*, *EF-1α*, *UBQ*, *Actin*, and *TIP41*. The FPKM values of these genes are relatively high and stable in different samples. (You can find the specific information of these candidate reference genes through the online website https://pan.baidu.com/s/1Vjc26DUB1DqwvePNe2ryYA, the extraction code is 9yy6). The *ORF Finder* (https://www.ncbi.nlm.nih.gov/orffinder/), the *NCBI* online analysis software, was used to make an open reading frame (ORF) prediction of the cDNA sequence of each gene. Quantitative specific primers were then designed using the *NCBI-Primer BLAST* (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) online software based on the gene sequences (Table 1 Suppl.). The parameters are set as follows: primer length is generally around 20 bases, the length of the PCR product is 80 to 200 bp, the range of primer melting temperature (Tm) is 57 to 60 °C, and the optimal temperature is 58 °C. All primer sequences were synthesized by *Sangon Biotech Co.* (Shanghai, China) and primer PCR products were detected by 1 % (m/v) agarose gel electrophoresis (Table 1 Suppl.).

Total RNA extraction and cDNA synthesis: Total RNA was extracted from each sample according to the instructions of RNA extraction kit (*TaKaRa*, Dalian, China), and the starting amount of the sample was 50 mg. The total RNA obtained was treated with DNase to remove genomic DNA contamination. The concentration and purity of total RNA were measured using an ultra-micro nucleic acid protein analyzer, RNA was quantified by absorbance at 260 nm, and purity was evaluated by absorbance ratio at 260/280 nm. The integrity of total RNA was detected by agarose gel electrophoresis. The obtained RNA bands are clear and free of DNA and other impurities and can be used in the next experiment. The first strand of cDNA was synthesized using reverse transcription kit *6210A* (*TaKaRa*), which requires approximately 1 μg of total RNA, 1 mm³ of oligo dT primer (concentration: 50 μM), 1 mm³ of dNTP, and 0.5 mm³ of RNase. Inhibitor (40 U mm^{-3}), 1 mm³ of *PrimeScript II* reverse transcriptase (200 U mm^{-3}), and deionized water were added to form a total system of 20 mm³. The reaction procedure was: 65 °C for 5 min; 42 °C for 60 min; 95 °C for 5 min; and 4 °C for 2 min to terminate the reaction.
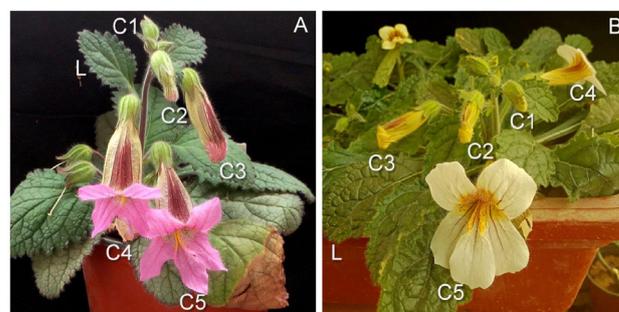

Fig. 1. Characteristics of buds and flowers in different growth periods (C1 to C5). *A - Rehmannia chingii*; *B - R. henryi.* L - leaf.

Real time qPCR: Reaction was performed according to the instructions of *SYBR® Premix Ex Taq™ II* kit (*TaKaRa*) and performed on a 96-well plate of the *Bio-Rad iQ5* real time PCR system (*Bio-Rad*, Hercules, USA), each reaction was repeated three times. The reaction system is followed: 12.5 mm³ of *SYBR® Premix Ex* enzyme, 1 mm³ of forward primer (10 μM), 1 mm³ of reverse primer (10 μM), 2.0 mm³ of cDNA template, and deionized water 8.5 mm³ for a total volume of 25 mm³. The qPCR reaction program was: 95 °C for 30 s; 95 °C for 5 s, 60 °C for 30 s, 40 cycles; after the last cycle, the set temperature was continuously increased between 60 °C and 95 C to obtain a melt curve. In addition to determining the Cq value of each candidate reference gene in different samples, a standard curve for each gene was also drawn to identify parameters such as the amplification efficiency (*E*) and correlation coefficient (R^2) of the primers. The cDNA solution obtained by reverse transcription was sequentially diluted 5, 10, 15, 20, and 25 times to draw a standard curve. Quantitative PCR was performed using the diluted cDNA as a template, and the obtained standard curve was analyzed by *CFX-Manager™* software. The parameters such as primer amplification efficiency and correlation coefficient were calculated. The amplification efficiency of primers was related to the slope of the standard curve, E = 10 $^{(-1/slope)}$ -1, and the applicable range was 95 ~ 110 %.

Data analysis: The Cq values of each candidate reference gene obtained by the qPCR analysis were transformed according to relevant requirements and then introduced into *geNorm*, *NormFinder*, and *BestKeeper* to analyze the stability of the candidate reference genes.

Validation of reference genes: 4-Coumarate-CoA ligase (4CL), cinnamic acid 4-hydroxylase (C4H), and phenylalanine ammonia-lyase (PAL) are the key enzymes in the phenylpropane metabolism pathway. 4CL and C4H are related to the synthesis of secondary metabolites such as lignin. PAL is a rate-limiting enzyme for the metabolism of phenylpropanes. Ethylene-responsive transcription factor (ERF) is involved in plant growth and development and plays a key role in response to plant stresses. EXPA8 is a member of the expansin family and an important part of plant cell walls. These key enzyme synthesis genes play an important role in the entire growth and development and secondary metabolism of plants. To verify the stability and reliability of reference genes in *R. chingii* and *R. henryi*, we screened *4CL*, *C4H*, *ERF*, and *PAL* genes in *R. chingii*, and *C4H*, *ERF*, *PAL*, and *EXPA8* in *R. henryi*. The qPCR was performed on these genes with the reference genes of *R. chingii* and *R. henryi*, respectively, and the relative expression of each gene was calculated according to the 2$^{-\Delta\Delta Cq}$ method. The quantitative results of these verification genes were compared with the corresponding FPKM values in the transcriptome data, and correlation analysis was further performed through the *DPS* data processing system (*v. 7.05*) to determine whether the reference gene is stable and reliable.

Results

Total RNA extracted from different tissues of *R. chingii* and *R. henryi* was detected by 1 % agarose gel, and 28S, 18S and 5S clear bands were seen (Fig. 1 Suppl.). The absorbance ratio A_{206}/A_{280} of all RNA was between 1.8 and 2.0, with an average value of 2.0 (SD = 0.05), indicating that the RNA had good integrity and quality, with no obvious degradation and could be used for cDNA synthesis.

Quantitative PCR analysis was performed using the diluted cDNA of *R. chingii* and *R. henryi*, and the qPCR products were all specific strips after detection by 1 % agarose gel electrophoresis (Fig. 2), and the fragment sizes were correct. Quantitative PCR results showed that the melt curve of each candidate reference gene also had a single main peak (Fig. 2 Suppl.). The standard curve of each candidate reference gene was obtained through the *CFX-Manager™* software, and the PCR amplification efficiency (*E*) and correlation coefficient R^2 (Table 1 Suppl.) of the primers were further obtained according to the standard curve, where the range of *E* needs to be 90 to 110 %. The above results indicate that qPCR primers of candidate reference genes have high specificity and amplification efficiency, and can be used in subsequent experiments.

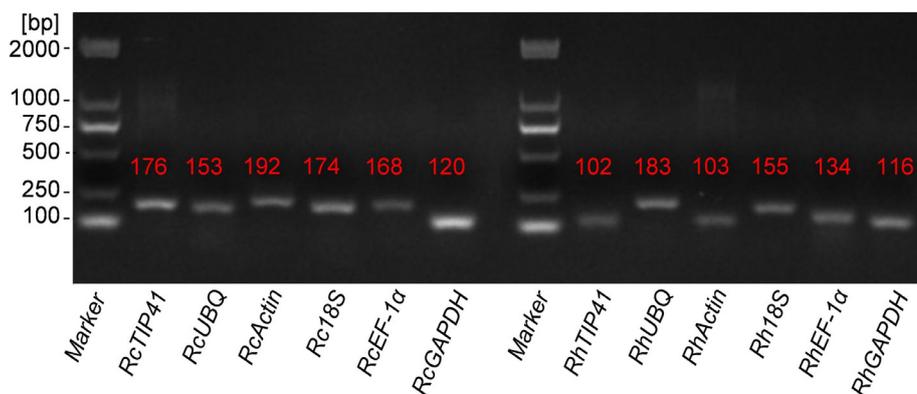Quantitative cycle (Cq) values of candidate reference

Fig. 2. Polymerase chain reaction products of candidate reference genes in *Rehmannia chingii* and *R. henryi.* For the names of the genes see Table 1.
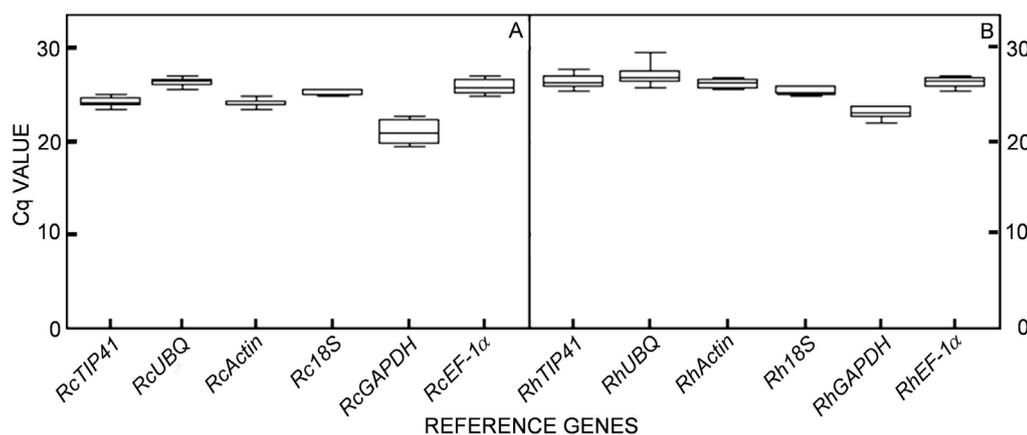
Fig. 3. Quantitative cycle values of candidate reference genes. *A - Rehmannia chingii, B - R. henryi*. The Cq values of each gene contain data from six different samples. For the names of the genes see Table 1. The *box-plots* contain means, interquartile ranges, non-outlier ranges, and outliers; *n* = 3.
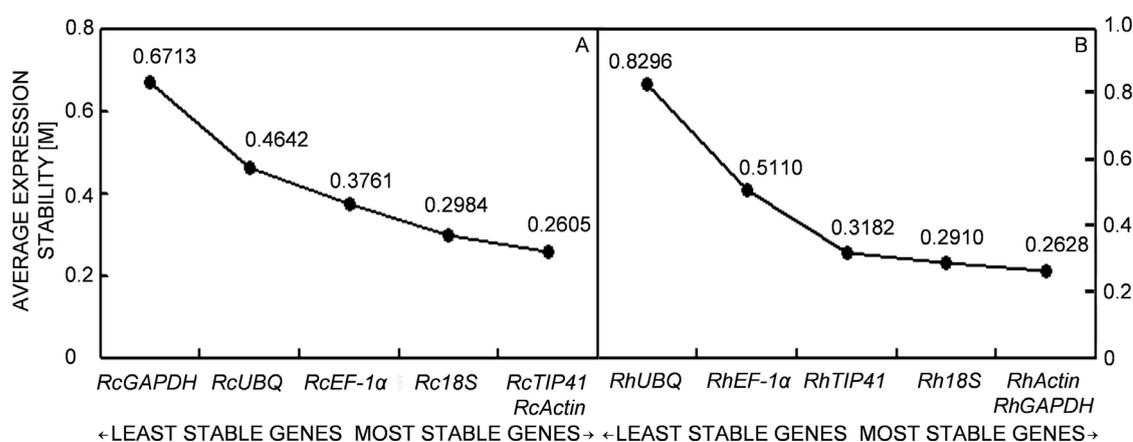

Fig. 4. Average expression stability of candidate reference genes calculated based on *geNorm*. *A - Rehmannia chingii* (*n* =18), *B - R. henryi* (*n* =18). A lower average expression stability (M value) indicates a higher stability. For the names of the genes see Table 1.

genes in all samples of *R. chingii* and *R. henryi* were shown in Fig. 3 and their Cq values are inversely related to the gene expression abundance. The Cq values of candidate reference genes fluctuated to a certain extent in all samples, indicating that the gene expression in these samples was not completely constant. From the results of qPCR, the Cq values of the six candidate reference genes in all samples were relatively consistent. The Cq values of *GAPDH* were relatively small, the Cq values of *UBQ* were relatively large, and the Cq values of *EF-1α* fluctuate greatly. The Cq values of *TIP41*, *Actin*, and *18S* in all samples were relatively moderate and concentrated. The fluctuation range of Cq values of *TIP41* gene was 23.57 to 26.87, the Cq values of the *Actin* gene were 23.62 to 27.03, and the Cq values of *18S* gene were 24.89 to 26.14.

Before using the *geNorm* algorithm for analysis, the Cq values of each gene need to be converted into relative expression by Q = $2^{-\triangle Cq}$, where the $\triangle$Cq value is equal to the Cq value of each corresponding candidate gene minus the minimum Cq value. The *geNorm* program calculates the stability measure (M) of each gene expression based on the average paired variation (V) of a specific gene relative

to all other reference genes. Generally, M < 1.5 is used as the stability threshold. The smaller the M value, the more stable the gene expression, and the more unstable conversely.

In addition, the $V_{n/n+1}$ value of the gene pair variation analysis calculated by the *geNorm* program can determine the optimal reference gene number, and the threshold is 0.15. When $V_{n/n+1}$ < 0.15, there is no need to introduce the n+1th reference gene. The average expression stability value (M) of each candidate reference gene in *R. chingii* and *R. henryi* was calculated by the *geNorm* program (Fig. 4). The default M threshold of the program is 1.5. According to the smaller M value, the higher the gene stability, the expression stability of each candidate reference gene was ranked. *GeNorm* analysis results show that the M values of the six candidate reference genes for both *R. chingii* and *R. henryi* were less than 1.5. In *R. chingii*, the highest expression stability had *RcTIP41* and *RcActin* (M = 0.26), and the lowest expression stability had *RcGAPDH* (M = 0.67). In *R. henryi*, *RhActin*, and *RhGAPDH* (M = 0.26) had the highest expression stability, and *RhUBQ* (M = 0.83) had the lowest expression
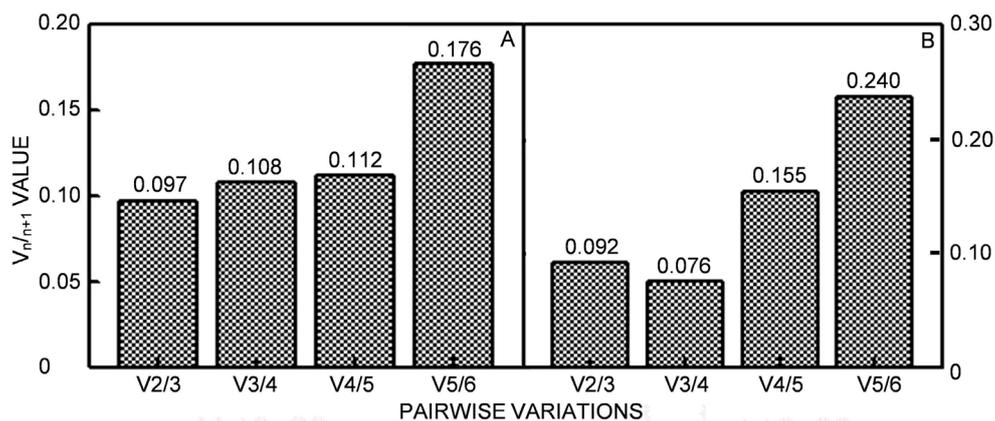
Fig. 5. Pairwise variation ($V_{n/n+1}$) analysis for selection of the number of real time quantitative PCR best reference genes with the *geNorm* algorithm. *A - Rehmannia chingii, B - R. henryi.*

stability.

The paired variation values ($V_{n/n+1}$) derived from the *geNorm* program were shown in Fig. 5. Taking $V_{n/n+1} = 0.15$ as the threshold, the $V_{2/3}$ of *R. chingii* and *R. henryi* are 0.097 and 0.092, respectively, which are both less than 0.15. Therefore, the optimal reference gene number during the quantitative normalization of *R. chingii* and *R. henryi* both are two, and there is no need to introduce a third reference gene.

NormFinder uses a conversion method similar to *geNorm* to calculate the expression stability of candidate reference genes. Unlike *geNorm*, the *NormFinder* program calculates the intra-group and inter-group variation and then combines the two results into the stability of each candidate gene (Schmidt and Delaney 2010).

The stability values (M) of six candidate reference genes were calculated by *NormFinder* in *R. chingii* and *R. henryi* (Table 1). According to the results of *NormFinder* analysis, the expression stability of six candidate reference genes in *R. chingii* and *R. henryi* were ranked from high to low. Among them, in *R. chingii*, *RcTIP41, RcEF-1α,* and *Rc18S* were relatively high in expression stability. The expression stability of *RhGAPDH, Rh18S*, and *RhActin* in *R. henryi* were relatively high. It can be seen that the results of *NormFinder* and *geNorm* programs were basically the

same in the stability calculation of reference genes of *R. chingii* and *R. henryi*.

The *BestKeeper* software can directly import the Cq value (geometric average of 3 repetitions) of the gene in the numerical input area. Through the obtained correlation coefficient (*r*), standard deviation (SD), and coefficient of variation (CV), the expression stability of candidate reference genes was analyzed (Li *et al.* 2019). The larger the correlation coefficient, the smaller the standard deviation, and the coefficient of variation, the better the stability of gene expression. The results are shown in Table 2. According to the combination of CV ± SD and *r* values, in *R. chingii*, the CV ± SD values of *Rc18S* and *RcTIP41* were relatively small, and the *r* values were relatively high. In *R. henryi*, *Rh18S* and *RhActin* had lower CV ± SD value and higher *r* value. Comprehensive analysis based on *BestKeeper* data showed that the genes with relatively stable expression in *R. chingii* were *Rc18S, RcActin,* and *RcTIP41*, and the most unstable gene was *RcGAPDH*. The relatively stable genes in *R. henryi* were *Rh18S* and *RhActin*, and *RhUBQ* had the most unstable expression. This result is basically consistent with the results of *geNorm* and *NormFinder*. Because the number of optimal reference genes was two, and combined with the analysis results of *geNorm*, *NormFinder*, and *BestKeeper*,

Table 1. The ranking of expression stability of candidate reference genes in *Rehmannia chingii* and *R. henryi* based on *NormFinder* analysis. *TIP41 - type 2A phosphatase activator TIP41, EF-1α - elongation factor 1-alpha, 18S - 18S ribosomal RNA, UBQ - ubiquitin, GAPDH - glyceraldehyde 3-phosphate dehydrogenase.*

Rank	*Rehmannia chingii*		*Rehmannia henryi*	
	gene name	stability value	gene name	stability value
1	*RcTIP41*	0.090	*Rh18S*	0.193
2	*RcEF-1α*	0.144	*RhActin*	0.264
3	*Rc18S*	0.183	*RhGAPDH*	0.298
4	*RcActin*	0.260	*RhTIP41*	0.300
5	*RcUBQ*	0.429	*RhEF-1α*	0.312
6	*RcGAPDH*	0.726	*RhUBQ*	0.990

Table 2 Expression stability of 12 candidate reference genes calculated by *BestKeeper*. CV - coefficient of variation, SD - standard deviation, *r* - coefficient of correlation. For the names of the genes see Table 1.

Rehmannia chingii				*Rehmannia henryi*			
gene name	CV	SD	*r*	gene name	CV	SD	*r*
Rc18S	1.13	0.29	0.95	*Rh18S*	1.65	0.42	0.66
RcActin	1.25	0.30	0.79	*RhActin*	1.67	0.44	0.70
RcUBQ	1.43	0.38	0.53	*RhEF-1α*	1.71	0.45	0.62
RcTIP41	1.76	0.43	0.97	*RhGAPDH*	2.10	0.49	0.80
RcEF-1α	2.54	0.66	0.98	*RhTIP41*	2.24	0.58	0.85
RcGAPDH	6.11	1.29	0.95	*RhUBQ*	3.00	0.81	0.29
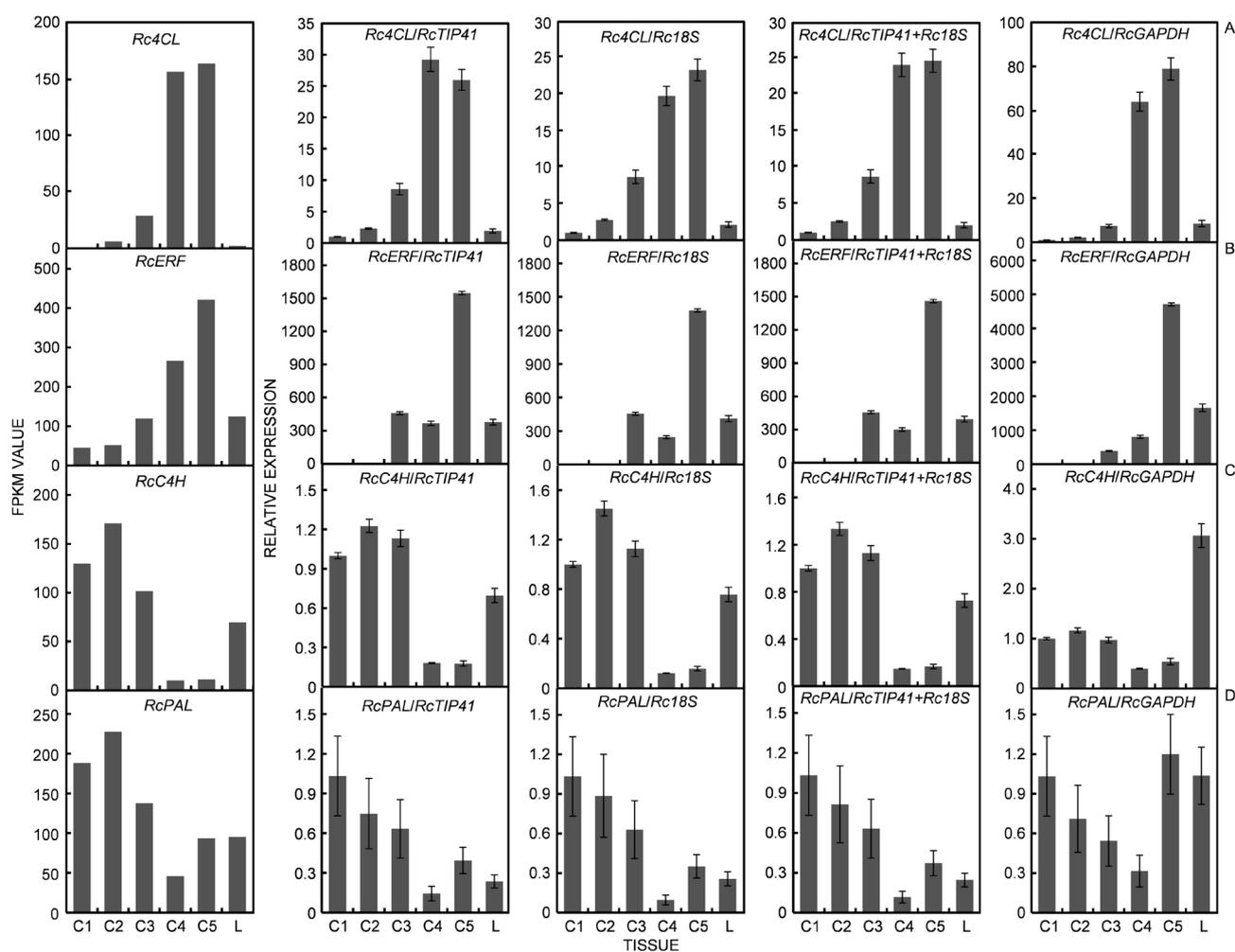
Fig. 6. The fragments per kilobase of transcript per million fragments mapped (FPKM) value of verified genes in *Rehmannia chingii* and their relative expressions using different reference genes. C1 to C5 - flowers (buds) at different developmental stages, L - leaf. *A - R. chingii 4-Coumarate-CoA ligase, B - R. chingii ethylene-responsive transcription factor, C - R. chingii cinnamic acid 4-hydroxylase, D - R. chingii phenylalanine ammonia-lyase.* For the names of other genes see Table 1. Means ± SDs, *n* = 3.

the optimal reference genes of *R. chingii* were *Rc18S* and *RcTIP41* and the optimal reference genes of *R. henryi* were *Rh18S* and *RhActin*.

Using *RcTIP41*, *Rc18S*, *RcTIP41* + *Rc18S*, and *RcGAPDH* as reference genes we verified the expressions

of the *Rc4CL*, *RcERF*, *RcC4H*, and *RcPAL* (Fig. 6). The correlation analysis of the qPCR results of the four verification genes under different reference genes conditions and the FPKM values of the transcriptome in *R. chingii*. The results show that when the stably
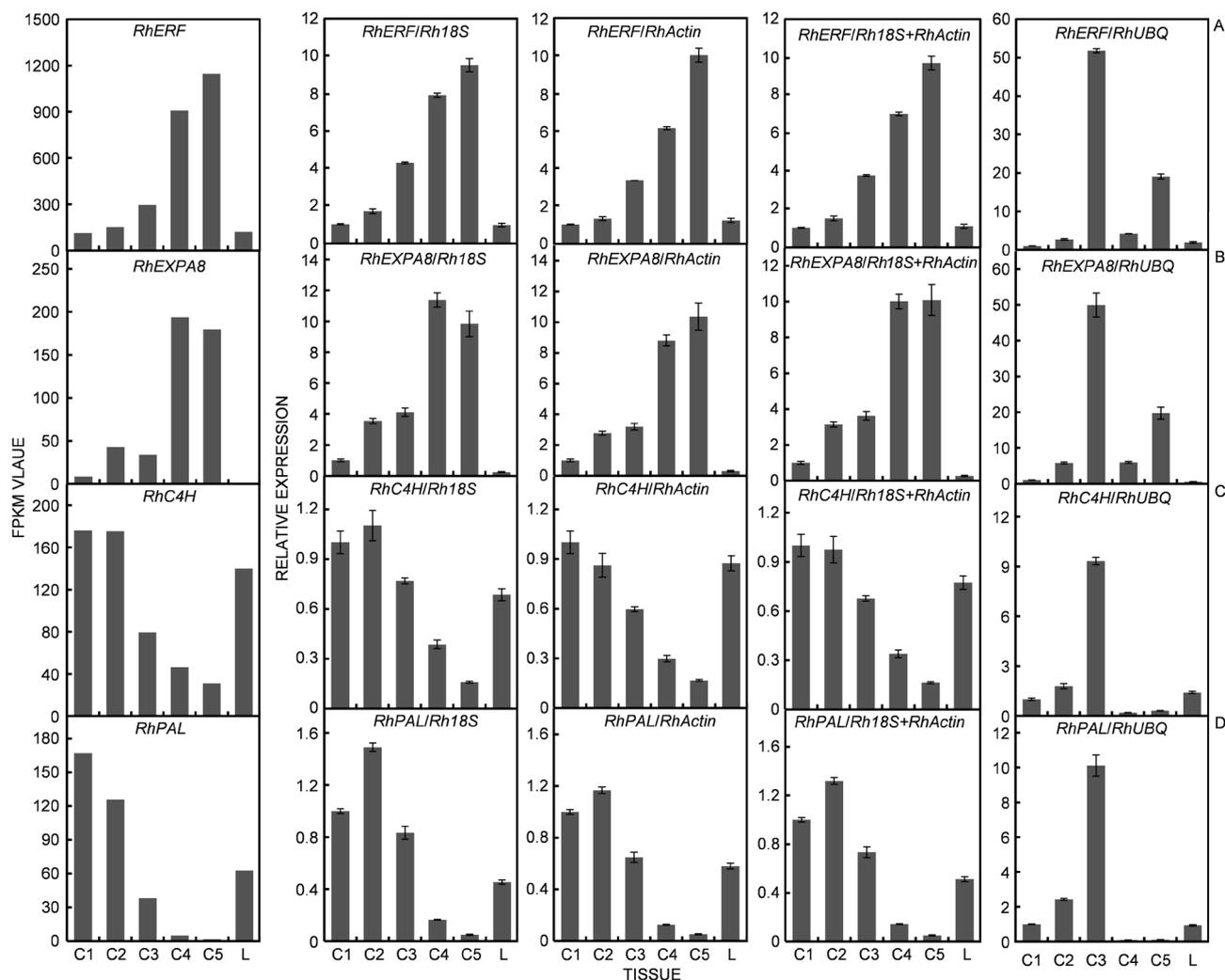
Fig. 7. The fragments per kilobase of transcript per million fragments mapped (FPKM) value of the verified genes in *Rehmannia henryi* and their relative expression using different reference genes. C1 to C5 - flowers (buds) at different developmental stages, L - leaf. *A - R. henryi cinnamic acid 4-hydroxylase, B - R. henryi phenylalanine ammonia-lyase, C - R. henryi ethylene-responsive transcription factor, D - R. henryi expansin-A8.* For the names of other genes see Table 1. Means ± SDs, *n* = 3.

expressed *RcTIP41, Rc18S, RcTIP41 + Rc18S* were used as reference genes for *R. chingii*, the changes in the relative expression of all the verification genes in all samples were well consistent with the FKPM values (Fig. 6). However, when qPCR was performed with *RcGAPDH* as the reference gene, the expression trends of *RcC4H* and *RcPAL* in different samples were inconsistent with their FPKM values of transcriptome analysis. Correlation analysis showed that relative expressions and FPKM values of the four verification genes were significantly correlated (*P* < 0.05) when *RcTIP41, Rc18S* or *RcTIP41 + Rc18S* were used as reference genes (Table 2 Suppl.). When *RcGAPDH* was used as a reference gene, the correlation between the relative expressions and FPKM values of *RcC4H* and *RcPAL* genes didn´t reach a significant level (Table 2 Suppl.). The results of verification experiments suggested that *RcTIP41, Rc18S* or *RcTIP41 + Rc18S* were good reference genes (reference gene combinations) for expression analysis in flower and leaf development of

R. chingii.

Similar to the verification results in *R. chingii*, when *Rh18S, RhActin, Rh18S + RhActin* were used as reference genes of *R. henryi*, the change trends in gene expression of *RhC4H, RhPAL, RhERF* and *RhEXPA8* were basically consistent with the results of the transcriptome (Fig. 7). When *RhUBQ* was used as a reference gene, the consistency of the changes in the expression levels of the four verification genes and the transcriptome data was relatively poor (Fig. 7). Correlation analysis results showed that when *Rh18S, RhActin* and *Rh18S + RhActin* were used as reference genes, the correlation between the relative expressions of the four verification genes (*RhERF, RhEXPA8, RhC4H* and *RhPAL*) and the FPKM values in the RNA-Seq data reached a significant level (*P* < 0.05) (Table 3 Suppl.). However, when *RhUBQ* was used as a reference gene, the correlation analysis results of the relative expressions of the four verification genes and FPKM values were not significant (Table 3 Suppl.).

This shows that *RhActin*, *Rh18S*, *RhActin + Rh18S* were suitable as reference genes (reference gene combinations) for gene expression level analysis during flowers and leaves development of *R. henryi*.

Discussion

Quantitative PCR is an important method for gene expression analysis, and the key to accurate analysis of gene expression by real time qPCR is to select the suitable and stable reference gene. The ideal reference gene is expressed in all cells and tissues of the organism (Tong *et al*. 2009), the stability of gene expression is not affected by external environmental conditions (Huggett *et al*. 2005), and the expression abundance of the reference gene and the target gene is similar (Paolacci *et al*. 2009).

GeNorm, *NormFinder*, and *BestKeeper* are all free statistical algorithms based on *Excel*, which are specifically used for reference gene stability calculations and are widely used and convenient. *GeNorm* and *NormFinder* need to convert the Cq value to the △Cq value to analyze the stability of the gene, and *BestKeeper* directly compares the standard deviation and the coefficient of variation of the Cq value to evaluate the stability of the gene. *GeNorm* analyzes the M values of candidate reference genes by pairwise comparison (Vandesompele *et al*. 2002), which can precisely control the error caused by the cDNA input because the systematic error can affect the two paired genes at the same time (Expósito-Rodríguez *et al*. 2008). In addition, *geNorm* can not only determine the most suitable reference genes, but also the number and combination of the most suitable genes. However, *geNorm* is more sensitive to co-regulatory genes, which makes it more inclined to select genes with higher expression similarity (Andersen *et al*. 2004). *NormFinder* can make up for the shortcomings of *geNorm*, it is not sensitive to co-coordination, and can analyze the differences within and between the sample groups to balance the sources of variation, but *NormFinder* does not consider the systematic errors generated during the sample preparation process (Expósito-Rodríguez *et al*. 2008, Paolacci *et al*. 2009). In addition, to analyze the reference genes, *BestKeeper* can also compare the expressions of target genes. Therefore, when using three different algorithms to comprehensively evaluate the stability of candidate genes, the results are more stable and reliable.

So far, no related research has been reported on diploid *R. chingii* and *R. henryi* reference gene screening, which has limited the application of qPCR in these organisms. In this study, *Actin*, *GAPDH*, *18S*, *UBQ*, *EF-1α*, and *TIP41*, which are six common endogenous reference genes in plants, were screened based on FPKM data of genes in the transcriptome as candidate genes (Guénin *et al*. 2009, Manoli *et al*. 2012). The results of qPCR and gene stability analysis showed that the expression abundance and stability of expression of the six candidate reference genes in different developmental stages of *R. chingii* and *R. henryi* were not consistent. This further validates the previous view that the stability of the expression of reference genes

in different species and tissues is not completely constant (Suzuki *et al*. 2000, Paolacci *et al*. 2009). In addition, the standardization of a single reference gene is prone to large errors (Thellin *et al*. 1999, Nicot *et al*. 2005). Based on the comprehensive evaluation of $V_{n/n+1}$, *geNorm*, *NormFinder*, and *BestKeeper* results, the number of optimal reference genes for *R. chingii* and *R. henryi* were determined to be two.

In *R. chingii*, the results of *geNorm* and *NormFinder* programs showed that *RcTIP41*, *RcActin*, and *Rc18S* gene expression stability was relatively high, but *BestKeeper* results showed that *r* (0.79) of *RcActin* gene was lower than that of *RcTIP41* (0.97) and *Rc18S* (0.95). The expression stability of *RcUBQ* and *RcGAPDH* genes were relatively poor, and qPCR showed that the expression abundance of *RcUBQ* gene was relatively low and the expression abundance of *RcGAPDH* gene was relatively high. Therefore, *RcUBQ* and *RcGAPDH* genes are not suitable as reference genes for *R. chingii*. In *R. henryi*, the results of *geNorm* and *NormFinder* showed that the gene expression stability of *Rh18S* and *RhActin* was the best, and it was the same as in *BestKeeper* analysis. While the expression stability of *RhEF-1α* and *RhGAPDH* was relatively poor, they were not suitable as reference genes. Although the stability of *RhGAPDH* gene expression in *R. henryi* was relatively high, it was also not suitable as a reference gene for *R. henryi* because of the high abundance of *RhGAPDH* expression. *18S* is one of the common reference genes. It has a highly conserved sequence and plays an important role in maintaining basic cellular metabolism and intracellular functions (Weigand *et al*. 2012, Yeap *et al*. 2014). It is often controversial that the expression of *18S* gene is too high, but in this study, the *18S* gene was moderately abundant in different tissues of *R. chingii* and *R. henryi*, 25.31 (SD = 0.35) and 25.48 (SD = 0.49), so it can be used as a reference gene.

Conclusions

Combining the expression abundance of candidate reference genes and the gene stability results obtained by *geNorm*, *NormFInder*, and *BestKeeper*, it was finally determined that the suitable reference genes for *R. chingii* were *RcTIP41* and *Rc18S*, and for *R. henryi* were *Rh18S* and *RhActin*. This study can provide appropriate references for the standardization and normalization of quantitative expressions of related genes in *R. chingii* and *R. henryi*, and lay a foundation for further research on the functions of key genes in *R. chingii* and *R. henryi*.

References

Andersen, C.L., Jensen, J.L., Ørntoft, T.F.: Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. - Cancer Res. **64**: 5245-5250, 2004.

Bustin, S.A.: Absolute quantification of mRNA using real-time

reverse transcription polymerase chain reaction assays. - J. mol. Endocrinol. **25**: 169-193, 2000.

Bustin, S.A.: Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. - J. mol. Endocrinol. **29**: 23-39, 2002.

Bustin, S.A., Benes, V., Nolan, T., Pfaffl, M.W.: Quantitative real-time RT-PCR – a perspective. - J. mol. Endocrinol. **34**: 597-601, 2005.

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, 2008.

Guénin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C., Gutierrez, L.: Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. - J. exp. Bot. **60**: 487-493, 2009.

Gutierrez, L., Mauriat, M., Pelloux, J., Bellini, C., Van Wuytswinkel, O.: Towards a systematic validation of references in real-time RT-PCR. - Plant Cell **20**: 1734-1735, 2008.

Huggett, J., Dheda, K., Bustin, S., Zumal, A.: Real-time RT-PCR normalisation; strategies and considerations. - Genes Immun. **6**: 279-284, 2005.

Kozera, B., Rapacz, M.: Reference genes in real-time PCR. - J. appl. Genet. **54**: 391-406, 2013.

Liu, Y.F., Shi, G.R., Wang, X., Zhang, C.L., Wang, Y., Chen, R.Y., Yu, D.Q.: Bioactive iridoid glycosides from the whole plants of *Rehmannia chingii*. - J. Natur. Prod. **79**: 428-433, 2016.

Li, L., Wang, K.Y., Zhao, M.Z., Li, S.K., Jiang, Y., Zhu, L., Chen, J., Wang, Y.F., Sun, C.Y., Chen, P., Lei, J., Zhang, M.P., Wang, Y.: Selection and validation of reference genes desirable for gene expression analysis by qRT-PCR in MeJA-treated ginseng hairy roots. - PLoS ONE **14**: e0226168, 2019.

Manoli, A., Sturaro, A., Trevisan, S., Quaggiotti, S., Nonis, A.: Evaluation of candidate reference genes for qPCR in maize. - J. Plant Physiol. **169**: 807-815, 2012.

Nicot, N., Hausman, J.F., Hoffmann, L., Evers, D.: Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. - J. exp. Bot. **56**: 2907-2914, 2005.

Paolacci, A.R., Tanzarella, O.A., Porceddu, E., Ciaffi, M.: Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. - BMC mol. Biol. **10**: 11, 2009.

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.

Schmidt, G.W., Delaney, S.K.: Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. - Mol. Genet. Genomics **283**: 233-241, 2010.

Suzuki, T., Higgins, P.J., Crawford, D.R.: Control selection for RNA quantitation. - Biotechniques **29**: 332- 337, 2000.

Tong, Z.G., Gao, Z.H., Wang, F., Zhou, J., Zhang, Z.: Selection of reliable reference genes for gene expression studies in peach using real-time PCR. - BMC mol. Biol. **10**: 71, 2009.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., Heinen, E.: Housekeeping genes as internal standards: use and limits. - J. Biotechnol. **75**: 291-295, 1999.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F.: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. - Genome Biol. **3**: research0034.1, 2002.

Weigand, A.M., Dinapoli, A., Klussmann-Kolb, A.: 18S rRNA variability map for Gastropoda. - J. Mollus. Stud. **78**: 151-156, 2012.

Yeap, W.C., Loo, J.M., Wong, Y.C., Kulaveerasingam, H.: 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 Cult. **116**: 55-66, 2014.

Zheng, C.J., Wu, Y., Zhu, J.Y., Zhao, X.X., Qin, L.P.: Chemical constituents of the aerial parts of Rehmannia chingii. - Chem. natur. Comp. **50**: 560-561, 2014.

Zhou, J., Shi, G.R., Liu, Y.F., Chen, R.Y., Yu, D.Q.: Five new iridoids from the whole plants of *Rehmannia henryi*. - J. asian nat. Products Res. **21**: 727-734, 2019a.

Zhou, J., Shi, G.R., Liu, Y.F., Chen, R.Y., Yu, D.Q.: Nine new compounds from the whole plants of *Rehmannia henryi*. - J. asian natur. Products Res. **21**: 399-408, 2019b.