

Validation of reference genes for real-time quantitative PCR normalization in soybean developmental and germinating seeds

Qing Li · Cheng-Ming Fan · Xiao-Mei Zhang ·
Yong-Fu Fu

Received: 21 February 2012 / Revised: 8 April 2012 / Accepted: 2 May 2012 / Published online: 16 May 2012
© Springer-Verlag 2012

Abstract Most of traditional reference genes chosen for real-time quantitative PCR normalization were assumed to be ubiquitously and constitutively expressed in vegetative tissues. However, seeds show distinct transcriptomes compared with the vegetative tissues. Therefore, there is a need for re-validation of reference genes in samples of seed development and germination, especially for soybean seeds. In this study, we aimed at identifying reference genes suitable for the quantification of gene expression level in soybean seeds. In order to identify the best reference genes for soybean seeds, 18 putative reference genes were tested with various methods in different seed samples. We combined the outputs of both geNorm and NormFinder to assess the expression stability of these genes. The reference genes identified as optimums for seed development were *TUA5* and *UKN2*, whereas for seed germination they were novel reference genes Glyma05g37470 and Glyma08g28550. Furthermore, for total seed samples it was necessary to combine four genes of Glyma05g37470, Glyma08g28550, Glyma18g04130 and *CYP* for normalization.

Key message We identified several reference genes that stably expressed in soybean seed developmental and germinating processes.

Keywords Soybean · Seed development · Seed germination · RT-qPCR · Normalization · Reference gene

Introduction

Seeds are essential for flowering plant reproduction because they protect, nourish, and contain the developmental embryo that represents the next sporophytic generation. In addition, seeds harvested from crops are the main food source for human beings and animals (Le et al. 2007). In the higher plants, seeds involve two important processes in the life cycle. One is seed development, and the other is seed germination. During seed development, seeds are formed and then filled with reserve food. Seed development is completed when storage compounds have accumulated, water content has decreased, ABA levels have increased, and desiccation tolerance and primary dormancy are established (Holdsworth et al. 2008). Upon imbibition massive transcriptome changes occur, which are regulated by ambient temperature, light conditions, and plant hormones. The hormones abscisic acid and gibberellins play a major role in regulating seed germination (Weitbrecht et al. 2011). Due to their special characteristics, research on seeds, especially their molecular regulation mechanism, is of significance for both fundamental and applied issues.

Now quantifying gene expression levels is increasingly important to unravel complex molecular regulation networks. Real-time quantitative PCR (RT-qPCR) is a widely used method compared with conventional gene expression

Communicated by H. Judelson.

Electronic supplementary material The online version of this article (doi:10.1007/s00299-012-1282-4) contains supplementary material, which is available to authorized users.

Q. Li · C.-M. Fan · X.-M. Zhang · Y.-F. Fu (✉)
MOA Key Lab of Soybean Biology (Beijing), National Key
Facility of Crop Gene Resource and Genetic Improvement,
Institute of Crop Sciences, Chinese Academy of Agricultural
Sciences, Beijing 100081, People's Republic of China
e-mail: fuyf@caas.net.cn

Q. Li
e-mail: liqing1986102@163.com

quantification methods, such as northern-blot analysis, ribonuclease protection assay, and semi-quantitative RT-PCR, due to its sensitivity, facility and dynamic quantification range. This approach also requires the same kind of normalization as the traditional methods (Vandesompele et al. 2002). Thus, to choose a proper and stably expressed reference genes for accurate normalization is critical for the reliable RT-qPCR results.

Housekeeping genes (HKGs), such as *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE*, *ACTIN*, *TUBULIN*, and *UBIQUITIN*, which were thought to be stably expressed, were often used as reference genes for normalization. However, the expression level of HKGs was found varied considerably as the case may be for different samples, developmental processes or experimental conditions (Graeber et al. 2011; Kulcheski et al. 2010; Lee et al. 2002; Suzuki et al. 2000). Using the unvalidated reference genes will have significant impact on the RT-qPCR results, even lead to wrong conclusions (Gutierrez et al. 2008; Hu et al. 2009; Jian et al. 2008). Therefore, it is prerequisite to identify and validate the best reference genes for the corresponding experimental systems. To date, several relevant softwares for identifying the best reference genes, such as geNorm (Vandesompele et al. 2002), NormFinder (Andersen et al. 2004), and Bestkeeper (Pfaffl et al. 2004) have been developed and widely used.

Most of traditional reference genes chosen for RT-qPCR normalization were assumed to be ubiquitously and constitutively expressed in vegetative tissues (Czechowski et al. 2005). However, seeds show distinct transcriptomes compared with the vegetative tissues (Cadman et al. 2006; Czechowski et al. 2005; Le et al. 2010; Schmid et al. 2005). Therefore, there is a need for re-validation of reference gene in samples of seed development and germination. Recently, Graeber et al. (2011) identified new candidate genes during seed germination based on transcriptome data of *Lepidium sativum* germinating seeds (Graeber et al. 2011). Based on stability analysis of homologous genes, they concluded that reference gene expression stability is higher for a given developmental process between distinct

species than for distinct developmental processes within a given single species. In another recent study, Dekkers et al. (2012) validated the expression stability of the putative reference genes in *Arabidopsis* and tomato seeds by exploiting seed microarray data (Dekkers et al. 2012). This study further supported the hypothesis that putative orthologs of reference genes derived from one species could serve the same purpose in other species (Czechowski et al. 2005). However, there is no validated reference gene for soybean seed development and germination up to now.

In this study, we aimed at identifying reference genes suitable for the quantification of gene expression level in seeds of soybean, which is an important seed crop and excellent model system to study seed development and germination. We evaluated 18 putative reference genes in total in the RT-qPCR experiment using different samples during soybean seed development and germination, and identified a number of genes that stably expressed (with a cut-off value of $M \leq 0.5$) in different seed developmental processes. Furthermore, some novel reference genes were mined and showed greater expression stability than the traditional soybean references. As a result, the use of combinations of these reference genes should provide a more reliable means of normalizing gene expression in soybean seed experiments.

Materials and methods-

Plant materials

The soybean cultivar Kennong18 (KN18) was grown in a growth chamber under short day conditions (8-h light/16-h dark) at a temperature 25–28 °C. Seed development progress was divided into 12 stages (S1–S12) based on seed length (Fig. 1), and each stage as a sample (Table S1). The seed lengths that correspond to approximated days after flowering (DAF) as specified in the Table S1. It should be noted that the range of DAF, which indicates early seed development, was slightly different between A81-356022

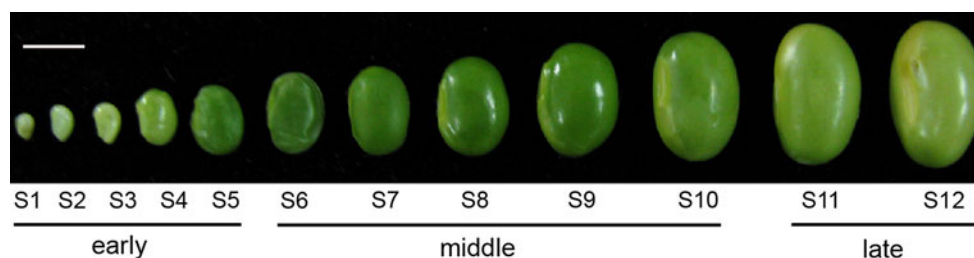


Fig. 1 The different stages of seed development progress. Soybean seed development was divided into 12 stages (S1–S12) based on seed length at 1-mm interval. Scale bar 5 mm. Early, middle, and late

developmental stages according to the previous report (Le et al. 2007) were shown below each stage

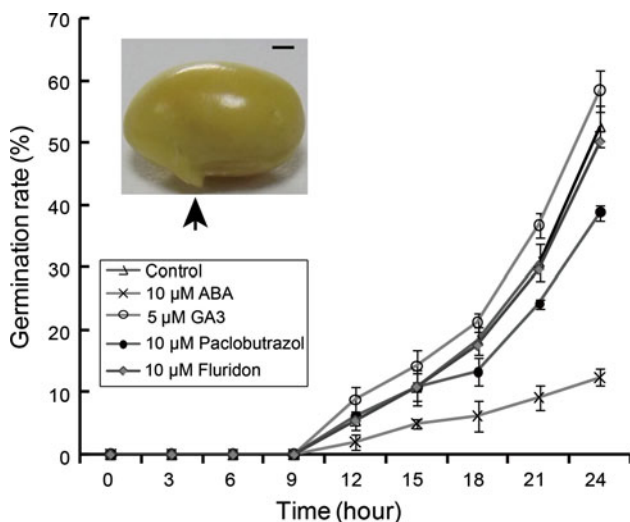


Fig. 2 Germination phenotype of the soybean seeds with different treatments. A seed was scored as germination if the primary root reached 1 mm in length (indicated by the arrow; scale bar 2 mm), and germination rates were calculated every 3 h in 24 h based on the mean \pm SD of three replicates

(the cultivar used for microarray assay) and KN18 (the cultivar used in this study). This difference may be caused due to different cultivars, different temperatures, different light periods and so on. For the germination assay, 80 hand-selected KN18 seeds of uniform size were placed in Petri dishes (12 cm in diameter) containing two layers of filter paper with 50 mL distilled water and allowed to germinate in the dark at 25 °C. The abiotic treatment was carried out at the same conditions except supplementing with 10 μ M ABA (Sigma-Aldrich A1049), 5 μ M GA3 (Sigma-Aldrich G7646), 10 μ M paclobutrazol (Sigma-Aldrich 46046), or 10 μ M fluridon (Sigma-Aldrich 45511), respectively. The seeds were scored as germination if the emerging primary root reached 1 mm in length, and germination rates were calculated every 3 h in 24 h based on the mean \pm SD of three replicates (Fig. 2). The seedling sample was at the stage with 5 mm in length of primary roots. For different hormone and inhibitor treatments, samples were harvested at 12, 18 and 24 h. All samples were immediately frozen in liquid nitrogen and stored at -80 °C until used.

Total RNA isolation and cDNA synthesis

Total RNA was extracted using the TRIzol[®] Reagent Kit (Invitrogen) according to the manufacturer's instructions. RNA integrity was checked by 2 % gel electrophoresis followed by quantity and purity determination with the NanoDrop[™] 2000C spectrophotometer (Thermo Scientific). Only high-quality samples with OD ratios of at least

1.9 (260/280 nm) and 1.5 (260/230 nm) were used for further analysis. Twice biologically replicated RNA samples were used for downstream applications. Before cDNA synthesis, 6 μ g of RNA was treated with RNase-free DNase I Amplification grade kit (Invitrogen) according to the manufacturer's instructions, then reverse transcription reaction was carried out with M-MLV reverse transcriptase (Invitrogen) according to the manufacturer's protocol. In brief, 22 μ l (6 μ g) of RNA, 0.9 μ l of 200 μ M oligodT18 and 2 μ l of 10 mM dNTP mix were added together to incubate at 65 °C for 5 min, then kept on ice for at least 3 min. After that, 8 μ l of 5 \times RT (reverse transcription) buffer, 4 μ l of 0.1 M DTT, 1.5 μ l of ribonuclease inhibitor (40 U/ μ l, Takara) were added together, and pre-incubated 2 min at 37 °C. Finally, 2 μ l of M-MLV reverse transcriptase (200 U/ μ L) was added and the mixture was incubated 50 min at 37 °C. The RT reaction was terminated by heating at 70 °C for 15 min. The residual RNA was digested with 1 μ l of RNaseH (5 U/ μ L, NEB). Then it was used as the template for PCR amplification. This amplification was performed using primer pair sets which span an intron to detect DNA contamination. All cDNA samples were tenfold diluted with sterile water before being used as a template in RT-qPCR analysis.

Primer design and amplification efficiency test

In this study, we used primers both available from the literature and designed by ourselves (see Table S2 for details). If possible, the primers were designed in such a way that they locate on different exons or directly spanning exon–exon junction of each cDNA. The length of primers was 19–24 bp and amplicons 80–150 bp. The T_m of the primers was between 57 and 61 °C. And a few primers showed slightly lower T_m temperatures but PCR efficiencies were similar to those of the other primers, indicating that this small difference in T_m did not affect amplification. Primer specificity was confirmed by blasting each primer sequence against soybean genome sequences at Phytozome (<http://www.phytozome.net/>) using the BLASTN algorithm. For each primer pair, including those cited from the literature and designed by ourselves, reaction efficiency and R^2 value were evaluated from a standard curve generated from a serial dilution of pooled cDNA. Mean Ct values of each mixed sample were plotted against the logarithm of the cDNA dilution factor. An estimate of PCR efficiency was derived from the formula: $[10^{(1/S)} - 1] \times 100$ %, where S represents the slope of the linear regression. For the primers designed by ourselves a single peak was observed in a melting curve, confirming the synthesis of a single product, which was further confirmed by analysis of the RT-qPCR products on a 2.0 % agarose gel (Fig. S1).

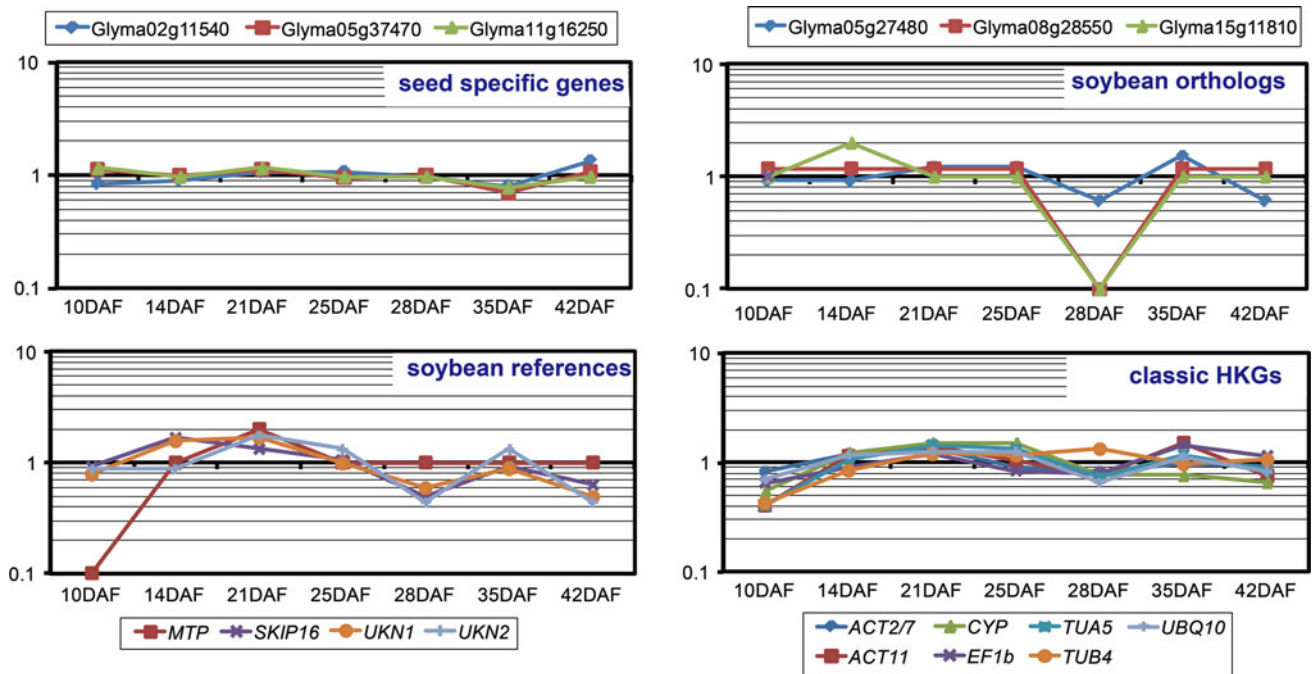


Fig. 3 Relative expression of putative reference genes during early seed development based on microarray data. These genes were divided into four sets: “seed specific genes” set, “soybean orthologs” set, “soybean references” set, and “classic HKGs” set (see text for

more detail). There were no data for Glyma18g04130 gene due to its undetectable expression level during the seed development. DAF indicated days after flowering

Real-time quantitative PCR

All quantitative PCR were performed in an ABI 7300 Real-Time PCR System (Applied Biosystem) using SYBR[®] Premix Ex Taq[™] KIT (Takara) to detect double-strand cDNA synthesis. Reactions were done in a tube containing 2.5 μ l of diluted cDNA (1:10), 7.5 μ l of SYBR[®] Premix Ex Taq[™], 0.3 μ l of ROX reference dye, 0.3 μ l of forward and reverse primers (10 μ M each), 4.1 μ l of sterile water. The conditions were set as the following: an initial polymerase activation step for 30 s at 94 $^{\circ}$ C, 40 cycles of 5 s at 94 $^{\circ}$ C for denaturation, 31 s at 60 $^{\circ}$ C for annealing and elongation. Threshold and baselines were manually determined using the ABI 7300 Real-Time PCR System SDS Software. The threshold was 0.2000. Baselines started at 3 and ended at 15. Routinely a melting curve analysis (between 60 and 95 $^{\circ}$ C with 0.5 $^{\circ}$ C increments for 15 s) was performed at the end of the RT-qPCR run. Each gene per sample was analyzed at least three technical replicates. For quality controls, three no-template controls for each primer pair were included per RT-qPCR plate, and sample data resulting from these primers were only taken when no-template control amplification was not evident.

Microarray and RT-qPCR data analysis

Microarray data-mined strategy was described in detail in “Results”. The expression ratio (Fig. 3) was calculated by

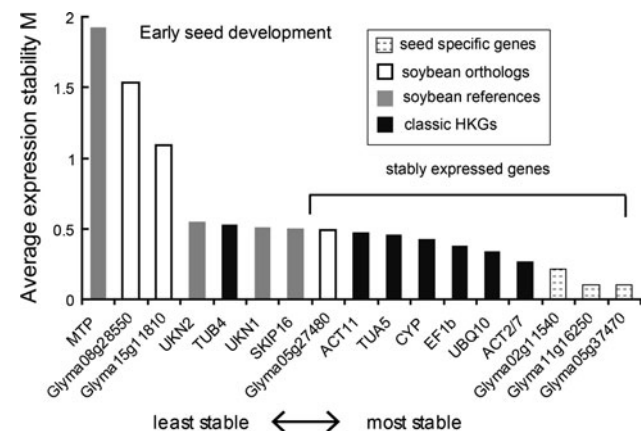


Fig. 4 Expression stability of putative references during early seed development analyzed by geNorm using microarray data. The genes were ranked by stepwise exclusion of the least stable genes. Genes with a high M value were less stably expressed ones compared to genes with a low M value. Genes had an M value ≤ 0.5 were considered as stably expressed (indicated by the frame)

dividing the expression per sample by the mean expression in all samples. For geNorm analysis (Fig. 4), the GeneChip expression data of these genes were converted into relative quantities via the expression of each sample divided by the max expression of all samples. If the relative amount for one gene in a particular sample was 0, we used 0.0001 instead of 0. For RT-qPCR data analysis, the coefficient of variation (CV) was calculated according to the following

formula: $CV = \text{standard deviation of Ct} / \text{average of Ct}$. Ct values were converted into relative quantities via the Ct method using the sample with the lowest Ct as calibrator and incorporating the calculated amplification efficiencies for each primer pair. The stability of reference gene expression was analyzed with the geNorm (v3.5) and NormFinder (v0.953) softwares using the relative quantities.

Results

Screening and identification putative soybean reference genes in early developmental seeds with microarray data

In order to select the most stably expressed genes in soybean seeds, we first searched publicly available soybean microarray database, including SoyBase (<http://www.soybase.org>), GENEVESTIGATOR (<https://www.genevestigator.com>), and Gene Network in Seed Development (<http://seedgenenetwork.net>). Then we mined the SoyBase database (<http://www.soybase.org>) for selection of candidate reference genes in an estimate of early developmental seeds using a similar approach as described by Czechowski et al. (2005). Each gene we calculated the mean expression value (MV) and the standard deviation (SD) during the early seed development. Then, the coefficient of variation (CV) was calculated as SD/MV . Genes with a low CV value were more stably expressed. A list of the partial genes with the lowest CV values was presented in Table S3. Three of them (Glyma02g11540, Glyma05g37470, Glyma11g16250) with significant difference of MVs were chosen as putative seed references and referred to “seed specific genes” set in this study. In addition, we added four soybean genes (Glyma05g27480, Glyma08g28550, Glyma15g11810, and Glyma18g04130), orthologs of *Arabidopsis* seed reference genes identified by Graeber et al. (2011) (At1g17210, At2g20000, At4g04320, and At2g04660, respectively). These four genes were referred to “soybean orthologs” set in this study. We also evaluated the expression stability of two sets of known reference genes, one including seven classic HKGs (*ACT2/7*, *ACT11*, *CYP*, *EF1b*, *TUA5*, *TUB4*, and *UBQ10*, referred to “classic HKGs” set here) and the other embracing four most stable soybean references (*SKIP16*, *UKN1*, *UKN2*, and *MTP*, referred to “soybean references” set here) in soybean total samples or various developmental stages (Hu et al. 2009).

To compare the variation in gene expression during early seed development, we calculated the expression ratio (the expression of per sample divided by the mean expression level in all samples) of these genes except Glyma18g04130 based on the microarray data (Fig. 3).

Figure 3 showed that both the “soybean references” set and the “soybean orthologs” set had considerable expression variation during early seed development, while the “classic HKGs” set showed more stable expression and the highest expression stability with the “seed specific genes” set. To further confirm the expression stability, the GeneChip expression data of these genes were calculated using geNorm. The average stability value M was calculated and genes were ranked according to their M values (the lower the M value, the higher the gene expression stability). This analysis identified ten stably expressed genes in total, considering their lower M values [≤ 0.5 , which is typical for stably expressed genes (Hellemans et al. 2007)] (Fig. 4). Among them, all three genes from the “seed specific genes” set were superior to other putative reference genes. The “classic HKGs” set also performed well, since they owned six genes in the ten of stably expressed genes. In contrast, the other two sets performed very poor with higher M value except Glyma05g27480. Taken together, these results implicated that the two sets of “seed specific genes” and “classic HKGs” may contain better reference candidates during the early and even whole seed development process.

There are no microarray data available for the middle or late stage of seed development and seed germination. Therefore, we did not carry out similar analysis for them.

Identification of putative soybean reference genes in developmental and germinating seeds by RT-qPCR

To confirm our speculation above and to validate the putative soybean reference genes in both developmental and germinating seeds, all of the 18 putative seed reference genes were identified in different seed samples (Table S1) by RT-qPCR experiment with primers both cited from the literature and designed by ourselves (Table S2). The amplification efficiencies of the all putative reference genes ranged from 86 to 110 % (Table S2), reflecting the high quality of PCR. Furthermore, all primer pairs showed an $R^2 > 0.99$ (Table S2), indicating a strong relationship between the detected Ct values and the corresponding relative amount of template in all the amplification reactions. In total 34 different soybean seed samples were used, which were further divided into two subsets (Table S1). One subset called “seed development” and included 12 developmental seed samples with different lengths (Fig. 1). The other subset named “seed germination” and contained ten untreated germinating seed samples (including one germinated seedling sample) and 12 treated germinating seed samples by ABA, GA3 and their corresponding inhibitors. Abscisic acid (ABA) and gibberellin (GA) are two antagonistic phytohormones that regulate seed germination. To test the effects of them and their corresponding

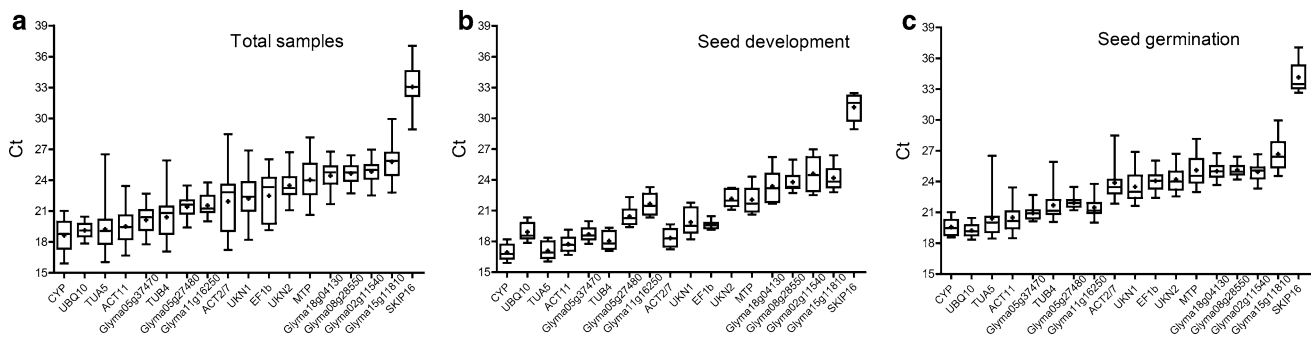


Fig. 5 Expression profiles of the putative reference genes. Values were given in the form of quantification cycle numbers (Ct values). Global expression levels (Ct values) in the different seed samples tested (**a**, **b**, and **c**) were shown as 25th and 75th quantiles (*horizontal*

lines), median (*emphasized horizontal lines*) and whiskers. Whiskers went from the minimal to maximal value. “+” indicated the mean Ct value

inhibitors, we compared the germination rate between the untreated and treated seeds (Fig. 2). We found that 10 μ M ABA can significantly inhibit seed germination. However, the germination rate was not increased when treated with its inhibitor 10 μ M fluridion. The germination was slightly promoted in response to 5 μ M GA3 treatment, while 10 μ M paclobutrazol which blocks GA biosynthesis can inhibit seed germination.

Firstly, we analyzed the expression levels of all 18 putative reference genes via their Ct values in total samples (Fig. 5a). The Ct values of these genes except *SKIP16* ranged from 15.9 to 30, which is within the range recommended for gene expression analysis by RT-qPCR (Karlen et al. 2007). *CYP* was highest expressed of the set with a mean Ct of 18.6, while *SKIP16* was expressed at the least level with mean Ct of 33.1. *SKIP16* was excluded owing to its mean Ct value higher than 30, avoiding interfering in the reaction. However, the above-mentioned data correspond to averages derived from total samples and further analysis showed that these genes exhibited more constant levels in both two subsets of “seed development” and “seed germination”, since the range of Ct values was much narrower at the single developmental process (Fig. 5b, c). In addition, the range of Ct values was relatively narrower in the “seed development” subset than that in the “seed germination” subset (Fig. 5b, c), indicating that gene expression in general was less variable during seed development than that during seed germination. Moreover, we found *Glyma18g04130* expressed not only in germinating seeds but also in developmental seeds—contrary to the result of microarray data. The reasons for this difference may be its low sensitivity of hybridization-based microarray chips and there is not a strict linear relationship between signal strength and transcript amount as that for RT-qPCR (Czechowski et al. 2005).

When we calculated their coefficient of variation (CV) of Ct values in all sample sets (Table S4), both “seed specific gene” and “soybean orthologs” sets performed

much better than the other two sets of reference genes, considering that they owned at least three genes in the top five of most stably expressed genes in all sample sets.

The calculation of the CV value was just an elementary in the validation of reference genes. It is also need to linearize the data, considering both the Ct value and amplification efficiency. Thus, the Ct values were converted into relative quantities via the Ct method with corrected primer efficiency, and then analyzed by geNorm (Fig. 6). We also used a cut-off value of ≤ 0.5 for the *M* value to identify the stably expressed genes. In total samples, *Glyma05g27480* and *Glyma08g28550* were the most stably expressed genes and *ACT2/7* still the least (Fig. 6a). This analysis identified 14 genes stably expressed in the subset of “seed development”, and the most stable genes were *TUA5* and *UKN2*, which were different from the result of microarray expression data during early seed development. But the “seed specific” set and “classic HKGs” set still performed well, and had eight genes in the top nine stably expressed genes, which was the similar to the result of microarray expression data (Fig. 6b). Five stably expressed genes were identified in the “seed germination” subset. Among them, top three (*Glyma05g37470*, *Glyma08g28550*, and *Glyma18g04130*) were from the “soybean orthologs” set or “seed specific genes” set (Fig. 6c). Similar results were generated in untreated germinating seed samples (Fig. S2). Surprisingly, the order of stability of the putative reference genes differed considerably in comparison of the rankings of two sample subsets. This indicated that seeds may show distinct transcriptomes at different developmental processes. To determine the optimal number of genes required for normalization, the pairwise variation (V_n/V_{n+1}) between the two sequential normalization factors (NF) (NF_n and NF_{n+1} , *n* for the number of genes used for normalization) was calculated (Fig. 6d). As reported by Vandesompele et al. (2002), a threshold value of 0.15 was adopted, below which represented *n* genes for normalization were adequate (Vandesompele et al. 2002). In the total

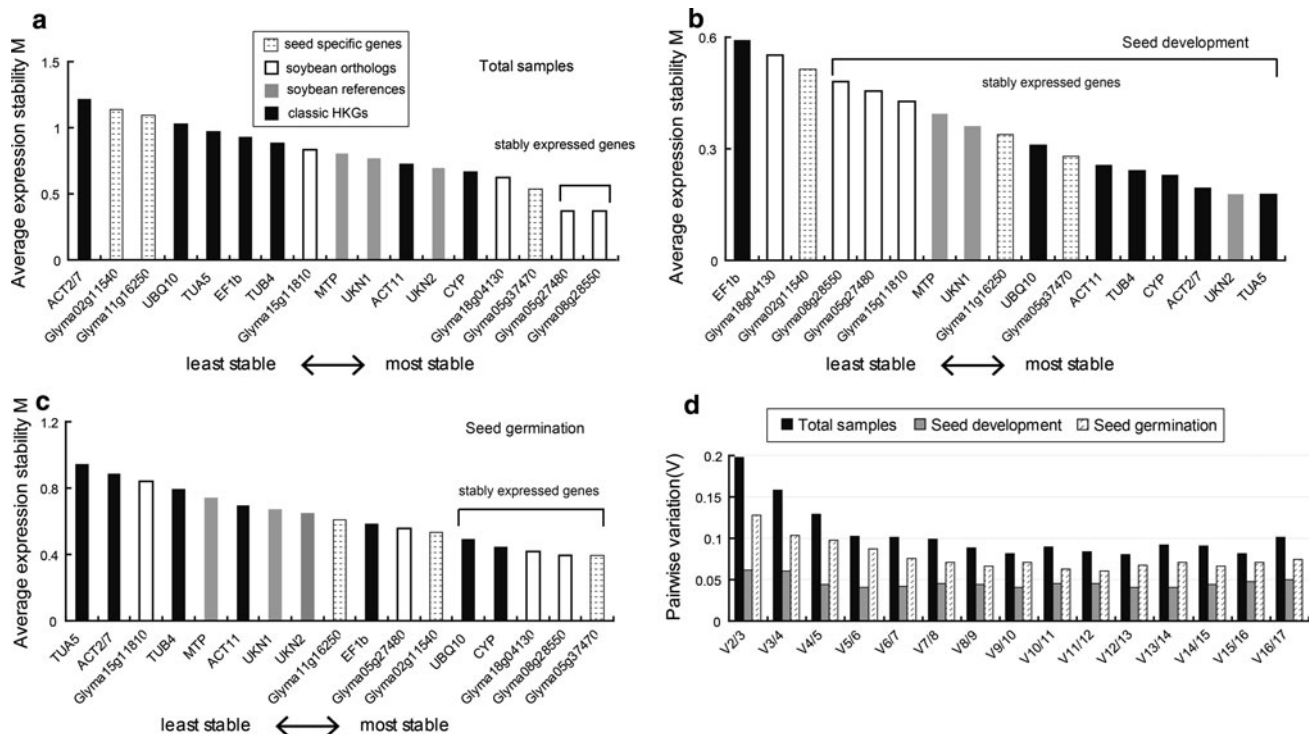


Fig. 6 Expression stability and pairwise variation of putative reference genes analyzed by geNorm using RT-qPCR data. Annotations for **a**, **b**, and **c** were same as Fig. 4. **d** Determination of the optimal

samples, four reference genes were needed for normalization, since the V3/V4 value was >0.15 and V4/5 value <0.15 . For the subset of “seed development”, the V2/V3 value was 0.06, suggesting two references were sufficient for normalization. And the same number of references was needed in “seed germination” subset, considering its V2/V3 value was <0.15 .

Stability of putative references expression was then re-analyzed using NormFinder, which is another software tool that ranks genes according to their stability of expression (Table 1). The ranking generated by this approach had some differences compared to that determined by geNorm, especially for the ranking of top five of most stable genes in total samples and “seed germination” subset. For example, *UKN2* was the most stable gene in total samples and “seed germination” subset for NormFinder analysis, but performed poor for geNorm analysis. Similarly, *Glyma05g27480* and *Glyma08g28550*, which were two top ranked genes for geNorm analysis in total samples, performed not well for NormFinder analysis. Despite a visible variation in the ranking of the most stable genes generated by geNorm and NormFinder algorithms, when compared the top five stably expressed genes in all three sample sets, at least three genes were identified by both programs. Taken together, we merged the outputs of both programs to assess the expression stability of these putative reference genes, that is, the reference genes identified as optimums

number of reference genes for normalization by pairwise variation (V_n/V_{n+1}). A threshold value of 0.15 was adopted, below which represented n genes adequate for normalization

for seed development were *TUA5* and *UKN2*, whereas for seed germination were *Glyma05g37470* and *Glyma08g28550*. Furthermore, for total samples it was necessary to combine four genes of *Glyma05g37470*, *Glyma08g28550*, *Glyma18g04130*, and *UKN2* for normalization.

Validation of putative reference genes

To test the putative reference genes identified above, we monitored the mean expression of two target genes during seed development and germination with different control genes. One gene was *IFS2* (*Glyma13g24200*), an isoflavone synthase gene, whose expression has been shown to increase during early stage (SoyBase, <http://www.soybase.org>) and middle to late stage of seed development (Gutierrez-Gonzalez et al. 2010). When normalized using the most stable gene *TUA5* or *UKN2* (Fig. 6b; Table 1) or both of them, the expression of *IFS2* gradually increased during seed development, peaking at the S11 or S12 stage (Fig. 7a). This expression pattern of *IFS2* was consistent with microarray data (SoyBase, <http://www.soybase.org>) and previous report (Gutierrez-Gonzalez et al. 2010). In contrast, normalization based on the least stable gene *EF1b* (Fig. 6b; Table 1), the transcript level of *IFS2* was relative constant during seed development, especially during S6–S9 stage (Fig. 7a).

Table 1 Ranking of putative references in the order of their expression stability calculated by NormFinder

Rank	Gene		
	Total samples	Seed development	Seed germination
1	<i>UKN2</i>	<i>TUA5</i>	<i>UKN2</i>
2	Glyma05g37470	<i>UKN2</i>	<i>EF1b</i>
3	<i>ACT11</i>	<i>ACT2/7</i>	Glyma05g37470
4	<i>CYP</i>	<i>TUB4</i>	Glyma18g04130
5	Glyma18g04130	<i>UBQ10</i>	Glyma08g28550
6	Glyma08g28550	<i>CYP</i>	<i>UKN1</i>
7	<i>UKN1</i>	Glyma11g16250	<i>ACT11</i>
8	Glyma15g11810	<i>ACT11</i>	Glyma11g16250
9	Glyma05g27480	Glyma15g11810	<i>CYP</i>
10	<i>MTP</i>	<i>UKN1</i>	Glyma05g27480
11	<i>TUB4</i>	Glyma05g37470	<i>UBQ10</i>
12	<i>TUA5</i>	<i>MTP</i>	<i>MTP</i>
13	<i>EF1b</i>	Glyma05g27480	<i>TUB4</i>
14	<i>UBQ10</i>	Glyma08g28550	Glyma15g11810
15	Glyma11g16250	Glyma18g04130	Glyma02g11540
16	Glyma02g11540	Glyma02g11540	<i>ACT2/7</i>
17	<i>ACT2/7</i>	<i>EF1b</i>	<i>TUA5</i>

The other target gene was *GmABI5* (Glyma10g08370), an ortholog of *ABI5* in *Arabidopsis*, whose expression pattern during germination is still unknown. The expression of *ABI5* in *Arabidopsis* seeds during germination was decreased, and this decline was inhibited by exogenous ABA (Xi et al. 2010). When using the two top ranked genes Glyma05g37470 and Glyma08g28550 (Fig. 6c) for normalization, the expression of *GmABI5* drastically decreased from 0 to 6 h of imbibitions and then relative constant after 6 h of imbibitions (Fig. 7b). Similar expression pattern was shown when normalized based on the least stable gene *TUA5* (Fig. 6c; Table 1), but difference was evident in estimated transcript abundance from 0 to 6 h of imbibitions (Fig. 7b). Furthermore, the reduction of *GmABI5* expression during germination was also inhibited by exogenous ABA (Fig. S3). The similar

expression patterns of *GmABI5* and *ABI5* implicated that they may play the similar role in seeds of different species. Taken together, these results proved that the superior reference genes identified in this study were suitable for normalization in soybean seeds.

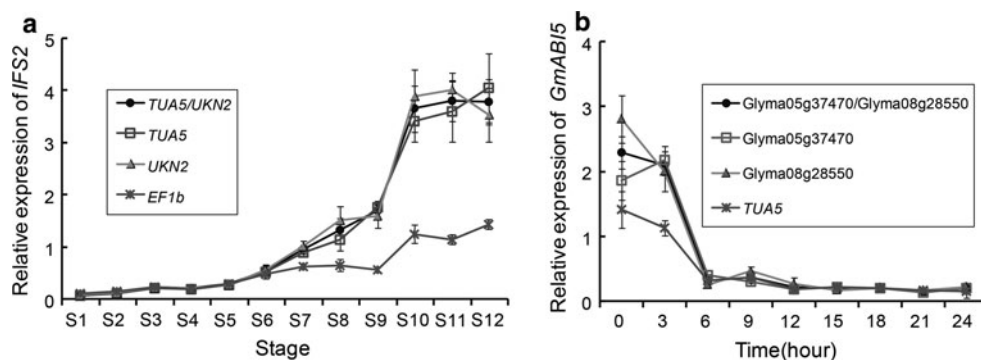
Discussion

To quantify the expression of target genes by RT-qPCR requires an accurate normalization. The reference genes used should be expressed at a constant level in the corresponding experimental system. However, no genes are universally stably expressed and all of them are regulated to a certain extent (Hruz et al. 2011). Thus, identification and validation of best reference genes for the corresponding study are important.

To screen novel candidate reference genes, we used two strategies in this study. One strategy was to exploit stably expressed genes from microarray data that has been successfully employed in several plant species like *Arabidopsis* (Czechowski et al. 2005; Dekkers et al. 2012), rice (Narsai et al. 2010) *Lepidium sativum* (Graeber et al. 2011) and soybean (Hu et al. 2009; Libault et al. 2008). The other strategy was employing putative orthologs of known reference genes in other species for the similar study; this was a strategy proposed by Czechowski et al. (2005) and proved available in several reports (Dekkers et al. 2012; Graeber et al. 2011; Hu et al. 2009). In total, we mined seven novel candidate references, among which there were three most stable genes for total samples and “seed germination” subset (Fig. 6a, c). In addition, most of these genes, although not the best, also performed well during seed development (Fig. 6b). These results showed that the two strategies were successfully employed in this study.

Each method can introduce bias, and we also observed some differences in the ranking of putative reference genes generated by geNorm and NormFinder. These differences in the ranking may be caused by the discrepancies between these two approaches as reported by Anderson et al. (2004). Therefore, it should merge the outputs of both

Fig. 7 Relative expression levels of *GmIFS2* and *GmABI5* normalizing with different reference genes. **a** Relative expression of *GmIFS2* during seed development. **b** Relative expression of *GmABI5* during seed germination



programs to minimize the bias, and further test the integrated best references using known target genes.

Accumulating data showed that seeds not only have very different transcriptomes from other tissues (Cadman et al. 2006; Czechowski et al. 2005; Le et al. 2010; Schmid et al. 2005) but also exhibit massive changes during seed development and germination (Howell et al. 2009; Le et al. 2010; Okamoto et al. 2010; Preston et al. 2009). These massive changes of seed transcriptomes were also verified by the expression profiling of putative reference genes in this study (Fig. 5). We found that the two subsets of “seed development” and “seed germination” exhibited more constant levels of tested transcripts than total samples. Therefore, four reference genes were needed for normalization in the total samples and only two references were needed in single sample subsets (Fig. 6d). In addition, the gene expression, in general, was less variable during seed development than that during seed germination (Fig. 5b, c), so that the stability ranking of putative references in two sample subsets varied tremendously (Fig. 6b, c). This obvious difference of ranking during seed development and germination was also existent in *Arabidopsis* seeds (Graeber et al. 2011).

It should be pointed that transcriptomes can differ considerably between specific seed tissues, which have been shown in soybean (Le et al. 2007), *Lepidium sativum* (Linkies et al. 2009), barley (Barrero et al. 2009), *Arabidopsis* (Le et al. 2010). Therefore, the stably expressed genes validated in whole seeds in this study may not be applied for specific seed tissues, such as embryo and endosperm. Moreover, these genes may not be used widely under the conditions that are not included in the present study, such as stratification, salt, cold and heat stress conditions. In these cases, using multiple reference genes across the experiment or re-evaluating novel reference genes is recommended. In brief, different tissues or experimental conditions should have their own special reference genes.

Conclusion

In this study, we have investigated the expression of 18 putative reference genes to identify the best reference genes suitable for normalization in soybean seeds in the RT-qPCR experiment. We identified several genes that stably expressed in seed developmental and germinating processes. Moreover, the expression level of some novel reference genes mined in this study was indeed more stable than the traditional ones. Taken together, using the combinations of these references would provide a more reliable means for accurate normalization of soybean seed experiments.

Acknowledgments This work was partly supported by Transgenic program (Nos 2009ZX08009-133B, 2011ZX08009-001 and 2011ZX08004-005), the Chinese National Key Basic Research “973” Program (2010CB125906), and the National Natural Science Funds (31000680 and 31000681).

Conflict of interest The authors declare that they have no conflict of interest.

References

- 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
- Barrero JM, Talbot MJ, White RG, Jacobsen JV, Gubler F (2009) Anatomical and transcriptomic studies of the coleorhiza reveal the importance of this tissue in regulating dormancy in barley. *Plant Physiol* 150:1006–1021
- Cadman CS, Toorop PE, Hilhorst HW, Finch-Savage WE (2006) Gene expression profiles of *Arabidopsis* Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. *Plant J* 46:805–822
- 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
- Dekkers BJ, Willems L, Bassel GW, van Bolderen-Veldkamp RP, Ligterink W, Hilhorst HW, Bentsink L (2012) Identification of reference genes for RT-qPCR expression analysis in *Arabidopsis* and tomato seeds. *Plant Cell Physiol* 53:28–37
- Graeber K, Linkies A, Wood AT, Leubner-Metzger G (2011) A guideline to family-wide comparative state-of-the-art quantitative RT-PCR analysis exemplified with a brassicaceae cross-species seed germination case study. *Plant Cell* 23:2045–2063
- Gutierrez L, Mauriat M, Guenin S, Pelloux J, Lefebvre JF, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C, Van Wuytswinkel O (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
- Gutierrez-Gonzalez JJ, Guttikonda SK, Tran LS, Aldrich DL, Zhong R, Yu O, Nguyen HT, Slepser DA (2010) Differential expression of isoflavone biosynthetic genes in soybean during water deficits. *Plant Cell Physiol* 51:936–948
- Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8:R19
- Holdsworth MJ, Bentsink L, Soppe WJ (2008) Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. *New Phytol* 179:33–54
- Howell KA, Narsai R, Carroll A, Ivanova A, Lohse M, Usadel B, Millar AH, Whelan J (2009) Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. *Plant Physiol* 149:961–980
- Hruz T, Wyss M, Docquier M, Pfaffl MW, Masanetz S, Borghi L, Verbrugge P, Kalaydjieva L, Bleuler S, Laule O, Descombes P, Gruissem W, Zimmermann P (2011) RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. *BMC Genomics* 12:156

- Hu R, Fan C, Li H, Zhang Q, Fu YF (2009) Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol Biol* 10:93
- Jian B, Liu B, Bi Y, Hou W, Wu C, Han T (2008) Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol Biol* 9:59
- Karlen Y, McNair A, Perseguers S, Mazza C, Mermod N (2007) Statistical significance of quantitative PCR. *BMC Bioinform* 8:131
- Kulcheski FR, Marcelino-Guimaraes FC, Nepomuceno AL, Abdelnoor RV, Margis R (2010) The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. *Anal Biochem* 406:185–192
- Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB (2007) Using genomics to study legume seed development. *Plant Physiol* 144:562–574
- Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamoto JK, Harada JJ, Goldberg RB (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci USA* 107:8063–8070
- Lee PD, Sladek R, Greenwood CM, Hudson TJ (2002) Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res* 12:292–297
- Libault MT, Bilgin D, Radwan O, Benitez M, Clough SJ, Stacey G (2008) Identification of four soybean reference genes for gene expression normalization. *Plant Genome* 1:44
- Linkies A, Muller K, Morris K, Tureckova V, Wenk M, Cadman CS, Corbineau F, Strnad M, Lynn JR, Finch-Savage WE, Leubner-Metzger G (2009) Ethylene interacts with abscisic acid to regulate endosperm rupture during germination: a comparative approach using *Lepidium sativum* and *Arabidopsis thaliana*. *Plant Cell* 21:3803–3822
- Narsai R, Ivanova A, Ng S, Whelan J (2010) Defining reference genes in *Oryza sativa* using organ, development, biotic and abiotic transcriptome datasets. *BMC Plant Biol* 10:56
- Okamoto M, Tatematsu K, Matsui A, Morosawa T, Ishida J, Tanaka M, Endo TA, Mochizuki Y, Toyoda T, Kamiya Y, Shinozaki K, Nambara E, Seki M (2010) Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of *Arabidopsis* using tiling arrays. *Plant J* 62:39–51
- 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
- Preston J, Tatematsu K, Kanno Y, Hobo T, Kimura M, Jikumaru Y, Yano R, Kamiya Y, Nambara E (2009) Temporal expression patterns of hormone metabolism genes during imbibition of *Arabidopsis thaliana* seeds: a comparative study on dormant and non-dormant accessions. *Plant Cell Physiol* 50:1786–1800
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37:501–506
- Suzuki T, Higgins PJ, Crawford DR (2000) Control selection for RNA quantitation. *Biotechniques* 29:332–337
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3:RESEARCH0034
- Weitbrecht K, Muller K, Leubner-Metzger G (2011) First off the mark: early seed germination. *J Exp Bot* 62:3289–3309
- Xi W, Liu C, Hou X, Yu H (2010) MOTHER OF FT AND TFL1 regulates seed germination through a negative feedback loop modulating ABA signaling in *Arabidopsis*. *Plant Cell* 22:1733–1748