Biotechnol Lett (2015) 37:1483–1493
DOI 10.1007/s10529-015-1810-8

ORIGINAL RESEARCH PAPER

Selection and validation of reliable reference genes in *Gossypium raimondii*

Runrun Sun · Qiuling He · Baohong Zhang · Qinglian Wang

Received: 12 December 2014 / Accepted: 5 March 2015 / Published online: 24 May 2015

Abstract
Objective To identify reliable reference genes for gene expression analysis in *Gossypium raimondii*.
Results Five different software tools, geNorm, Norm Finder, BestKeeper, ReFinder and ΔCt method were employed to analyze the qRT-PCR data systematically of 12 housekeeping genes. SAD and TUA11 showed relatively stable expression levels in all tissues (i.e. leaves, shoots, buds, and sepals). We then limited our analysis to each plant part and identified tissue-specific reference genes. Our results showed TUA11, TUB6 and EF1a, EF1a, MZA and GAPC2, MZA, GAPC2, SAD and TUA11, and UBQ and MZA were reliable reference genes in leaves, shoots, buds, and sepals, respectively.
Conclusion Some genes were commonly identified as candidate reference genes in more than two tissue, while others were tissue-specific. Thus, our study allows choosing an appropriate control gene based on sampling for gene expression analysis.

Keywords Gene expression · *Gossypium raimondii* · qRT-PCR · Reference genes

Abbreviations

EF-1a	Elongation factor-1A
UBQ	Polyubiquitin
ACT	Actin
TUA	α-Tubulin
TUB	β-Tubulin
GAPC2	Glyceraldehyde-3-phosphate dehydrogenase C-2
PTB	Polypyrimidine tract-binding protein homolog
PP2A	Catalytic subunit of protein phosphatase 2A (PP2A)
SAD	Stearoyl-ACP desaturase
GM (CP)	The geometric mean of CP
AR (CP)	The arithmetic mean of CP
Min (CP) and Max (CP)	The extreme values of CP
SD (±CP)	The standard deviation of the CP
CV (% CP)	The coefficient of variance expressed as a percentage on the CP level
r	Pearson correlation coefficient
SD	Standard deviation

Electronic supplementary material The online version of this article (doi:10.1007/s10529-015-1810-8) contains supplementary material, which is available to authorized users.

R. Sun · B. Zhang · Q. Wang (✉)
Henan Institute of Sciences and Technology, Xinxiang 453003, Henan, People's Republic of China
e-mail: zhangb@ecu.edu

R. Sun · Q. He · B. Zhang
Department of Biology, East Carolina University, Greenville, NC 27858, USA

Introduction

Gene expression analysis is a fundamental step in biological research. Common experimental techniques, including northern blot hybridization, microarray and quantitative real-time reverse transcription PCR (qRT-PCR), are employed to evaluate gene expression levels. qRT-PCR is the most widely used method for quantifying differential expression. It is a powerful technique due to its high sensitivity, specificity, accuracy and broad quantification range (Bustin 2005; Bustin et al. 2005). Nevertheless, many factors should be taken into account for reliable analysis, including the amount of starting material (i.e., RNA concentration), cDNA quality, primer specificity, and accurate normalization. Therefore, it is vital to select stable reference genes to control for biological and technical variations for each sample to obtain accurate results (Mahoney et al. 2004).

The selection of reference genes is critical for analyzing expression of genes of interest. Housekeeping genes are considered to be constitutively and stably expressed regardless of growth conditions or developmental stage. Therefore, they were considered as the best reference genes. Some of the popular housekeeping gene used for qRT-PCR analysis in plants and animals include elongation factor-1A (EF-1a), polyubiquitin (UBQ), actin (ACT), and α-tubulin and β-tubulin (TUA and TUB, respectively), Glyceraldehyde-3-phosphate dehydrogenase C-2 (GAPC2), Polypyrimidine tract-binding protein homolog (PTB), catalytic subunit of protein phosphatase 2A (PP2A), stearoyl-ACP desaturase (SAD) (Andersen et al. 2004; Brunner et al. 2004; Artico et al. 2010; Kim et al. 2003; Goidin et al. 2001). Nevertheless, the idea of universal housekeeping reference genes has been disproved as they are affected by different experimental conditions (Ding et al. 2004; Nicot et al. 2005; Czechowski et al. 2005). To identify a reliable reference, several computational programs have been developed including geNorm (Vandesompele et al. 2002), NormFinder (Andersen et al. 2004), BestKeeper (Pfaffl et al. 2004), ReFinder (Xie et al. 2012) and ΔCt method (Silver et al. 2006), all of which are based on different statistical algorithms.

Identifying reliable reference genes have been performed in different plants, including *Arabidopsis thaliana* (Remans et al. 2008), wheat (Paolacci et al. 2009), barley (Faccioli et al. 2007), rice (Jain et al. 2006), soybean (Jian et al. 2008), grape (Reid et al. 2006), tomato (Exposito-Rodriguez et al. 2008), potato (Nicot et al. 2005) and peach (Tong et al. 2009). However, few studies focused on the reference genes of cotton (Wang et al. 2013; Artico et al. 2010). Cotton is one of the most important crop plants, which provides high quality of natural fiber and edible oil and brought great economic value. Several groups, including our own, are interested in understanding the genetic pathways of fiber development (Xie et al. 2015) and oil biosynthesis. For that, gene expression analysis is inevitable. *Gossypium hirsutum* and *G. barbadense* share a common ancestor, *G. raimondii,* whose genome has been sequenced (Paterson et al. 2012). *G. raimondii* has become a cotton model species for genetic and functional investigations; however, there have been no reports on suitable reference genes in it. Additionally, a single housekeeping gene is insufficient to normalize the expression level of genes under different experiment conditions (Gutierrez et al. 2008b; Schmittgen and Zakrajsek 2000). Therefore, in this study, we have evaluated 12 housekeeping genes in four different tissues of *G. raimondii* for qRT-PCR data normalization. We found that the expression of SAD and TUA11 are relatively stable in different tissues among all of the samples.

Materials and methods

Plant material and growth conditions

Gossypium raimondii was grown in the greenhouse in East Carolina University as our previous report (Ma et al. 2015). The flower buds were collected which the diameter was approximately 1 cm, and the sepal was separated from the bud. At the same time, the first-outstretched leaves and shoots on the top of plant were collected. Four tissues were frozen in liquid nitrogen immediately and stored at −80 °C for further analysis.

Selecting housekeeping genes and primer design

To identify suitable reference genes, the first critical step is to choose a series of housekeeping genes. In this study, 12 housekeeping genes previously used as reference genes in cotton (*G. hirsutum*) or other model plants were selected, including TUA10 (α-tubulin 10),

TUA11 (α-tubulin 11), TUB6 (tubulin 6), MZA (clathrin adaptor complexes medium subunit family protein), GAPC2 (glyceraldehyde-3-phosphate dehydrogenase C-2), ACT (actin), EF1a (elongation factor 1A), PTB (polypyrimidine tract-binding protein homolog), UBQ (ubiquitin extension protein), PP2A (catalytic subunit of protein phosphatase 2A), SAD (stearoyl-ACP desaturase), FBX (F-box family protein) (Artico et al. 2010). Primers were designed by Primer-blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) based on the obtained cDNA sequences obtained from the Joint Genome Institute (JGI) Phytozome v9.1: *Gossypium raimondii* v2.1 website (http://www.phytozome.net/cotton.php).

RNA extraction and quantitative real-time PCR

According to our previous report (Zhang and Pan 2009), 0.03–0.1 g tissues were prepared to RNA extraction. Total RNA was extracted from four tissue using mirVana miRNA Isolation Kit (Ambion, Austin TX) according to the production introduction. The quantity and quality of total RNA were determined by NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA), which is based on the ratios of $A_{260/280}$ and $A_{260/230}$. The TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA) was used to generate the 1st strand of cDNA for an individual gene. For each sample, 800 ng RNA was used for reverse transcription. The reaction volume was 15 μl, including 0.15 μl dNTPs, 1.5 μl reverse transcription buffer, 1 μl MultiScribe Reverse Transcriptase, 0.19 μl RNase inhibitor, 800 ng total RNAs. Reverse transcription started with 30 min at 16 °C for primer binding with the RNA, 30 min at 42 °C for elongation, followed by 5 min at 85 °C for denaturation, and was finally held at 4 °C. Before qRT-PCR, the products were diluted with 80 μl DNase/RNase-free water and stored in −20 °C.

qRT-PCR was performed to determine the expression levels of 12 housekeeping genes in four different *G. raimondii* tissues. Gene-specific reverse and forward primers were used for each target gene (Supplementary Table 1). All reactions were performed using the Applied Biosystems ViiA 7 Real Time PCR System (Foster City, CA, USA). PCR reactions were performed under the following program: 10 min at 95 °C and 45 cycles of the thermal cycling of 15 s at 95 °C for denaturation and 60 s at 60 °C for annealing and extension in a 384-well plate. Each reaction was performed in triplicate (technical replicates) on three individual samples (biological replicates). SPSS (20) was used to generate boxplots using Ct values for the 12 housekeeping genes. Differences in gene expression were assessed using one-way analysis of variance (ANOVA). Relative quantities were calculated in Excel before being imported to the gene expression stability programs.

Determination of gene stability

Five common statistical approaches: geNorm, ΔCt method, NormFinder, BestKeeper, and RefFinder were preformed to identify the stability of each reference gene.

geNorm (Vandesompele et al. 2002) determines the gene stability by pairwise comparisons among all candidate genes. The expression stability value (M value) is calculated, and the gene with highest M-value is eliminated. The process is then repeated until there are only two genes left. The lowest M value is recommended as the optimum reference gene (Vandesompele et al. 2002).

The comparative ΔCt method was the only one that uses an excel sheet without any other applet. It is similar to geNorm and also depends on the pairwise comparison (Silver et al. 2006). Briefly, the mean and SD of each pair candidate genes and the average SD of each gene are calculated, and the gene with lowest SD is considered as the reliable reference gene. Boxplots were done using SPSS, different colors represent different pairs.

NormFinder uses an ANOVA-based model (Andersen et al. 2004). The method calculates a stability value for all candidate genes tested that is based on the combined estimate of intra- and inter-group expression variations of the genes.

BestKeeper was another Excel-based application to evaluate the gene expression stability (Pfaffl et al. 2004). The expression level of all candidate genes was detected via three variables: the standard deviation (Goetz et al. 2006), coefficient of correlation (r) and coefficient of variance (CV). The mean of Ct values for each sample across all candidate genes are combined to form the BestKeeper index. Subsequently, each candidate gene is tested in a pair-wise way via Pearson correlation coefficients, the coefficient of determination (r^2) and the P value. The result is a ranked order of their stability. The highest ranked gene is the most stable. Any candidate gene with the SD

value higher than 1 is considered as least reliable and should be excluded.

ReFinder (http://www.leonxie.com/referencegene.php) applies four computational programs, NormFinder, BestKeeper, GeNorm, and the comparative ΔCt, to produce a comprehensive stability value for each gene (Xie et al. 2012). The Ct value of each gene was input directly. Finally, ReFinder calculates the geometric mean of each gene to reach its overall final ranking. A lower geometric mean of ranking value indicates more stable expression.

Results

Descriptive statistics for the expression levels of 12 reference gene candidates

RNA transcript levels of the 12 candidate reference genes were assessed in four different tissues: young leaves, shoots, buds and sepals. All tested genes were expressed except FBX; which was subsequently removed the following analyses (Fig. 1). Based on our results, SAD and TUA11 were the most stable reference genes in the different organs and overall level. PTB, UBQ and PP2A showed higher variations in their expression levels. Among the 12 tested candidate reference genes, SAD mRNA was the most abundant whereas UBQ mRNA was the least abundant in all four tissues. The Ct median values across the candidate housekeeping genes ranged from 20.67 to 44.3 (Table 1). The wide range of Ct values imply that these candidate genes had different expression levels in the four tissues examined. SAD, TUA11, TUA10 and MZA had the lower Ct values with the $C_{t\ median}$ of 20.67, 22.37, 31.53 and 31.66, respectively. However, PTB, UBQ and GAPC2 showed largest Ct values with the $C_{t\ median}$ of 42.01, 44.3, and 40.85, respectively. Genes with SD values exceeding 1 were considered as unstable. The least variable reference genes were TUA11, TUB6 and SAD with the SD value of 0.74, 0.96 and 0.8, respectively. Conversely, PTB, UBQ and PP2A were the most variable genes (SD value is 3.59, 3.71 and 3.49, respectively). Although some of these genes showed a lower expression variation between the different tissue samples, it is still necessary to perform further analyses in order to identify the most suitable combination of these genes candidates for normalizing gene expression.

geNorm analysis of candidate reference gene stability

The transformed data of reference gene candidates were calculated using geNorm based on the geometric mean of the SD, which was defined as the M value for a putative reference gene. The stability of each reference gene was ranked by its M value, and genes with the lowest M values were considered the most stable in expression. $M < 1.5$ was used as criterion for stable gene expression (Zhang et al. 2012). The results of geNorm showed that the M value of most reference genes was less than 1.5 except UBQ, PTB and PP2A that were Fig. 2 treated as least stability genes (Fig. 2). Among those reference genes, the most stable genes were TUA11 and SAD with the same M value of 0.47. Overall, TUA11, SAD, TUB6, GAPC2, EF1A were good reference genes. In decreasing order, the M-based ranking of reference genes from the most stable (lowest M value) to the least stable one (highest M value) was: SAD, TUA11, TUB6, GAPC2, EF1A, ACT, MZA, TUA10, PP2A, PTB and UBQ.

We also analyzed the 11 reference candidates using geNorm, which ranks reference gene candidates based on the pairwise variation $Vn/n + 1$. Interestingly, the V-value for all the gene pairs was higher than 0.15 (Fig. 3). This suggests that the use of multiple housekeeping genes as reference genes is not required for reliable normalization.

Reference gene ranking based on comparative ΔCt method

ΔCt method compares differential expression of 'gene pairs'. If the dCt value of two genes remains constant in different samples, then both genes are stably coexpressed (Cassan-Wang et al. 2012). Our results showed that SAD, TUB6, TUA11 were better reference genes than others (Table 2), while GAPC2, EF1a, ACT, MZA, TUA10, PP2A, PTB and UBQ were ranked below, respectively (Table 2, Fig. 4). This was similar to the result generated by geNorm.

Reference gene ranking based on NormFinder

The geNorm and ΔCt methods evaluate the optimal reference genes between tissues. NormFinder was the only applet to rank the best candidate reference genes according to their minimal combined inter- and intra-

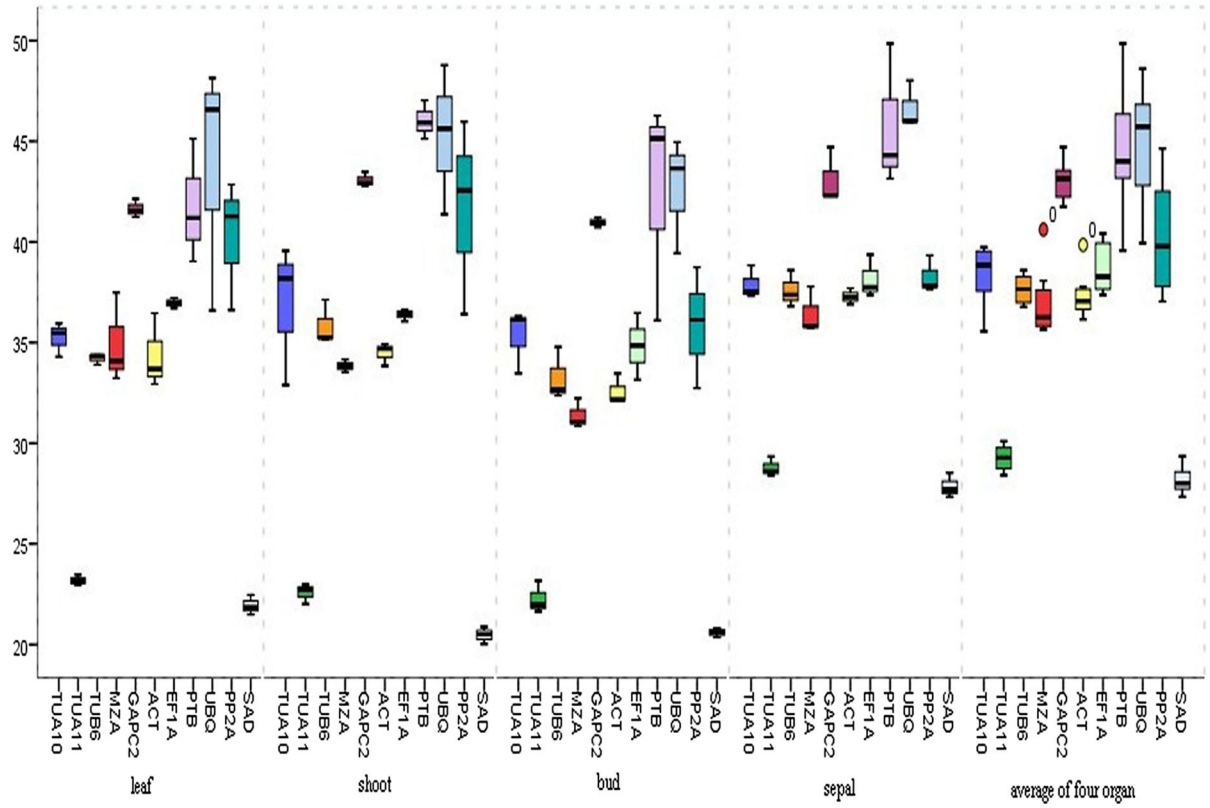

Fig. 1 *Boxplots* showing the threshold cycles (Liang et al. 2006) of 11 reference genes in four different tissues as well the average of four organs. 50 % of the values are included in the box. The median is indicated with a black line in the center of the box, the 75th and 25th percentiles are represented by *upper* and *lower line*, respectively. The *y*-axis represents the ΔCt values of three biotechnology replicates of each tissues, while the *x*-axis shows the eleven reference gene distribution in leaf, shoot, bud, sepal and the average of four organs

Table 1 Descriptive statistics of the raw Ct values of each gene among all samples

Gene abbreviation	Number	Minimum	Maximum	Mean	SD	Median
TUA10	12	30.74	36.13	34.61	1.69	35.13
TUA11	12	21.21	23.47	22.32	0.74	22.37
TUB6	12	32.37	34.80	33.54	0.96	33.54
MZA	12	30.87	37.49	32.45	1.94	31.66
GAPC2	12	38.99	42.94	40.66	1.25	40.85
ACT	12	31.53	36.47	33.01	1.28	32.74
EF1a	12	33.15	37.23	34.92	1.61	34.36
PTB	12	36.10	49.81	42.58	3.59	42.01
UBQ	12	36.60	48.14	43.26	3.71	44.30
PP2A	12	32.73	42.84	37.13	3.49	36.38
SAD	12	19.79	22.47	20.85	0.80	20.67

tissue expression variation based on normalization factor (NF) calculation. So, stability of expression was then re-analyzed using the NormFinder algorithm. The most reliable candidate reference gene were SAD, TUA11, EF1A, TUB6, GAPC2, ACT with the stability value of 0.19, 0.28, 0.38, 0.44, 0.45 and 0.98, which may serve as the top six reliable reference genes. The rank of six least stable reference candidates

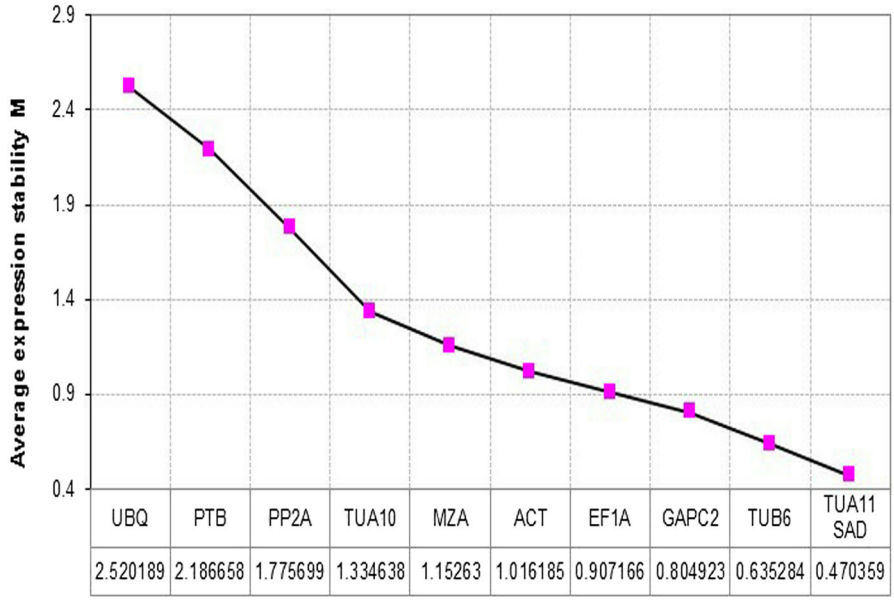

Fig. 2 The stability of reference genes as analysed by geNorm amongst all tissues. Mean expression stability (M) was computed following stepwise exclusion of the least stable gene across all tested tissues. The most stable genes are on the *right* and the least stable genes on the *left*

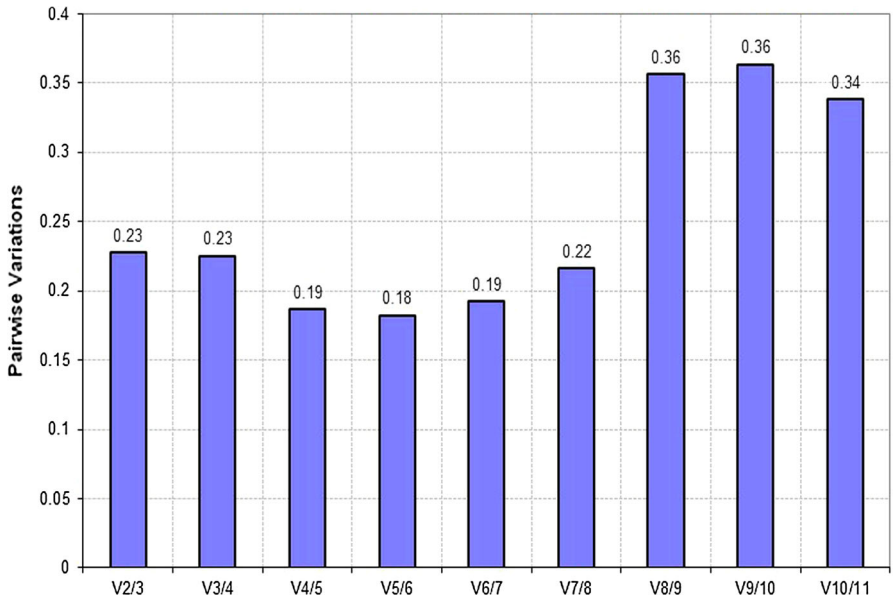

Fig. 3 The reference ranking based on pairwise variation value (V-value) as calculated by geNorm. The y-axis represents pairwise variation (Vn/Vn + 1) which was calculated between Vn (normalization factors) and Vn + 1 to determine the optimal number of reference genes

(MZA, TUA10, PP2A, PTB and UBQ) were the same as described in the method of geNorm, ΔCt and NormFinder. In addition, UBQ was consistently ranked as the least stable gene in all statistical methods.

Reference gene ranking based on BestKeeper

This applet ranked candidate reference genes based on the Ct values of each gene, standard deviation (Goetz et al. 2006) and coefficient of variation (CV). SD and CV were considered as two criteria to deduce stable reference gene candidates. Genes with SD greater than 1 are considered unstable. The results of BestKeeper analysis are shown in Table 3. In our study, TUA11, SAD, ACT and TUB6 showed an SD value lower than 1 (with the SD value of 0.62, 0.63, 0.86 and 0.87, respectively) indicating that these genes were suitable as reference genes. Furthermore, pairwise correlation and regression analysis were performed to assess the

Table 2 A summary of pair wise comparison of means and SD for reference gene using ΔCt method; the last column are the average of SD of each gene pairs

Gene		Pair1	Pair2	Pair3	Pair4	Pair5	Pair6	Pair7	Pair8	Pair9	Pair10	Avg. SD
TUA10	Mean	−12.29	−1.08	−2.17	6.04	−1.60	0.30	7.97	8.65	2.51	−13.76	
	SD	1.56	1.63	2.27	1.92	1.95	2.14	4.45	4.47	3.98	1.69	2.61
TUA11	Mean	11.21	10.12	18.33	10.69	12.59	21.11	20.94	14.80	−1.48	12.29	
	SD	0.63	1.45	1.04	1.27	1.04	4.44	3.83	3.09	0.47	1.56	1.88
TUB6	Mean	−1.09	7.27	−0.52	1.38	9.05	9.73	3.59	−12.69	1.08	−11.21	
	SD	1.49	1.23	1.30	1.12	3.26	3.72	3.46	0.80	1.63	0.63	1.86
MZA	Mean	8.21	0.57	2.47	10.14	10.82	6.57	−11.60	2.17	−10.12	1.09	
	SD	1.61	1.84	1.32	4.09	4.43	4.24	1.26	2.27	1.45	1.49	2.40
GAPC2	Mean	−7.64	−5.74	1.93	2.61	−3.53	−19.81	−6.04	−18.33	−7.27	−8.21	
	SD	1.06	1.05	3.27	3.38	3.59	0.92	1.92	1.04	1.23	1.61	1.91
ACT	Mean	1.90	9.33	10.25	4.12	−12.16	1.60	−10.69	0.52	−0.57	7.64	
	SD	1.45	3.66	3.19	3.57	1.09	1.95	1.27	1.30	1.84	1.06	2.04
EF1a	Mean	7.67	8.35	2.21	−14.07	−0.30	−12.59	−1.38	−2.47	5.74	−1.90	
	SD	3.30	3.63	2.97	1.03	2.14	1.04	1.12	1.32	1.05	1.45	1.91
PTB	Mean	0.68	−5.59	−21.73	−7.97	−21.11	−9.05	−10.14	−1.93	−9.33	−7.67	
	SD	4.18	5.10	3.65	4.45	4.44	3.26	4.09	3.27	3.66	3.30	3.94
UBQ	Mean	−6.14	−22.41	−8.65	−20.94	−9.73	−10.82	−2.61	−10.25	−8.35	−0.68	
	SD	5.65	3.74	4.47	3.83	3.72	4.43	3.38	3.19	3.63	4.18	4.02
PP2A	Mean	−16.28	−2.51	−14.80	−3.59	−6.57	3.53	−4.12	−2.21	5.59	6.14	
	SD	3.03	3.98	3.09	3.46	4.24	3.59	3.57	2.97	5.10	5.65	3.87
SAD	Mean	13.76	1.48	12.69	11.60	19.81	12.16	14.07	21.73	22.41	16.28	
	SD	1.69	0.47	0.80	1.26	0.92	1.09	1.03	3.65	3.74	3.03	1.77

inter-gene relations. A strong correlation was seen for EF1a ($r = 0.948$). However, based on BestKeeper index, EF1a had a high SD value in comparison to other candidates ($SD = 1.436$). Therefore, EF1a was excluded from candidate reference gene list. The result of BestKeeper agreed with those of ΔCt method, geNorm and NormFinder, although their rank order was slightly altered. When evaluated across all experimental samples, TUA11 was ranked as the best reference gene.

Comprehensive ranking

ReFinder integrates the currently major computational programs (geNorm, NormFinder, BestKeeper, and the comparative ΔCt method) to compare and rank the candidate reference genes. ReFinder confirmed that SAD (1.49) and TUA11 (1.41) were the most reliable reference gene among different tissues in *G. raimondii* (Table 4). The candidate reference genes were ranked (from the highest to lowest stability) by the ReFinder as TUB11>SAD>TUB6>GAPC2>EF1a>MZA>-TUA10>PP2A>PTB>UBQ. According to the results of all programs, TUB11 and SAD were considered as the most stable reference genes, and UBQ was the least stable one.

The expression profiles of commonly used house-keeping genes were tissue-specific (Table 5). In leaves, TUA11, TUB6 were considered as the most reliable reference genes, MZA was selected as the most reliable reference gene in shoot, bud and sepal, whereas EF1a and UBQ were the best stable reference candidates in shoot and sepal, respectively. The overall ranking showed that TUA11 and SAD were the most reliable reference genes in all different tissues.

Discussion

The gene expression analysis is a common way for analyzing gene function in different tissues or under

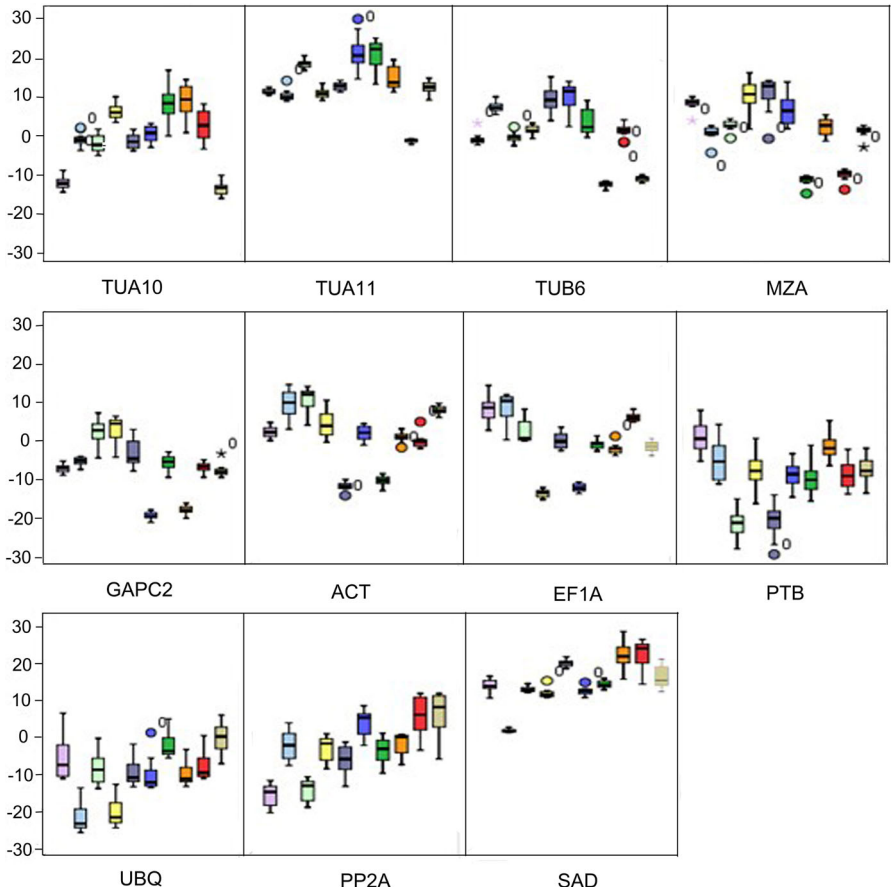

Fig. 4 A boxplot graph representing the gene expression values of 11 reference gene candidates. The ΔCt values between each gene pair/group are shown in y-axis, while the x-axis represents 11 reference candidates. Different "gene pairs" are shown as *different colors*

Table 3 Ranking of reference gene based on BestKeeper

Gene	Number	GM(CP)	AR(CP)	Min (CP)	Max (CP)	SD ± (CP)	CV (% CP)	(r)	P Value	Ranking based on	
										SD	(r)
TUA10	12	34.575	34.614	30.737	36.314	1.357	3.919	0.239	0.453	TUA11	EF1a
TUA11	12	22.314	22.325	21.210	23.469	0.615	2.757	0.918	0.001	SAD	TUA11
TUB6	12	33.525	33.538	32.369	34.797	0.875	2.609	0.787	0.002	ACT	SAD
MZA	12	32.396	32.446	30.866	37.487	1.458	4.494	0.829	0.001	TUB6	MZA
GAPC2	12	40.639	40.656	38.986	42.939	1.032	2.537	0.737	0.006	GAPC2	TUB6
ACT	12	32.992	33.013	31.531	36.468	0.864	2.617	0.487	0.109	TUA10	GAPC2
EF1a	12	34.883	34.917	33.152	37.226	1.436	4.112	0.948	0.001	EF1a	PP2A
PTB	12	42.445	42.583	36.102	49.805	2.667	6.263	0.369	0.236	MZA	ACT
PP2A	12	36.981	37.129	32.729	42.845	2.940	7.918	0.634	0.027	PTB	PTB
SAD	12	20.836	20.850	19.790	22.469	0.631	3.028	0.899	0.001	PP2A	TUA10

different experimental conditions. Several experimental methods have been used to measure gene expression including qRT-PCR. However, the choice of a reference gene is critical for reliable analyses and conclusions. Usually, housekeeping genes were commonly used as reference genes because of their conservation and abundant expression in different organisms. Nevertheless, no single housekeeping gene

Table 4 A summary of comprehensive ranking and other four-method ranking of 11 reference genes in *G. raimondii*

Comprehensive ranking (ReFinder)		geNorm		ΔCt method		BestKeeper				NormFinder	
Gene	Stability value	Gene	M-value	Genes	Mean SD	Gene	SD(C)	Gene	Coefficient of correlation(r)	Gene	Stability value
SAD	1.19	SAD	0.47	SAD	1.77	TUA11	0.62	EF1a	0.95	SAD	0.19
TUA11	1.41	TUA11	0.47	TUB6	1.86	SAD	0.63	TUA11	0.92	TUA11	0.28
TUB6	3.46	GAPC2	0.64	TUA11	1.88	ACT	0.86	SAD	0.9	EF1a	0.38
GAPC2	4.47	TUB6	0.81	GAPC2	1.91	TUB6	0.87	MZA	0.83	TUB6	0.44
EF1a	4.79	EF1a	0.91	EF1a	1.91	GAPC2	1.03	TUB6	0.79	GAPC2	0.45
ACT	5.05	ACT	1.02	ACT	2.04	TUA10	1.36	GAPC2	0.74	ACT	0.98
MZA	7.24	MZA	1.15	MZA	2.4	EF1a	1.44	PP2A	0.63	MZA	1.44
TUA10	7.44	TUA10	1.34	TUA10	2.61	MZA	1.46	ACT	0.49	TUA10	2.01
PP2A	9.24	PP2A	1.78	PP2A	3.87	PTB	2.67	PTB	0.37	PP2A	3.36
PTB	9.74	PTB	2.19	PTB	3.94	PP2A	2.94	TUA10	0.24	PTB	3.46
UBQ	11	UBQ	2.52	UBQ	4.02					UBQ	3.68

has a consistent expression in different tissues under different conditions (Artico et al. 2010). It is advisable to identify stable reference genes for target gene expression normalization. In past years, numerical reports had been published to evaluate reliable reference genes in different plants under various experiment conditions (Cassan-Wang et al. 2012; Chaouachi et al. 2013; Scholtz and Visser 2013; Serra et al. 2012; Zhu et al. 2013; Zhu et al. 2012). It has been stated that the stability of a reference gene must be validated with each experimental setup (Hruz et al. 2011). Several genes including ACTIN, GAPDH, TUA, and UBQ are used as single control genes in more than 90 % of the published expression studies (de Jonge et al. 2007). However, recent studies have indicated that these common reference genes are not always stably expressed when tested in other species or in a wider range of experimental treatments (Artico et al. 2010; Chen et al. 2011; Gutierrez et al. 2008a; Zhu et al. 2012). Here, we aimed to validate a collection of reference genes in every tissue qRT-PCR studies in G. *raimondii*. Our results indicated that

Table 5 The comprehensive stability of different tissues analysis by ReFinder

Ranking	Leaf		Shoot		Bud		Sepal	
	Gene	Geomean of rank value	Gene	Geomean of rank value	Gene	Geomean of rank value	Gene	Geomean of rank value
1	TUA11	1.41	EF1a	1.57	MZA	1.73	UBQ	2.59
2	TUB6	1.57	MZA	2.21	GAPC2	1.97	MZA	2.83
3	EF1a	2.45	GAPC2	2.59	SAD	2.63	PP2A	3.13
4	SAD	4.16	TUA11	3.50	TUA11	2.99	EF1a	3.35
5	GAPC2	4.47	SAD	3.56	TUB6	4.95	TUB6	4.40
6	TUA10	6.00	PTB	6.09	EF1a	5.24	ACT	5.20
7	ACT	7.00	ACT	6.24	ACT	6.09	SAD	5.45
8	MZA	8.00	TUB6	8.00	TUA10	8.00	TUA11	5.47
9	PTB	9.49	TUA10	9.24	UBQ	9.24	GAPC2	7.33
10	PP2A	9.49	UBQ	9.74	PP2A	9.74	TUA10	7.95
11	UBQ	11.00	PP2A	11.00	PTB	11.00	PTB	11.00

UBQ was the worst candidate for normalization in *G. raimondii*.

To our knowledge, this is the first systematic study to validate a set of candidate reference genes for qRT-PCR in *G. raimondii*. In this study, we selected a series of candidate reference genes for which sequence information was available in the *G. raimondii* genome database. After determination of primer amplification efficiencies, 12 candidates were selected for evaluation of their normalization potential along a group of samples from different tissues. Our results suggested that TUA11 and SAD were the best reference genes according to the average expression stability (M) or stability values acquired by ΔCt method, geNorm, BestKeeper, NormFinder and ReFinder. When the outcomes of five algorithms were compared, only slight differences were observed in ranking. However, for different tissues, TUA11, TUB6 and SAD were identified as the reliable reference genes in leaves. EF1a, MZA and GAPC2 were the less variable reference genes in shoots. MZA, GAPC2, SAD and TUA11 were the best reference genes in buds, while UBQ and MZA were considered as the most stable reference genes in sepals. These results showed that combined reference genes in overall level might solve this problem. In addition, using multiple reference genes might improve, yet is not necessary, for reliable gene expression analysis in *G. raimondii.*

Acknowledgments This project is partially supported by National Natural Science Foundation of China (Grant Number: 31170263). We appreciate Faten Taki for her proofread of this manuscript.

Supplementary Table 1 Properties of twelve reference gene candidates.

References

Andersen CL, Jensen JL et al (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 64:5245–5250

Artico S, Nardeli SM et al (2010) Identification and evaluation of new reference genes in Gossypium hirsutum for accurate normalization of real-time quantitative RT-PCR data. BMC Plant Biol 10:49

Brunner AM, Yakovlev IA et al (2004) Validating internal controls for quantitative plant gene expression studies. BMC Plant Biol 4:14

Bustin SA (2005) Real-time, fluorescence-based quantitative PCR: a snapshot of current procedures and preferences. Expert Rev Mol Diagn 5:493–498

Bustin SA, Benes V et al (2005) Quantitative real-time RT-PCR–a perspective. J Mol Endocrin 34:597–601

Cassan-Wang H, Soler M et al (2012) Reference genes for high-throughput quantitative reverse transcription-PCR analysis of gene expression in organs and tissues of Eucalyptus grown in various environmental conditions. Plant Cell Physiol 53:2101–2116

Chaouachi M, Alaya A et al (2013) Development of real-time PCR method for the detection and the quantification of a new endogenous reference gene in sugar beet "Beta vulgaris L.": GMO application. Plant Cell Rep 32:117–128

Chen L, Zhong HY et al (2011) Validation of reference genes for RT-qPCR studies of gene expression in banana fruit under different experimental conditions. Planta 234:377–390

Czechowski T, Stitt M et al (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol 139:5–17

de Jonge HJ, Fehrmann RS et al (2007) Evidence based selection of housekeeping genes. PLoS One 2:e898

Ding J, Jia J et al (2004) Validation of a rice specific gene, sucrose phosphate synthase, used as the endogenous reference gene for qualitative and real-time quantitative PCR detection of transgenes. J Agri Food Chem 52:3372–3377

Exposito-Rodriguez M, Borges AA et al (2008) Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biol 8:131

Faccioli P, Ciceri GP et al (2007) A combined strategy of "in silico" transcriptome analysis and web search engine optimization allows an agile identification of reference genes suitable for normalization in gene expression studies. Plant Mol Biol 63:679–688

Goetz M, Vivian-Smith A et al (2006) AUXIN RESPONSE FACTOR8 is a negative regulator of fruit initiation in Arabidopsis. Plant Cell 18:1873–1886

Goidin D, Mamessier A et al (2001) Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. Anal Biochem 295:17–21

Gutierrez L, Mauriat M et al (2008a) The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotech J 6:609–618

Gutierrez L, Mauriat M et al (2008b) Towards a systematic validation of references in real-time RT-PCR. Plant Cell 20:1734–1735

Hruz T, Wyss M et al (2011) RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC Genom 12:156

Jain M, Nijhawan A et al (2006) Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. Biochem Biophy Res Commun 345:646–651

Jian B, Liu B et al (2008) Validation of internal control for gene expression study in soybean by quantitative real-time PCR. BMC Mol Biol 9:59

Kim BR, Nam HY et al (2003) Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice. Biotech Lett 25:1869–1872

Liang M, Haroldsen V et al (2006) Expression of a putative laccase gene, ZmLAC1, in maize primary roots under stress. Plant, Cell Environ 29:746–753

Ma J, Guo TL et al (2015) Expression profiles of miRNAs in *Gossypium raimondii*. J Zhejiang University Sci B. doi:10.1631/jzus.B1400277

Mahoney DJ, Carey K et al (2004) Real-time RT-PCR analysis of housekeeping genes in human skeletal muscle following acute exercise. Physiol Genomics 18:226–231

Nicot N, Hausman JF et al (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot 56:2907–2914

Paolacci AR, Tanzarella OA et al (2009) Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. BMC Mol Biol 10:11

Paterson AH, Wendel JF et al (2012) Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492:423–427

Pfaffl MW, Tichopad A et al (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. Biotech Lett 26:509–515

Reid KE, Olsson N et al (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol 6:27

Remans T, Smeets K et al (2008) Normalisation of real-time RT-PCR gene expression measurements in Arabidopsis thaliana exposed to increased metal concentrations. Planta 227:1343–1349

Schmittgen TD, Zakrajsek BA (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophy Method 46:69–81

Scholtz JJ, Visser B (2013) Reference gene selection for qPCR gene expression analysis of rust-infected wheat. Physiol Mol Plant P 81:22–25

Serra IA, Lauritano C et al (2012) Reference genes assessment for the seagrass Posidonia oceanica in different salinity, pH and light conditions. Mar Biol 159(6):1269–1282

Silver N, Best S et al (2006) Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol Biol 7:33

Tong Z, Gao Z et al (2009) Selection of reliable reference genes for gene expression studies in peach using real-time PCR. BMC Mol Biol 10:71

Vandesompele J, De Preter K et al (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:RESEARCH0034

Wang M, Wang Q et al (2013) Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.). Gene 530:44–50

Xie FL, Jones DC et al (2015) Small RNA sequencing identifies miRNA roles in ovule and fiber development. Plant Biotechnol J. doi: 10.1111/pbi.12296 (in press)

Xie F, Xiao P et al (2012) miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol Biol 80:75–84

Zhang B, Pan X (2009) Expression of microRNAs in cotton. Mol Biotech 42:269–274

Zhang Y, Chen D et al (2012) Selection of reliable reference genes in Caenorhabditis elegans for analysis of nanotoxicity. PLoS One 7:e31849

Zhu X, Li X et al (2012) Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. PLoS One 7:e44405

Zhu JF, Zhang LF et al (2013) Reference gene selection for quantitative real-time PCR normalization in *Caragana intermedia* under different abiotic stress conditions. PLoS One 8:e53196