

# Selection of Reliable Reference Genes for Gene Expression Studies Using Real-Time PCR in Tung Tree during Seed Development

Xiaojiao Han<sup>1,2</sup>, Mengzhu Lu<sup>1,3</sup>, Yicun Chen<sup>1,2</sup>, Zhiyong Zhan<sup>2</sup>, Qinjin Cui<sup>2</sup>, Yangdong Wang<sup>1,2\*</sup>

**1** State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing, People's Republic of China, **2** Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang, People's Republic of China, **3** Research Institute of Forestry, Chinese Academy of Forestry, Beijing, People's Republic of China

## Abstract

Quantitative real-time PCR (RT-qPCR) has become an accurate and widely used technique to analyze expression levels of selected genes. It is very necessary to select appropriate reference genes for gene expression normalization. In the present study, we assessed the expression stability of 11 reference genes including eight traditional housekeeping genes and three novel genes in different tissues/organs and developing seeds from four cultivars of tung tree. All 11 reference genes showed a wide range of Ct values in all samples, indicating that they differently expressed. Three softwares – geNorm, NormFinder and BestKeeper – were used to determine the stability of these references except for *ALB* (2S albumin), which presented a little divergence. The results from the three softwares showed that *ACT7* (Actin7a), *UBQ* (Ubiquitin), *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) and *EF1 $\alpha$*  (elongation factor 1- $\alpha$ ) were the most stable reference genes across all of the tested tung samples and tung developing seeds, while *ALB* (2S albumin) was unsuitable as internal controls. *ACT7*, *EF1 $\beta$*  (elongation factor1-beta), *GAPDH* and *TEF1* (transcription elongation factor 1) were the top four choices for different tissues/organs whereas *LCR69* did not favor normalization of RT-qPCR in these tissues/organs. Meanwhile, the expression profiles of *FAD2* and *FADX* were realized using stable reference genes. The relative quantification of the *FAD2* and *FADX* genes varied according to the internal controls and the number of internal controls. The results further proved the importance of the choice of reference genes in the tung tree. These stable reference genes will be employed in normalization and quantification of transcript levels in future expression studies of tung genes.

**Citation:** Han X, Lu M, Chen Y, Zhan Z, Cui Q, et al. (2012) Selection of Reliable Reference Genes for Gene Expression Studies Using Real-Time PCR in Tung Tree during Seed Development. PLoS ONE 7(8): e43084. doi:10.1371/journal.pone.0043084

**Editor:** Christian Schönbach, Kyushu Institute of Technology, Japan

**Received:** May 11, 2012; **Accepted:** July 16, 2012; **Published:** August 17, 2012

**Copyright:** © 2012 Han et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The work presented here was supported by the State Forestry Administration Foundation of China (2009-4-23). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: wyd11111@126.com

## Introduction

Tung tree (*Vernicia fordii* Hemsl.), a subtropical round-crowned deciduous tree, belongs to a species of the genus *Vernicia* in the spurge (Euphorbiaceae) family. Tung oil extracted from seeds is considered to be one of the high-value industrial oils [1], used widely in production of cloth, shoes, waterproofing masonry, clothing, paper, and biodiesel [1,2]. With 80%  $\alpha$ -eleostearic acid of tung oil, a high degree of unsaturation, tung oil is regarded as a conjugated drying oil [3,4]. Following the development and maturation of tung tree seeds, the content of fatty acids gradually increases. The peak periods of fatty acid synthesis are during the middle of August and the middle of September in China [5]. The construction of cDNA library of tung seeds and the release of expressed sequence tag (EST) databases have greatly promoted the study of genes involved in fatty acids synthesis such as delta-12 fatty acid desaturase (*FAD2*), delta 12 fatty acid conjugase (*FADX*), diacylglycerol acyltransferase 1 (*DGAT1*) and diacylglycerol acyltransferase 2 (*DGAT2*) [6,7]. Therefore, the understanding of expression patterns of some key genes will help elucidate the mechanism involved in fatty acids synthesis of tung seeds.

Methods to detect gene expression level include Northern blot, semi-quantitative PCR (semi-PCR), RNase protection analysis

(RPA), gene chips and quantitative real-time PCR (RT-qPCR). RT-qPCR has become a very powerful method to detect and quantify gene transcription levels due to its high sensitivity, specificity, reproducibility and accuracy [8–10]. Besides, in many situations, it is the only method for detecting mRNA levels of low copy number target genes of interest [11]. However, the factors of RNA stability, quality, quantity, retrotranscription efficiencies and PCR reaction can affect the reliability of testing result for RT-qPCR [8,12]. Thus, normalization for transcript levels of test genes is crucial to minimize technical variations [12].

One of the methods used to normalize RT-qPCR data is to select appropriate reference genes for controlling the experimental possible errors generated during the detection process [12]. Ideal reference genes are expected to be stable at an expression level across various experimental conditions such as plant developmental stages, tissue types and external stimuli [13]. The most commonly used reference genes include  $\beta$ -actin (*ACT*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), 18S ribosomal RNA (*18S rRNA*), 25S ribosomal RNA (*25S rRNA*), polyubiquitin (*UBQ*), ubiquitin conjugating enzyme (*UBC*), translation elongation factor (*TEF*), cyclophilin (*CYC*), elongation factor 1- $\alpha$  (*EF1 $\alpha$* ) and tubulin (*TUB*) etc. [8,14,15]. However, recent studies have shown

**Table 1.** Candidate reference genes, primers and different parameters derived from RT-qPCR analysis.

Gene Symbol	Gene name	GenBank accession number	Primer sequence(5'→3') (forward/reverse)	Tm (°C)	Amplicon Length (bp)	Amplification efficiency (%)	R <sup>2</sup>
<i>ACT7</i>	Actin7a	JQ680035	CGATGAAGCACAGT CCAAAAGGTTGAGA GGAGCCTCAGTG	82.36	170	100.03	0.9984
<i>ALB</i>	2S albumin	JQ680046	TAAGGCAACAAATGG CTTCACATCGAAA CCCTGAAGACG	87.06	166	97.91	0.9997
<i>EF1<math>\alpha</math></i>	Elongation factor 1-alpha	JQ680036	GCCTGGTATGGTTGT GACCT GGATCATCC TTGGAGTTGGA	84.01	180	97.2	0.9995
<i>EF1<math>\beta</math></i>	Elongation factor 1-beta	JQ680037	CGAATCAGGCTCAA GTCTC CACCTTTG CCACCAATTCTT	84.72	218	98.14	0.9992
<i>GAPDH</i>	Glyceraldehyde-3-phosphate dehydrogenase	JQ680038	CTGCTAAGGCTGTT GGGAAG TCCCTCT GACTCCTCCTTGA	83.55	168	97.41	0.9999
<i>LCR69</i>	Low molecular weight cysteine-rich 69	JQ680039	CCTCCTTCTTCTG CTGCTTG GTAAC CCTCGGCAGTCTCCT	84.85	155	96.15	1
<i>SAMDC</i>	S-adenosylmethionine decarboxylase	JQ680043	CCTGGAGCTCAGTC GTATCC CCAAACC AGTCATGCACATC	83.16	208	93.53	0.9987
<i>TCTP</i>	Translationally controlled tumor protein	JQ680040	GAAGGGGCAGATG AAGATGAGAGAG CAGGAACCTGGTTGC	82.95	214	95.58	0.9996
<i>TEF1</i>	Transcription elongation factor 1	JQ680042	GTTGTCCTTCTG CAACCAT AACGTT GTAAACCCGCTCAC	81.48	179	95.01	0.9997
<i>UBC</i>	Ubiquitin-conjugating enzyme E2	JQ820248	CCATTTCCAAGGT GTTGCTT GGCAG CACTGTTAACCCATT	83.31	165	93.40	0.9993
<i>UBQ</i>	Ubiquitin	JQ680041	CCGTGGTGGCTGTT AAGTTT AAGGCCA TTTCAACATCCTG	80.71	193	94.47	0.9993
<i>FAD2</i>	delta-12 fatty acid desaturase	AF525534	AGCATCCGCTGGTT CTCTAA GCAAGAA CACCAGCATCAGA	83.25	212	99.87	0.9994
<i>FADX</i>	delta 12 fatty acid conjugase	AF525535	GGAAGCAGAAGC GTGAAACAGGTGG TGCAATGGAGTAG	83.01	170	94.7	0.9991

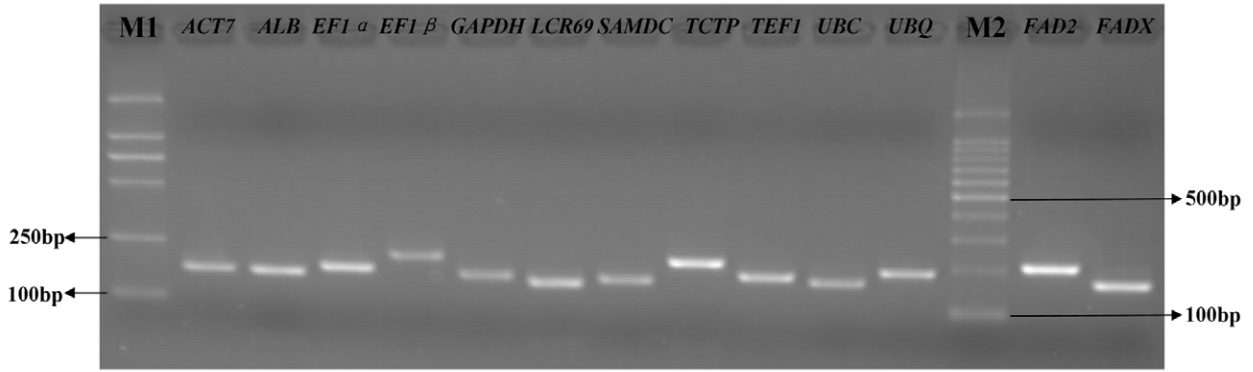
doi:10.1371/journal.pone.0043084.t001

that some of these references might not be stably expressed under different experimental conditions [16]. For example, *UBC16* expression in leaves of the lily plants was quite stable under various treatments, whereas its expression was rather variable in the roots [17]. In Chinese cabbage, *EF1 $\alpha$*  is the most suitable reference genes among different tissues, but *GAPDH* is the best choice for experiment under conditions of drought stress and downy mildew infection [18]. Therefore, it is necessary to screen and identify novel reference genes from expressed sequence tags databases (EST), transcriptome data, Microarray analysis and cDNA libraries [19]. *Expressed1*, *SNAD* and *CACS* from transcriptome sequence data in buckwheat, for instance, are revealed as the most stable in different plant structures [20].

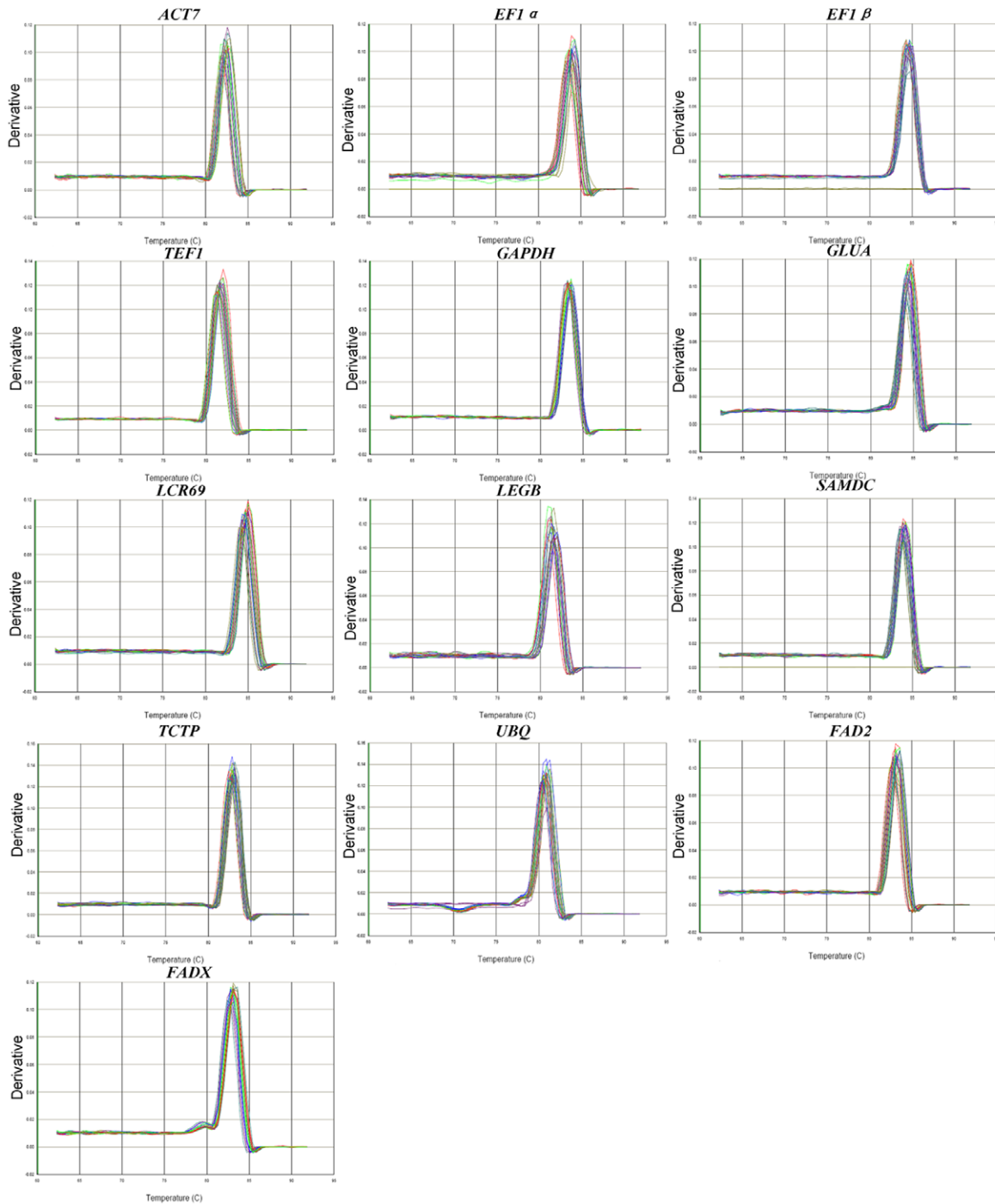
As far as is known, selection and identification of stable reference genes refer to many plant species and cultivated varieties [19]. In the tung tree, albumin (*ALB*) and ubiquitin ligase (*UBC*) have been used as reference genes in developing seeds [6,21]; however, the stability of both genes has not yet been accessed.

Thus, identification of reliable reference genes for RT-qPCR will benefit further studies on the tung seeds development and different tissue/organs at the transcription level. In the present study, we aimed to identify potential reference genes suitable for transcript normalization in tung developing seeds and different tissue/organs. The expression profiles of 11 reference genes including *ACT7* (actin7a), *ALB* (2S albumin), *EF1 $\alpha$*  (elongation factor1-alpha), *EF1 $\beta$*  (elongation factor1-beta), *TEF1* (transcription elongation factor 1), *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase), *LCR69* (low molecular weight cysteine-rich 69), *SAMDC* (S-adenosylmethionine decarboxylase), *TCTP* (translationally controlled tumor protein), *UBC* (ubiquitin-conjugating enzyme E2) and *UBQ* (ubiquitin), were studied in seven different tissue/organs and six different development stages of seeds collected from four cultivars of tung tree, and the expression stability of these genes was subsequently evaluated using geNorm [22], Bestkeeper [23] and NormFinder [24]. Furthermore, the expression patterns of two target genes *FAD2* and *FADX* were

a



b



**Figure 1. Gene specificity and amplicon size.** (a) Agarose gel (2%) electrophoresis showing amplification of a specific PCR product of the expected size for each gene. (b) Melting curves of 11 reference genes and 2 target genes showing single peaks. M1 and M2 represent 2000 bp and 100 bp DNA ladder marker respectively.  
doi:10.1371/journal.pone.0043084.g001

investigated using the selected references, which may be helpful to reveal their roles in fatty acids synthesis.

## Results

### Verification of amplicons, primer specificity, Ct data collection and gene-specific PCR amplification efficiency

A total of 11 reference genes from the tung tree kernel uncut cDNA library were selected as candidates for normalization of gene expression measures. Gene name, accession number, gene description, primer sequences, amplicon length, amplification efficiency,  $T_m$  values and correlation coefficients were listed in Table 1. The melting temperatures ( $T_m$ ) of all PCR products ranged from 80.71°C for *UBQ* to 84.85°C for *LCR69*. Amplification efficiency ( $E$ ) of PCR reactions varied from 93.40% for *UBC* to 100.03% for *ACT7*, and correlation coefficients ( $R^2$ ) ranged between 0.9984 for *ACT7* and 1 for *LCR69*, respectively (Table 1). The amplifications were confirmed by the presence of a single band of expected size for each primer pairs in 2% agarose gel electrophoresis (Figure 1a) and by the single peak melting curves of the PCR products (Figure 1b). No primer dimers or other PCR products were generated from non-specific amplification (Figure 1a), and no products were detected in negative controls.

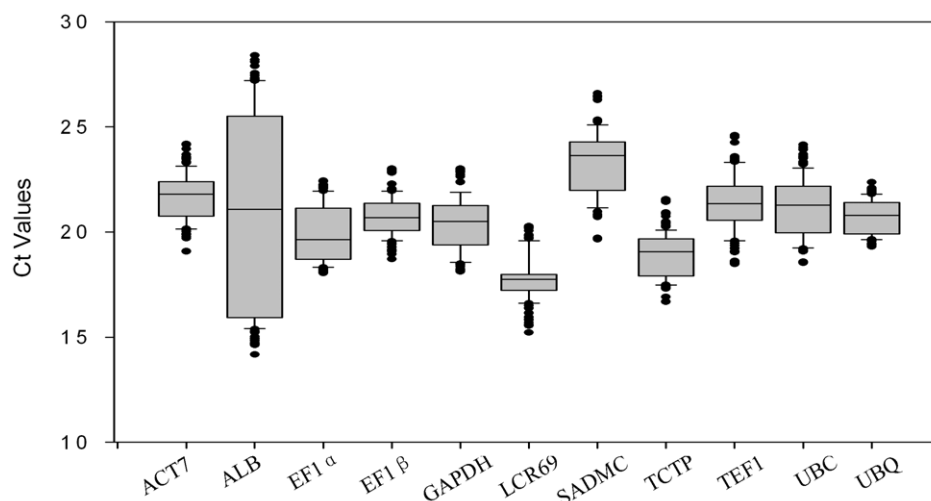
The cycle threshold (Ct) values were obtained from each reaction with 11 primer pairs. To reveal the differences in transcript levels between 11 reference genes, it is necessary to assess the Ct range and calculate the coefficient of variance for each gene across all samples. As expected, the average Ct values varied between the different genes ranging from 16 to 27 cycles (Figure 2). *LCR69* with narrow variance was the most abundant reference transcript, reaching threshold fluorescence only 17 to 19 amplification cycles, while *SADMC* was the least abundant transcript with Ct of 24. The expression stability was also detected by calculating coefficient of variance (CV) of Ct values. A seed storage protein *ALB* had large variances in their transcript levels,

and the CV values were more than 11 for all samples, indicating that the gene was unstable. On the other hand, *EF1 $\beta$* , *GAPDH*, *TEF* and *UBQ* had narrow variances in their transcript expressions.

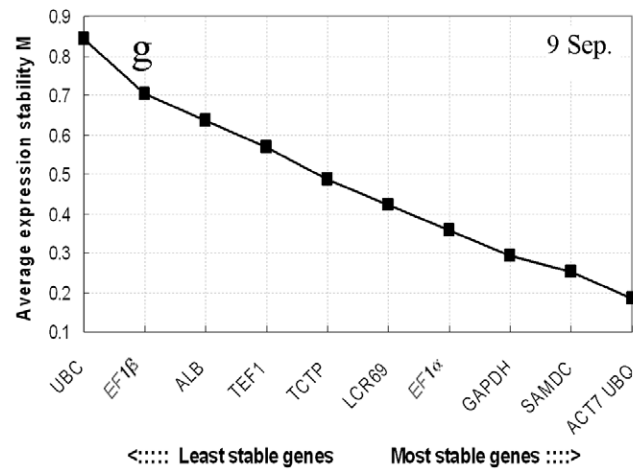
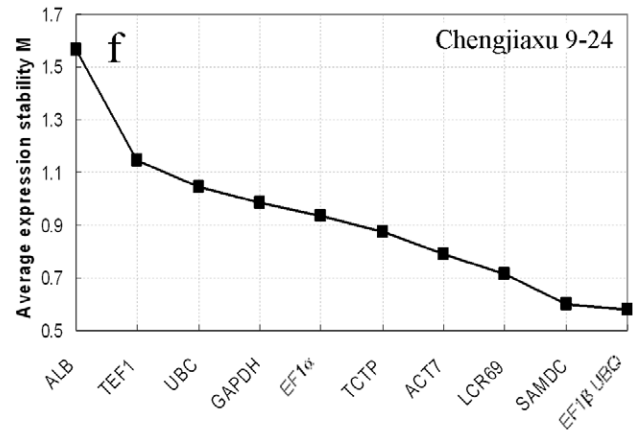
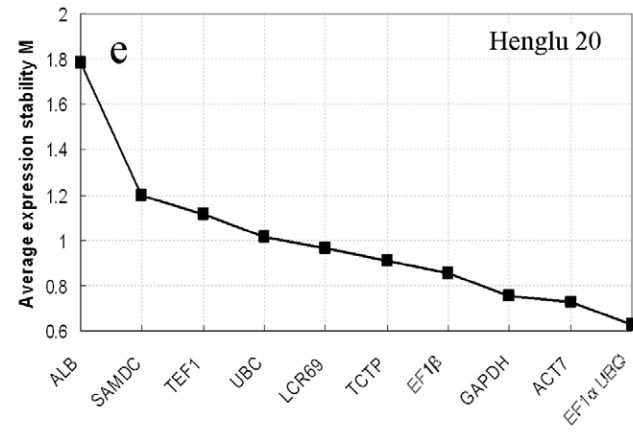
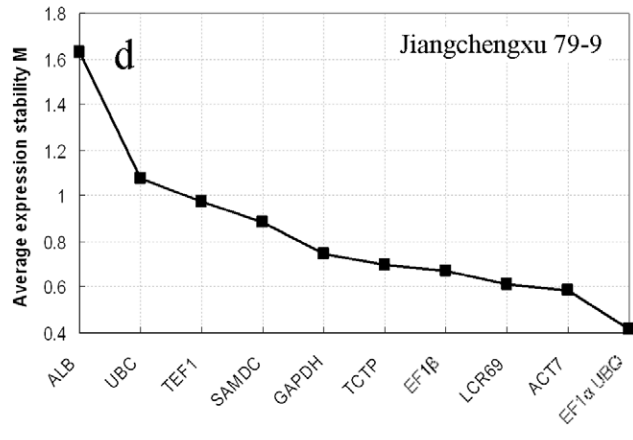
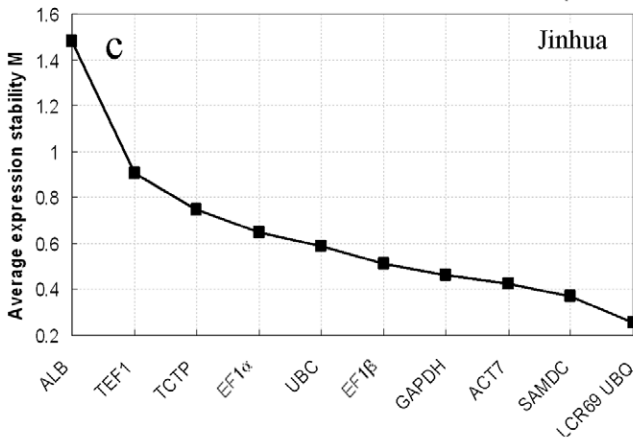
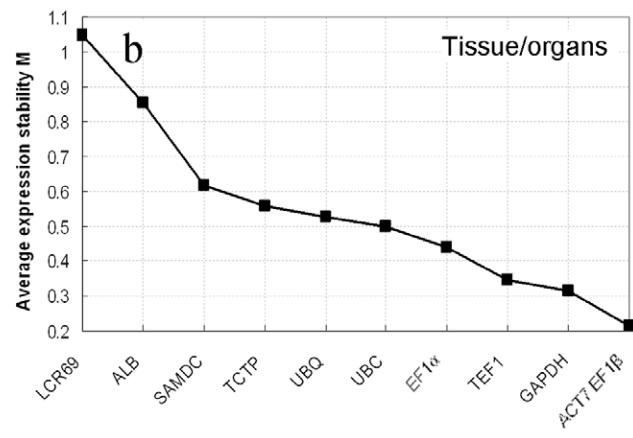
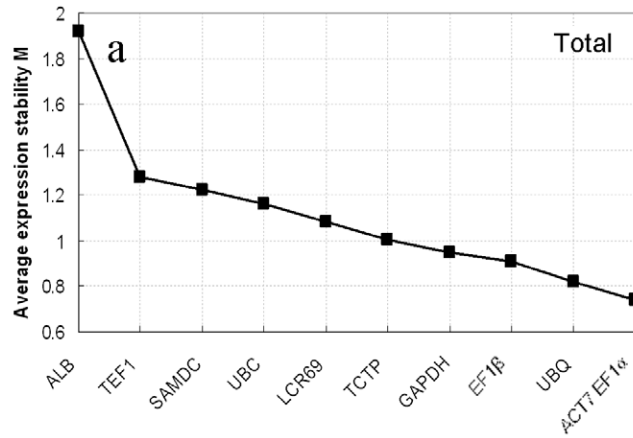
### Expression stability of candidate reference genes

Three different software programmes were used to evaluate the expression stability of the candidate reference genes: geNorm [22], Bestkeeper [23] and NormFinder [24]. Ct data were collected across all samples. Ct values were used directly for stability calculations for BestKeeper analysis, and these data were transformed to relative quantities using the delta-Ct method for geNorm and NormFinder analysis.

**a) geNorm analysis.** geNorm was used to rank the reference genes by calculating gene expression stability value  $M$  based on the average pairwise expression ratio. The most stable reference gene has the lowest  $M$  value, while the least stable one presents the highest  $M$  value. The program recommends using  $M$  value below the threshold of 1.5 to identify reference genes with stable expression. In our analysis of four cultivars, all genes except *ALB* had  $M < 1.5$  as a criterion to consider the tested genes as rather stable (Figure 3). When all the results from all 31 samples of *V. fordii* were combined, *EF1 $\alpha$*  and *ACT7* had the highest expression stability (the lowest  $M$  value), whereas a seed storage protein *ALB* was revealed less stability and other eight genes were placed at the intermediate positions between the two extremums (Figure 3a). Stability rank of the 11 tested reference genes in the seven tissue/organs confirmed that all genes had  $M$  values below the threshold of 1.5, and *EF1 $\beta$*  and *ACT7* had the highest expression stability (Figure 3b). For the two cultivars “Jiangchengxu 79-9” and “Henglu 20”, *EF1 $\alpha$*  and *UBQ* were the most stably expressed genes (Figure 3d and Figure 3e). In contrast, for the cultivar “Jinhua”, *LCR69* and *UBQ* were the most stably expressed genes (Figure 3c). Besides, the geNorm analysis also indicated that *EF1 $\beta$*  and *UBQ* were the most stably expressed reference genes in the



**Figure 2. RT-qPCR Ct values for reference genes.** Expression data displayed as Ct values for each reference gene in all tung samples. A line across the box is depicted as the median. The box indicates the 25th and 75th percentiles, whisker caps represent the maximum and minimum values, dots represent outliers.  
doi:10.1371/journal.pone.0043084.g002



<::::: Least stable genes      Most stable genes ::::::>

**Figure 3. Average expression stability values (M) calculated by geNorm.** (a) all 31 samples, (b) different tissue/organs, (c) the cultivar “Jinhua”, (d) the cultivar “Jiangchengxu 79-9”, (e) the cultivar “Henglu 20”, (f) the cultivar “Chengjiayu 9–24”, (g) four cultivars from 9 September. Lower average expression stability (M value) indicates more stable expression. doi:10.1371/journal.pone.0043084.g003

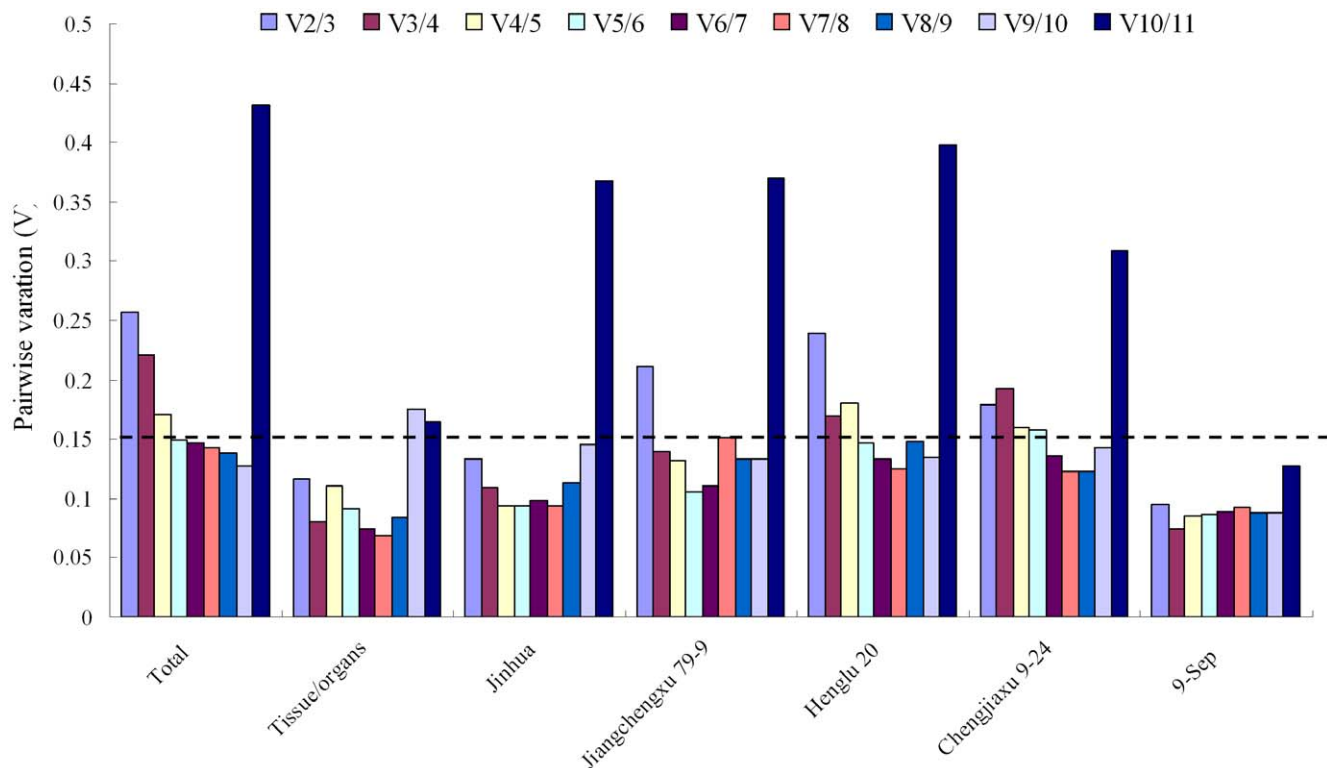
cultivar “Chengjiayu 9–24” (Figure 3f). Therefore, there was a little change of the reference ranks in four different cultivars. Furthermore, reference gene expression stability was analyzed in the samples of four different cultivars from 9 September (Figure 3g). *ACT7* and *UBQ* had the lowest M value with the highest expression stability. Overall, all of the tested reference genes except *ALB* showed relatively high stability with low M values of less than 1.3 (Figure 3).

To evaluate the optimal number of reference genes required for accurate normalization, the pairwise variation ( $V_n/V_{n+1}$ ) was calculated using geNorm between consecutively ranked normalization factors. Generally, 0.15 was used as a cutoff value to determine the optimal number of reference genes [22]. In our data sets, the paired variable coefficient in all samples indicated that the inclusion of the fourth reference gene hardly contributed to the variation of the normalization factor, whereas two stable reference genes *ACT7* and *EF1 $\beta$*  or *ACT7* and *UBQ* ( $V_2/3 < 0.15$ ) in different tissue/organs or different cultivars from 9 September would be sufficient for normalizing gene expression (Figure 4). When all 31 samples were taken together to determine the number of reference genes, the pairwise variation of  $V_2/3$  was higher than 0.15 (0.257), as were  $V_3/4$  (0.221) and  $V_4/5$  (0.171). The  $V_5/6$  value was 0.148, indicating that at least five reference genes should be included for gene expression studies in all the samples of *V. fordii*.

**b) NormFinder analysis.** NormFinder is another Excel application, which ranks the candidate genes according to stability

index M based on the average pairwise variation of a gene compared to the rest of the studied genes [22]. The more stably expressed genes exhibited the lower average expression stability values (M values). The stability value of each gene was calculated by NormFinder as shown in Table 2. This analysis method identified that *ACT7*, *EF1 $\beta$* , *UBQ* and *GAPDH* were the most appropriate for use as a reference gene in all samples. For the different tissues/organs and the cultivar “Jiangchengxu 79-9”, *GAPDH*, *TEF1*, *UBQ* and *ACT7* had the most stable expression. In the three cultivars “Jinhua”, “Henglu 20”, “Chengjiayu 9–24”, *LCR69*, *UBQ*, *GAPDH* and *ACT7* had the most stable expression and were the ideal reference genes. In four cultivars collected from 9 September, *ACT7* and *UBQ* were the stable reference genes. For all tested samples, *ALB* was the least stable reference genes. Due to unstable expression according to the results of geNorm and NormFinder analysis, the candidate *ALB* was discarded from subsequent analysis.

**c) Bestkeeper analysis.** Bestkeeper, an Excel-based software tool, evaluates the most stably expressed genes based on the coefficient of correlation ( $r$ ) to the BestKeeper index by calculating the Ct set standard deviation (SD) and coefficient of variance (CV). BestKeeper analysis revealed that the best correlations were obtained for *ACT7* (0.960), *EF1 $\alpha$*  (0.883), *GAPDH* (0.892) and *UBQ* (0.829) with  $P$  value of 0.001 for all samples (Table 3). In the four cultivars, *ACT7*, *UBQ*, *EF1 $\alpha$*  and *GAPDH* showed the highest



**Figure 4. Pairwise variation (V) calculated by geNorm to determine the optimal number of reference genes.** The average pairwise variations  $V_n/V_{n+1}$  was analyzed between the normalization factors  $NF_n$  and  $NF_{n+1}$  to indicate the optimal number of reference genes required for RT-qPCR data normalization in different samples. doi:10.1371/journal.pone.0043084.g004



**Table 2.** Ranking of candidate reference genes in order of their expression stability as calculated by NormFinder.

Rank	Total	Tissue/organs	Jinhua	Jiangchengxu 79-9	Henglu 20	Chengjiayu 9-24	9 September
1	<i>ACT7</i> (0.121)	<i>GAPDH</i> (0.023)	<i>LCR69</i> (0.089)	<i>ACT7</i> (0.054)	<i>UBQ</i> (0.066)	<i>UBQ</i> (0.178)	<i>ACT7</i> (0.064)
2	<i>EF1β</i> (0.341)	<i>TEF1</i> (0.265)	<i>UBQ</i> (0.191)	<i>GAPDH</i> (0.321)	<i>GAPDH</i> (0.192)	<i>EF1β</i> (0.202)	<i>UBQ</i> (0.064)
3	<i>UBQ</i> (0.403)	<i>UBQ</i> (0.277)	<i>ACT7</i> (0.206)	<i>UBQ</i> (0.341)	<i>ACT7</i> (0.299)	<i>LCR69</i> (0.340)	<i>SAMDC</i> (0.175)
4	<i>GAPDH</i> (0.438)	<i>ACT7</i> (0.292)	<i>SAMDC</i> (0.222)	<i>TEF1</i> (0.456)	<i>TEF1</i> (0.491)	<i>ACT</i> (0.447)	<i>GAPDH</i> (0.212)
5	<i>TEF1</i> (0.459)	<i>UBC</i> (0.297)	<i>GAPDH</i> (0.229)	<i>SAMDC</i> (0.464)	<i>EF1β</i> (0.493)	<i>SAMDC</i> (0.474)	<i>EF1α</i> (0.326)
6	<i>EF1α</i> (0.593)	<i>TCTP</i> (0.299)	<i>EF1β</i> (0.328)	<i>EF1α</i> (0.504)	<i>EF1α</i> (0.520)	<i>GAPDH</i> (0.519)	<i>LCR69</i> (0.362)
7	<i>LCR69</i> (0.697)	<i>EF1α</i> (0.335)	<i>UBC</i> (0.552)	<i>LCR69</i> (0.552)	<i>TCTP</i> (0.776)	<i>TEF1</i> (0.711)	<i>TEF1</i> (0.419)
8	<i>TCTP</i> (0.712)	<i>EF1β</i> (0.347)	<i>EF1α</i> (0.641)	<i>EF1β</i> (0.617)	<i>LCR69</i> (0.813)	<i>TCT</i> (0.789)	<i>TCTP</i> (0.460)
9	<i>SAMDC</i> (0.909)	<i>SAMDC</i> (0.627)	<i>TEF1</i> (0.670)	<i>TCTP</i> (0.847)	<i>SAMDC</i> (0.862)	<i>EF1α</i> (0.841)	<i>ALB</i> (0.530)
10	<i>UBC</i> (0.979)	<i>ALB</i> (1.155)	<i>TCTP</i> (0.975)	<i>UBC</i> (0.923)	<i>UBC</i> (0.962)	<i>UBC</i> (0.935)	<i>EF1β</i> (0.651)
11	<i>ALB</i> (3.336)	<i>LCR69</i> (1.248)	<i>ALB</i> (2.804)	<i>ALB</i> (2.802)	<i>ALB</i> (3.030)	<i>ALB</i> (2.351)	<i>UBC</i> (0.972)

doi:10.1371/journal.pone.0043084.t002

correlations. The results of BestKeeper analysis showed little differences from those obtained from geNorm and Normfinder.

### Reference gene validation

The use of different reference genes to calculate relative expression data could have a significant influence on the final normalized results. To detect the effect of reference gene on the outcome of a practical experiment, the relative expression patterns for two functional genes *FAD2* and *FADX* were evaluated using different reference genes in seven different tissue/organs and six different developmental stages of tung seeds from the cultivar “Jiangchengxu 79-9” (Figure 5). The most stable references *ACT7*, *TEF1* and *EF1β* in the different tissue/organs selected by geNorm and Bestkeeper were used as internal controls. *FAD2* was expressed in all tested tissues/organs, with a higher level in leaves and petals (Figure 5a1), while *FADX* expression was restricted to leaves (Figure 5b1). However, *FAD2* and *FADX* were expressed at a lower level when using the least stable reference *LCR69* as internal control. When stable references *UBQ*, *ACT7*, *EF1α*, and the combination of the three references were used as internal controls respectively, the expression patterns of *FAD2* and *FADX* showed a low expression in earlier stages of tung seeds development (16 July and 26 July). Both *FAD2* and *FADX* exhibited a similar expression pattern with an increase from 11 August to 9 September (Figure 5a2 and Figure 5b2). When the least stable reference gene *ALB* was used for normalization, the

two target genes were expressed in lower levels in tung developing seeds, and showed a reverse result compared to the stable references for normalization. Thus, the use of unsuitable references can lead to over- or underestimation of relative transcript abundance. These results reinforce the importance of validating reference genes prior to experimental applications.

### Discussion

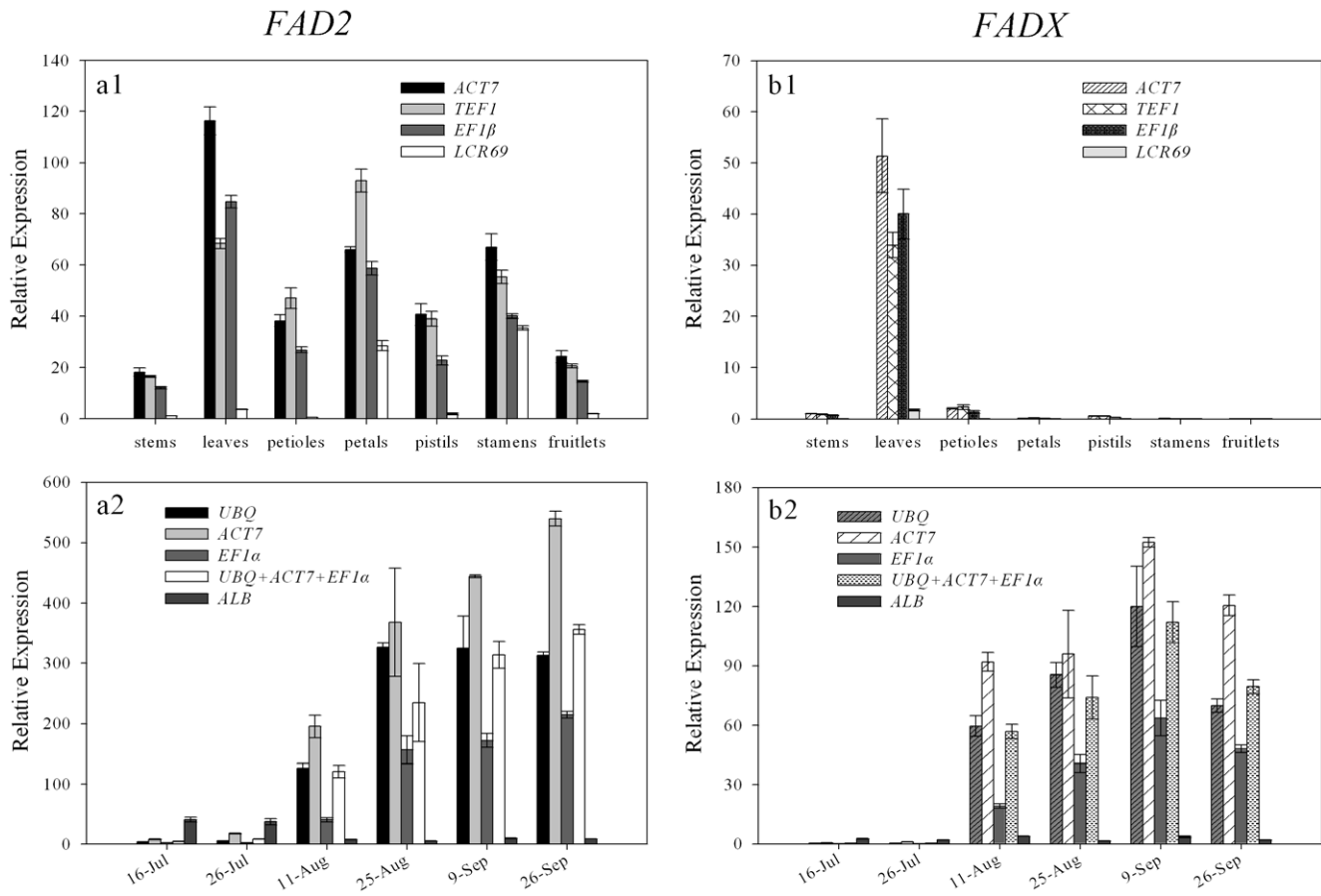
At present, quantitative real-time PCR has significantly improved the detection and quantification of expression profiles of target genes due to its high throughput, sensitivity, specificity, accuracy and broad quantification range [9,10]. It is very necessary to screen appropriate internal reference genes for gene expression normalization during target genes expression analyzed. A stable expressed reference gene should produce constant Ct values under different experimental conditions such as plant developmental stages, tissue types and external stimuli [13]. Here, the stability of expression of three novel and eight traditional reference genes was evaluated in different tissue/organs and developing seeds from four different cultivars of tung tree. *EF1β*, *GAPDH*, *TEF*, *UBQ* and *LCR69* with narrow Ct values were stable in the developing seeds. A seed storage protein *ALB* had large variances in their transcript levels, indicating that the reference gene was unstable.

Among recent studies on qRT-PCR, commonly used traditional reference genes, e.g. *ACT*, *GAPDH*, *18S rRNA*, *25S rRNA*, *UBQ*,

**Table 3.** Statistics results by BestKeeper software for ten selected genes based on Ct values.

coeff. of corr. [r] (p-value)	<i>ACT7</i>	<i>EF1α</i>	<i>EF1β</i>	<i>GAPDH</i>	<i>LCR69</i>	<i>SAMDC</i>	<i>TCTP</i>	<i>TEF1</i>	<i>UBC</i>	<i>UBQ</i>
Total	0.960 (0.001)	0.883 (0.001)	0.774 (0.001)	0.892 (0.001)	0.435 (0.015)	0.725 (0.001)	0.626 (0.001)	0.577 (0.001)	0.641 (0.001)	0.829 (0.001)
Tissue/organs	0.980 (0.001)	0.861 (0.013)	0.980 (0.001)	0.980 (0.001)	0.153 (0.741)	0.840 (0.018)	0.896 (0.006)	0.985 (0.001)	0.884 (0.008)	0.847 (0.016)
Jinhua	0.903 (0.014)	0.786 (0.064)	0.743 (0.091)	0.816 (0.048)	0.842 (0.036)	0.897 (0.015)	0.761 (0.079)	0.283 (0.587)	0.761 (0.079)	0.093 (0.003)
Jiangchengxu 79-9	0.955 (0.003)	0.964 (0.002)	0.718 (0.108)	0.871 (0.024)	0.848 (0.033)	0.674 (0.143)	0.560 (0.248)	0.643 (0.168)	0.459 (0.361)	0.904 (0.013)
Henglu 20	0.988 (0.001)	0.858 (0.029)	0.849 (0.033)	0.928 (0.008)	0.538 (0.27)	0.647 (0.164)	0.711 (0.113)	0.823 (0.044)	0.63 (0.181)	0.917 (0.010)
Chengjiayu 9-24	0.982 (0.001)	0.942 (0.005)	0.975 (0.001)	0.917 (0.010)	0.947 (0.004)	0.872 (0.024)	0.807 (0.052)	0.657 (0.157)	0.614 (0.194)	0.849 (0.033)

doi:10.1371/journal.pone.0043084.t003



**Figure 5. Expression levels of *FAD2* and *FADX* in different tissues/organs and seeds development of the cultivar “Jiangchengxu 79-9”.** (a1 and a2) Expression levels of *FAD2* in different tissues/organs and seeds development, (b1 and b2) Expression levels of *FADX* in different tissues/organs and seeds development. Genes were normalized to individual and/or combined reference genes. Error bars show mean standard error calculated from two biological replicates. doi:10.1371/journal.pone.0043084.g005

*UBC*, *TEF*, *CYC*, *EF1α*, *TUB* were considered to be stable and suitable in various tissues [14,25], since these genes are present in all cell types and necessary for basic cell survival. Nevertheless, numerous researches have already shown that the expression of these traditional genes might also be variational [26–28]. Thus, normalization with multiple reference genes is becoming popular and standard in plant research [23,24]. The present study demonstrates the importance of screening reference genes. geNorm analysis is used to determine the optimal number of stable reference genes for accurate normalization [29]. Generally, 0.15 was used as a cutoff value to confirm the optimal number of reference genes [22]. However, this is not an absolute rule and depends on the dataset tested. A higher V value is considered in other reports [30–32]. In the present study, when all samples were taken together to determine the number of reference genes, the pairwise variation of V2/3, V3/4 and V4/5 were higher than 0.15 (Figure 3). The V5/6 value was 0.148, thus the result shows that five genes are included to support gene expression studies. This indicates that the combination of multiple references is necessary to normalize gene expression for all the samples of *V. fordii*.

When gene expression stability in all samples was analyzed by geNorm, the most stable genes were *ACT7* and *EF1α*, followed by *UBQ* (Figure 3). The genes encoding actin, elongation factor 1-alpha and ubiquitin are often considered as reliable reference genes under various experimental conditions. For example, *ACT11*

and *EF1α* exhibited a stable expression pattern across different tissues in the water lily [17]. *ACT7* is one of stable reference genes for developing embryos in *Brassica napus* [33]. Besides, *ACT* also had the highest expression stability across leaf and root tissues in chicory [34] and tomato [35]. In addition, *EF1α* is a stable reference gene in dandelion [36], grape berry development [25], different developmental stages and under mild Cd stress conditions in poplar leaves [37], and also in cucumber [38]. *UBQ* exhibited the most stable expression across all samples of *Arabidopsis* [26]. However, *UBQ10* was the most variable reference gene, and should be avoided as an internal control in rice, soybean and the development of grape berry [25,29,39]. An ubiquitin tag is reported to mark particular proteins for proteolytic elimination, but it can also have nonproteolytic functions [40], thus its wide range of function lead to the variable expression of ubiquitin in different plants. According to the results of geNorm and NormFinder, a seed storage protein *ALB* was ranked in the last position in all samples and developing seeds from the four cultivars. In different tissue/organs, *ALB* was the least abundant transcript with Ct values of 26–28, indicating that the expression level of the reference gene was very low. When using a stable reference *UBQ* as internal control, *ALB* was not detected in the developing seeds in July. The transcript level rapidly increased from 11 August, and slightly declined on 9 September (data not shown). The results showed that *ALB* was not suitable for reference



gene in tung tree. In animals, albumin is present in all nucleated cell types and is necessary for basic cell survival and considered to be stable in various tissues [23]. This indicates that there is a great difference in the expression of *ALB* between plants and animals, thus the reference gene is suitable for animals but not for plants.

The results from geNorm and NormFinder analysis showed some differences, especially in the top ranked genes. However, the output of both programs very consistently listed the same genes showing unstable expression patterns. This little divergence probably reflects differences in the statistical algorithms. It was also reported that there were discrepancies between NormFinder and geNorm in other studies. For example, in *citrus*, the *FBOX/SAND* pair was selected as the least variable among all reference genes by geNorm, but the most stable reference gene according to NormFinder was *UPL7* [41].

To validate the suitability of the reference genes we identified in this study, the expression profiles of *FAD2* and *FADX* were assessed in different tissue/organs and tung developing seeds of the cultivar “Jiangchengxu 79-9”. In tung oil biosynthesis, *FAD2* desaturates oleic acid (18:1 $\Delta^9$ ) to produce linoleic acid (18: 2 $\Delta^9, 12$ ), then *FADX* converts linoleic acid to eleostearic acid (18:3 $\Delta^9, 11, 13$ ) [42,43]. The data showed that the use of the most stable reference genes *UBQ*, *ACT7*, *EF1 $\alpha$*  or the combination of stable references resulted in the trend consistency of the relative transcript abundance of *FAD2* and *FADX*. However, the relative transcript abundance presented a reduction when the most variable reference gene *LCR69* or *ALB* used as an internal control (Figure 5). These results suggest that the incorrect use of reference genes may introduce bias in the analysis and lead to the misinterpretation of data.

## Conclusions

In summary, 11 reference genes were evaluated in different tissue/organs and different development stages of tung seeds. We also concluded traditional housekeeping genes that outperformed novel reference genes. The results showed *ACT7*, *UBQ*, *GAPDH* and *EF1 $\alpha$*  were suggested as good candidate genes used as reference genes for normalization in gene expression studies. In this constitution, we identify and validate optimal reference genes for RT-qPCR normalization with consideration of different tissues/organs and seed development stages.

## Methods

### Plant materials

Tung fruits of four different cultivars, including “Jinhua,” “Jiangchengxu 79-9,” “Henglu 20” and “Chengjiayu 9–24” were collected from the National Gene Pool (constructed in 1979) of Tung Tree in Dongfanghong Forest Farm, Zhejiang Province, China. Seven different tissue/organs, including stems, leaves, petioles, petals, pistils, stamens and fruitlets (30 days after flowering) were collected from the cultivar “Jiangchengxu 79-9”. No specific permits were required for the farm to select samples. The farm is not privately-owned in any way and the field studies did not involve protected species. Samples of the six different developmental stages of tung fruits during the increasing periods of fatty acids were taken in 2011: 16 July, 26 July, 11 August, 25 August, 9 September and 26 September. Seeds removed from fruits were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until needed for RNA extraction. All samples were collected in two replicates.

### RNA extraction and first strand cDNA synthesis

Frozen seeds were hand-shelled, and kernels were ground to a fine powder in liquid nitrogen with a pestle and mortar. About 100 mg of this powder was used for RNA extraction. Total RNA was isolated using the RN38 EASYspin plus Plant RNA kit (Aidlab Biotech, Beijing, China). Purified RNA was quantified with NanoDrop2000 spectrophotometer (Thermo, Wilmington, USA), and loaded on a denaturing 1.0% (p/v) agarose gel to check concentration and integrity. Only RNA samples with 260/280 wavelength ratio between 1.8 and 2.1 and 260/230 wavelength ratio greater than 2.0 were used for cDNA synthesis. cDNA synthesis was performed with 3  $\mu\text{g}$  total RNA using the superscript III first strand synthesis system followed by the RNase H step (Invitrogen, Carlsbad, USA), according to the protocol of the manufacturer in a total volume of 20  $\mu\text{l}$ . cDNAs were diluted 1:30 with nuclease-free water for RT-qPCR.

### Primer design and PCR conditions

The 11 candidate genes including eight traditional housekeeping genes (*ACT7*, *EF1 $\alpha$* , *EF1 $\beta$* , *TEF1*, *GAPDH*, *SAMDC*, *UBC* and *UBQ*) and three novel reference genes (*ALB*, *LCR69* and *TCTP*) were selected from the tung tree kernel uncut cDNA library (Table 1). Gene sequences were deposited in the GenBank (accession numbers are listed in Table 1). All reference genes were named based on similarity to known nucleotide sequences using BLAST with a score value higher than 100 and identity ranging from 81% to 94%. Primer pairs were designed using Primer3 (<http://frodo.wi.mit.edu/primer3/>) with the following parameters:  $T_m$  around  $60^{\circ}\text{C}$  and product size range 155–218 base pairs, primer sequences with a length of 19 to 21 nucleotides with an optimum at 20 nucleotides, and a GC content of 45% to 55%. To check all primers specificity, real-time PCR was performed on cDNA and products were analyzed by electrophoresis on 2% agarose gel and ethidium bromide staining.

Real-time PCR reactions were performed in 96-well plates with a 7500 Real Time PCR System (Applied Biosystems, CA, USA) and a SYBR<sup>®</sup> Premix Ex Taq<sup>™</sup> Kit (TaKaRa, Tokyo, Japan) as described by Phillips et al. [44]. PCR reactions were prepared in 20  $\mu\text{l}$  volumes containing 2  $\mu\text{l}$  of 30-fold diluted synthesized cDNA, 10  $\mu\text{l}$  2 $\times$  SYBR<sup>®</sup> Premix Ex Taq<sup>™</sup>, 0.4  $\mu\text{l}$  10  $\mu\text{M}$  forward primer, 0.4  $\mu\text{l}$  10  $\mu\text{M}$  reverse primer, 0.4  $\mu\text{l}$  50 $\times$  RO $\times$  reference dye and 6.8  $\mu\text{l}$  sterile distilled water. Negative PCR control with no templates was performed for each primer pair. The cycling conditions were recommended by the manufacturer (30 s at  $95^{\circ}\text{C}$ , 40 cycles of  $95^{\circ}\text{C}$  for 5 s, and  $60^{\circ}\text{C}$  for 34 s). Specificity of amplicons was verified by melting curve analysis (60 to  $95^{\circ}\text{C}$ ) after 40 PCR cycles. The final threshold cycle (Ct) values were the mean of eight values including two biological replicates for each treatment and four technical replicates.

### Analysis of gene expression stability

Standard curves were constructed to calculate the gene-specific PCR efficiency from 10-fold series dilution of the mixed cDNA template for each primer pair. The correlation coefficients ( $R^2$ ) and slope values can be obtained from the standard curve, and the corresponding PCR amplification efficiencies ( $E$ ) were calculated according to the equation  $E = (10^{-1/\text{slope}} - 1) \times 100$  [45].

Gene expression stability was evaluated by applying three different statistical approaches: geNorm (ver. 3.5) [22], Bestkeeper (ver. 1.0) [23] and NormFinder (ver. 0.953) [24]. Real-time RT-qPCR data was exported into an Excel datasheet (Microsoft Excel 2003) and Ct values were converted according to the requirements of the software. Each of these approaches generates a measure of

reference gene stability, which can be used to rank the order of stability for reference genes.

## Acknowledgments

The work presented here was supported by the State Forestry Administration Foundation of China (2009-4-23).

## References

- Brown K, Keeler W (2005) The history of tung oil. *Wildland Weeds* 9: 4–24.
- Shang Q, Jiang W, Lu H, Liang B (2010) Properties of tung oil biodiesel and its blends with 0# diesel. *Bioresour Technol* 101: 826–828.
- Sonntag N (1979) Composition and characteristics of individual fats and oils. *Bailey's industrial oil and fat products* 1: 289–477.
- Thanamongkollit N, Soucek MD (2008) Modification of tung oil for bio-based coating [The Graduate Faculty of the University of Akron In Partial Fulfillment of the Requirements for the Degree Master of Science]. Akron (Ohio): Department of Science, University of Akron.
- Fang JX, He F (1998) Tung oil trees in China. China Forestry Publishing House, Beijing: 104–127.
- Chen Y, Zhou G, Wang Y, Xu L (2010) F-BOX and oleosin: additional target genes for future metabolic engineering in tung trees? *Ind Crop Prod* 32: 684–686.
- Shockey JM, Gidda SK, Chapital DC, Kuan J-C, Dhanoa PK, et al. (2006) Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *Plant Cell* 18: 2294–2313.
- Bustin SA (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* 29: 23–39.
- Bustin SA, Benes V, Nolan T, Pfaffl MW (2005) Quantitative real-time RT-PCR—a perspective. *J Mol Endocrinol* 34: 597–601.
- Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. *Nat Protoc* 1: 1559–1582.
- Huggett J, Dheda K, Bustin S, Zumla A (2005) Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity* 6: 279–284.
- Udvardi MK, Czechowski T, Scheible WR (2008) Eleven golden rules of quantitative RT-PCR. *Plant Cell* 20: 1736–1737.
- Banda M, Bommineni A, Thomas RA, Luckinbill LS, Tucker JD (2008) Evaluation and validation of housekeeping genes in response to ionizing radiation and chemical exposure for normalizing RNA expression in real-time PCR. *Mutat Res* 649: 126–134.
- Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, et al. (2004) Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* 37: 112–119.
- Kim BR, Nam HY, Kim SU, Kim SI, Chang YJ (2003) Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice. *Biotechnol Lett* 25: 1869–1872.
- Gutierrez L, Mauriat M, Guenin S, Pelloux J, Lefebvre JF, et al. (2008) 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 Biotechnol J* 6: 609–618.
- Luo H, Chen S, Wan H, Chen F, Gu C, et al. (2010) Candidate reference genes for gene expression studies in water lily. *Anal Biochem* 404: 100–102.
- Qi J, Yu S, Zhang F, Shen X, Zhao X, et al. (2010) Reference gene selection for real-time quantitative polymerase chain reaction of mRNA transcript levels in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Plant Mol Biol Rep* 28: 597–604.
- Kumar V, Sharma R, Trivedi PC, Vyas GK, Khandelwal V (2011) Traditional and novel references towards systematic normalization of qRT-PCR data in plants. *Aust J Crop Sci* 5: 1455–1468.
- Demidenko NV, Logacheva MD, Penin AA (2011) Selection and validation of reference genes for quantitative real-time PCR in buckwheat (*Fagopyrum esculentum*) based on transcriptome sequence data. *PLoS One* 6: e19434.
- Pastor S (2011) Determining biological roles of four unique *Vernicia fordii* acyl-CoA Binding Proteins. University of New Orleans Theses and Dissertations Paper 1337.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034.1–0034.11.
- Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. *Biotechnol Lett* 26: 509–515.
- Andersen CL, Jensen JL, Orntoft 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: 5245–5250.
- Reid KE, Olsson N, Schlosser J, Peng F, Lund ST (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: 27–37.
- 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 Physiol* 139: 5–17.
- Nicot N, Hausman JF, Hoffmann L, Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot* 2907–2914.
- Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, et al. (2008) Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. *Planta* 227: 1343–1349.
- Jian B, Liu B, Bi Y, Hou W, Wu C, et al. (2008) Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol Biol* 9: 59–72.
- Kuijk EW, Du Puy L, Van Tol HTA, Haagsman HP, Colenbrander B, et al. (2007) Validation of reference genes for quantitative RT-PCR studies in porcine oocytes and preimplantation embryos. *BMC Dev Biol* 7: 58.
- De Ketelaere A, Goossens K, Peelman L, Burvenich C (2006) Technical note: validation of internal control genes for gene expression analysis in bovine polymorphonuclear leukocytes. *J Dairy Sci* 9: 4066–4069.
- Fernandez P, Di Rienzo JA, Moschen S, Dosio GAA, Aguirrezabal LAN, et al. (2011) Comparison of predictive methods and biological validation for qPCR reference genes in sunflower leaf senescence transcript analysis. *Plant Cell Rep* 30: 63–74.
- Chen X, Truksa M, Shah S, Weselake RJ (2010) A survey of quantitative real-time polymerase chain reaction internal reference genes for expression studies in *Brassica napus*. *Anal Biochem* 405: 138–140.
- Maroufi A, Bockstaele Ev, Loose Md, van Bockstaele E, de Loose M (2010) Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. *BMC Mol Biol* 11: 15–26.
- Lovdal T, Lillo C (2009) Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. *Anal Biochem* 387: 238–242.
- Martin RC, Hollenbeck VG, Dombrowski JE (2008) Evaluation of Reference Genes for Quantitative RT-PCR in *Lolium perenne*. *Crop Sci* 48: 1881–1887.
- Basa B, Solti A, Sarvari E, Tamas L (2009) Housekeeping gene selection in poplar plants under Cd-stress: comparative study for real-time PCR normalisation. *Funct Plant Biol* 36: 1079–1087.
- Wan H, Zhao Z, Qian C, Sui Y, Malik AA, et al. (2009) Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber. *Anal Biochem* 399: 257–261.
- Jain M, Nijhawan A, Tyagi AK, Khurana JP (2006) Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem Biophys Res Commun* 345: 646–651.
- Hochstrasser M (2000) Evolution and function of ubiquitin-like protein-conjugation systems. *Nat Cell Biol* 2: E153–157.
- Mafra V, Kubo KS, Alves-Ferreira M, Ribeiro-Alves M, Stuart RM, et al. (2012) Reference genes for accurate transcript normalization in citrus genotypes under different experimental conditions. *PLoS One* 7: e31263.
- Dyer JM, Chapital DC, Kuan JC, Mullen RT, Turner C, et al. (2002) Molecular analysis of a bifunctional fatty acid conjugase/desaturase from tung. Implications for the evolution of plant fatty acid diversity. *Plant Physiol* 130: 2027–2038.
- Dyer JM, Mullen RT (2008) Engineering plant oils as high-value industrial feedstocks for biorefining: the need for underpinning cell biology research. *Physiol Plantarum* 132: 11–22.
- Phillips MA, D'Auria JC, Luck K, Gershenzon J (2009) Evaluation of candidate reference genes for real-time quantitative PCR of plant samples using purified cDNA as template. *Plant Mol Biol Rep* 27: 407–416.
- Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, et al. (2004) Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun* 313: 856–862.

## Author Contributions

Conceived and designed the experiments: XH YW. Performed the experiments: XH ML. Analyzed the data: XH ZZ QC. Contributed reagents/materials/analysis tools: XH ML YC ZZ QC. Wrote the paper: XH ML YC.