



Identification and evaluation of endogenous control genes for use in quantitative RT-PCR during wheat (*Triticum aestivum* L.) grain filling

D. Wu¹, J. Dong^{1,2}, Y.J. Yao¹, W.C. Zhao¹ and X. Gao^{1,2}

¹College of Agronomy, Northwest A&F University, Yangling, China

²Wheat Engineering Research Center of Shaanxi Province, Shaanxi, China

Corresponding author: X. Gao

E-mail: gx@nwsuaf.edu.cn

Genet. Mol. Res. 14 (3): 10530-10542 (2015)

Received February 5, 2015

Accepted May 15, 2015

Published September 8, 2015

DOI <http://dx.doi.org/10.4238/2015.September.8.15>

ABSTRACT. The use of appropriate reference genes is essential for the generation of accurate and biologically meaningful results from quantitative real-time PCR (qRT-PCR) analysis. However, studies have found that the expression of most commonly used reference genes is not always independent of the tissues, treatments, or developmental stages studied. geNormPlus, NormFinder, and BestKeeper, were applied and the expression stability of nine candidate genes was evaluated in different data sets during wheat grain development. Varying degrees of diversity in either single or multiple reference genes were observed among the results generated from the different computer programs, parameters, and data sets. Therefore, the reliability of identified reference genes in the flag leaf and the complete set of samples was estimated by monitoring the expression dynamics of three *NAM* genes (*TaNAM-A1*, *TaNAM-B1*, and *TaNAM-B2*). The results suggest that a single control gene identified by geNormPlus for use with the complete set of samples, and multiple reference genes selected by geNormPlus and NormFinder exclusively for the flag leaf outperformed others owing to the consistent results with previous analyses of these genes, which

were normalized against a verified single control gene. Given the limit of NormFinder in gene numbers of multiple reference genes, robust quantification can be achieved by normalizing against *Ta27922* or multiple reference genes chosen by geNormPlus for individual tissues.

Key words: *Triticum aestivum* L.; Quantitative real-time PCR; Grain filling; Endogenous control gene

INTRODUCTION

Owing to its higher sensitivity and specificity in contrast to reverse transcription-polymerase chain reaction (RT-PCR), northern blotting and microarray (Kriegova et al., 2008), qRT-PCR has become a widely used method to quantify the dynamic expression of genes of interest, especially those expressed at low levels (Kumar et al., 2011). Nevertheless, normalization is required during qRT-PCR analysis to control for variations introduced throughout the experimental process. A preferred technique used for normalization is the use of endogenous control genes (Argyropoulos et al., 2005). Several studies have demonstrated that different reference genes can change the outcomes of a quantitative study (Dheda et al., 2005; Barsalobres-Cavallari et al., 2009). Therefore, an accurate result can only be achieved if the normalization standards used are abundant and stably expressed genes. Commonly used reference genes, or housekeeping genes (HKGs), include those encoding actin, rRNA, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), elongation factor (EF), α -tubulin, and ubiquitin, which are necessary for maintaining basic cellular processes and thus assumed not to be differently regulated under different experimental conditions. However, some of these genes are far from ideal. For example, the rRNA and mRNA contents are not balanced in rat (Solanas et al., 2001), and rRNA transcription is variable under different biotic treatments (Johnson et al., 1995; Warner, 1999). This is also true in plants. Numerous studies have revealed that these “HKGs” are not sufficiently stable under some treatments (Thellin et al., 1999), during development (Bustin, 2002), or even in different tissues (Czechowski et al., 2005) due to the variable nature of plant genes. In wheat, *18S rRNA* showed uniform expression in *P. graminis* f. sp. *tritici*-infected cultivars compared with *P. triticina*- and *P. striiformis*-infected cultivars (Scholtz and Visser, 2013). *GAPDH* was constitutively expressed throughout the whole developmental process in chickpea and sugarcane, whereas it was the most unstable gene during wheat development (Czechowski et al., 2005). Additionally, the expression of β -actin fluctuated in *yellow mosaic virus*-infected wheat (Liu et al., 2013). Accordingly, previous reports have emphasized that an ideal all-purpose inner control gene does not exist, and that the screening of proper reference genes under certain experimental condition is crucial. Paolacci et al. (2009) retrieved 26 research articles published from January 1996 to March 2008, which utilized qRT-PCR to quantify gene expressions or to confirm the results of screening technologies of differentially expressed genes in wheat. However, almost three quarters of these articles used genes encoding 18S rRNA (eight times, 30%), actin (seven times, 27%), and α -tubulin (five times, 19%) for normalization purpose, and moreover, 15 of these were conducted without their expressions being verified as stable. Even more recently, quantitative analyses have seldom been carried out with verification of the endogenous control gene. Furthermore, although an accumulating number of studies have involved reference gene mining, these fail to meet the needs of diverse kinds of experiments.

The grain-filling period is critical in the production of high yields of wheat. Approximately 70-90% of the final grain dry weight is determined at this stage (Wang et al., 2008). However, the genetic regulation of dry weight accumulation during grain filling is complex and highly coordinated, and the knowledge of genes involved in this process is far from sufficient to shed light on the underlying molecular mechanisms. There is no doubt that gene expression profiling could provide valuable information, whereas inaccurate outcomes of qRT-PCR analysis would result in bias or errors. Accordingly, the present study aimed to evaluate the global expression stability of nine candidate reference genes and to define the optimal single and multiple reference genes for ten data sets (different tissues and the complete set of samples) during wheat grain filling using three freely available computer programs. These results should provide normalization standards for further gene expression assessments that are carried out during this developmental stage.

MATERIAL AND METHODS

Plant materials

The seeds of bread wheat cultivar Chinese Spring were grown at the Wheat Breeding Center of Northwest A&F University in Yangling, Shaanxi Province (34°26'N, 108°14'E). The flag leaf, flag leaf sheath, penultimate leaf, penultimate leaf sheath, peduncle, glume, palea, and lemma, rachis, and kernel were collected every five days from the date of first bloom until 30 days after anthesis (DAA). Each tissue was harvested from three tillers at each time point.

RNA extraction and cDNA synthesis

RNA was extracted using an RNA extraction kit (Biotek, Beijing, China), following the manufacturer protocol. The integrity of RNA was examined by formaldehyde polyacrylamide gel electrophoresis. The DNA was removed using DNase I (Takara, Shiga, Japan) and the RNA was subsequently quantified by Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). RNA (1 µg) was used for cDNA synthesis with M-MLV reverse transcriptase (Takara). A total of 20 µL reverse transcription products were diluted with 80 µL ddH₂O. A pair of primers specifically targeting an intron region of the protein disulfide isomerase (PDI) gene was applied to detect any DNA contamination in the cDNA. The primer sequences were listed as followed: PDI-F: 5'-CTTATGCAATCCAATGATGG-3' and PDI-R: 5'-TCTATGGTTCTG AAGAGGACC-3' (Paolacci et al., 2009).

Quantitative real-time PCR

The DNA-free cDNA was used to perform quantitative analyses. The reaction was run on a CFX96™ Real-Time PCR Detection System (Bio-Rad, CA, USA) following the program: initial denaturation at 95°C for 30 s, 40 cycles at 95°C for 5 s, 55°C for 30 s, and 72°C for 30 s. The reaction mixture (20 µL) contained 1 µL cDNA, 10 µL SYBR® Premix Ex Taq™ II (TaKaRa), and 0.6 µL of each primer at a final concentration of 300 nM. Each reaction was performed in duplicate, and the average threshold cycle (Ct) value was used for subsequent analysis. Nine genes were chosen as candidate genes, all of which were previously verified as being highly stable under different research conditions in wheat (Paolacci et al., 2009; Long

et al., 2010; Giménez et al., 2011). The specificity of each primer pair was confirmed by melting curve analysis and sequencing of qRT-PCR amplicons. The amplification efficiencies of all primer pairs were estimated using five 10-fold dilutions and validated as being between 90 and 105% (Ginzinger, 2002). Information about reference candidate genes is listed in Table 1.

Table 1. Information relating to the candidate reference genes (Paolacci et al., 2009; Long et al., 2010; Giménez et al., 2011).

Gene	Annotation	Primer Pair (5' to 3')
<i>Ta54227</i>	Cell division control protein, AAA - superfamily of ATPases	F: CAAATACGCCATCAGGGAGAACATC R: CGCTGCCGAAACCACGAGAC
<i>Ta2291</i>	ADP- ribosylation factor	F: GCTCTCCAACAACATTGCCAAC R: GCTTCTGCCTGTCACATACGC
<i>Ta2776</i>	RNase L inhibitor-like protein	F: CGATTCAGAGCAGCGTATTGTTG R: AGTTGGTCGGGTCTTCTAAATG
<i>Ta27922</i>	Unknown EST, most likely a cyclin-like protein	F: CAGGTGCTTGTTCCTATG R: GCAGCCTCTTTCCTATCGTTC
<i>Ta53967</i>	Vacuolar-H ⁺ -pumping ATPase 16-kDa proteolipid subunit 2-like protein	F: AGTGCTCGGTATCTACGG R: TGCGAAGATGAGGATGAGG
<i>Ta54171</i>	Superoxide dismutase [Cu-Zn]	F: TGAGCAAGAGCACTGGAAC, R: CGTTGGTCGGCGAAGATG
<i>Ta35284</i>	Protein transport protein	F: AGCAATTGCGACAATTATTACAAG R: CTCACAGAAGACCTGGAAGC
<i>EF</i>	Elongation factor 1 alpha-subunit	F: CAGATTGGCAACGGCTACG R: CGGACAGAAAACGACCAAG
<i>Ta53919</i>	S-adenosylmethionine decarboxylase precursor	F: GGCTGGACAAGAAGAAG R: ATGGATGGTGGAGACG

Stability of reference candidate genes

Three widely used software programs-geNormPlus (<http://www.biogazelle.com/geNormPlus>), BestKeeper (<http://www.gene-quantification.de/BestKeeper.html>) and NormFinder (<http://www.multid.se/genex/hs410.htm>)-were employed to assess the expression stability of candidate genes.

The relative quantity was imported as raw data for both geNormPlus and NormFinder, which was calculated according to the $2^{-\Delta C_t}$ method using the lowest Ct value as the calibrator (Livak and Schmittgen, 2001). geNormPlus, with a minor change in the algorithm of geNorm, identifies optimal reference genes by determining the standard deviation (SD) of the logarithmically transformed expression ratios of a gene to other candidate genes (Vandesompele et al., 2002; Maroufi et al., 2010). The M value, which is the average pair-wise variation of a particular gene with all other genes, is hence set to evaluate the stability. The threshold M value for stable genes depends on the heterogeneity of the sample set. Normally, 0.5 is considered to be a strict threshold for homogeneous samples, and 1 applies to more heterogeneous sample sets. The candidate genes with the lowest M values can be regarded as the most stable. geNormPlus also evaluates the optimal gene numbers for multiple reference genes (Hellemans et al., 2007), the pair-wise variation (V_n/V_{n+1}) between two sequential normalization factors is hence calculated (Vandesompele et al., 2002). This process starts from the top two genes on the ranking list of expression stability and proceeds until the V_n/V_{n+1} drops below 0.15, which means that an additional gene is unnecessary to attain a robust and reliable normalization factor.

NormFinder assesses the expression stability using an ANOVA-based model and also takes the inter- and intra-group expression variations of the genes tested into consideration. Therefore, misinterpretations caused by co-regulated genes can be avoided (Andersen et al.,

2004). A lower stability value indicates higher expression stability with less intra- and inter-group variation (Andersen et al., 2004).

BestKeeper employs a pair-wise correlation analysis to identify the optimal reference gene based on the Ct value (Pfaffl et al., 2004). BestKeeper determines variation in gene expression based on the SD and Pearson correlation coefficient (r) between each gene. The most stable gene exhibits the lowest SD value (usually less than 1) and higher correlations with other genes at significant levels (Pfaffl et al., 2004). The BestKeeper index (BI) is the geometric mean of Ct values of genes that are considered stable for a respective sample set, and genes that composed multiple reference genes normally have significantly higher correlations with BI.

Evaluations of reference genes identified

Single and multiple reference genes that were identified for the complete sample set and from the flag leaf by different programs and indexes were evaluated.

The expression dynamics of *TaNAM-A1* (KM820886), *TaNAM-B1* (KM820887), and *TaNAM-B2* (KM820888) were analyzed in the flag leaf only, the orthologs of which were previously studied using *actin* as the reference gene and its constitutive expression was confirmed by a transcriptome analysis (Uauy et al., 2006). In a previous study, the transcripts of their orthologs increased slightly before 10 DAA, and then rose significantly towards 25 DAA (Uauy et al., 2006). Information relating to the specific primers is listed in [Table S1](#).

RESULTS

Expression levels of reference candidate genes

According to the Ct values, the expression levels of nine candidate genes are shown across all samples (Figure 1). In general, all of the Ct values ranged from 18 to 30 cycles, in which *Ta54171* (20.92 ± 1.30) and *Ta27922* (26.98 ± 1.09) had the highest and the lowest average Ct values, respectively. Based on the coefficient of variation (CV) of the Ct value, *Ta27922* exhibited less Ct dispersion than other genes in the complete set of samples, penultimate leaf, flag leaf, and the peduncle, and that in the flag leaf sheath, rachis, palea, and lemma was *EF* ([Table S2](#)). Additionally, the genes with the lowest CV values in the remaining tissues included those in the penultimate leaf sheath (*Ta2291*), glume (*Ta54227*), and the kernel (*Ta2776*) were indistinct.

Estimation of the stability of candidate reference genes

geNormPlus analysis

The results indicated that all of the candidate reference genes were stably expressed ($0.5 < M < 1$) in the complete set of samples during grain filling, in which *Ta27922* was the most constant. Furthermore, the best combination of genes (BCTG) for multiple reference genes was *Ta27922*, *Ta2291*, *Ta2776*, and *EF* (data not shown).

Optimal inner control genes were also estimated for the respective tissues (Figure 2). The results showed that the stability rankings of candidate genes and the BCTGs differed greatly between tissues. None of the candidate genes in the flag leaf and the sheaths showed relatively steady expression. *Ta2291* was identified as a suitable reference gene in the penulti-

mate leaf, while that for the rachis, palea, and lemma was *Ta53919*. In the remaining tissues, *Ta53967* outperformed others.

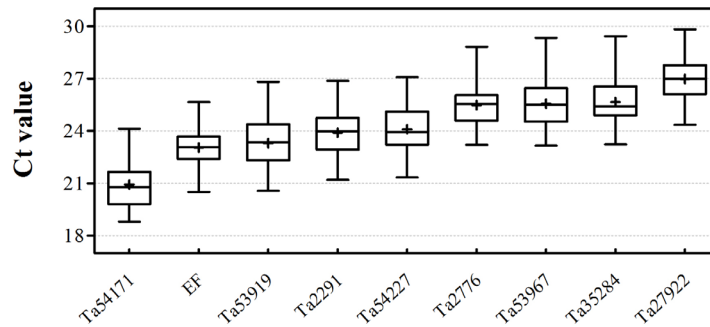


Figure 1. Expression levels of reference candidate genes across all of the samples. The genes were sorted based on the average Ct values. The average (crosses), median (horizontal lines), 25 to 75th percentile (boxes), and expression ranges (whiskers) are shown for each gene.

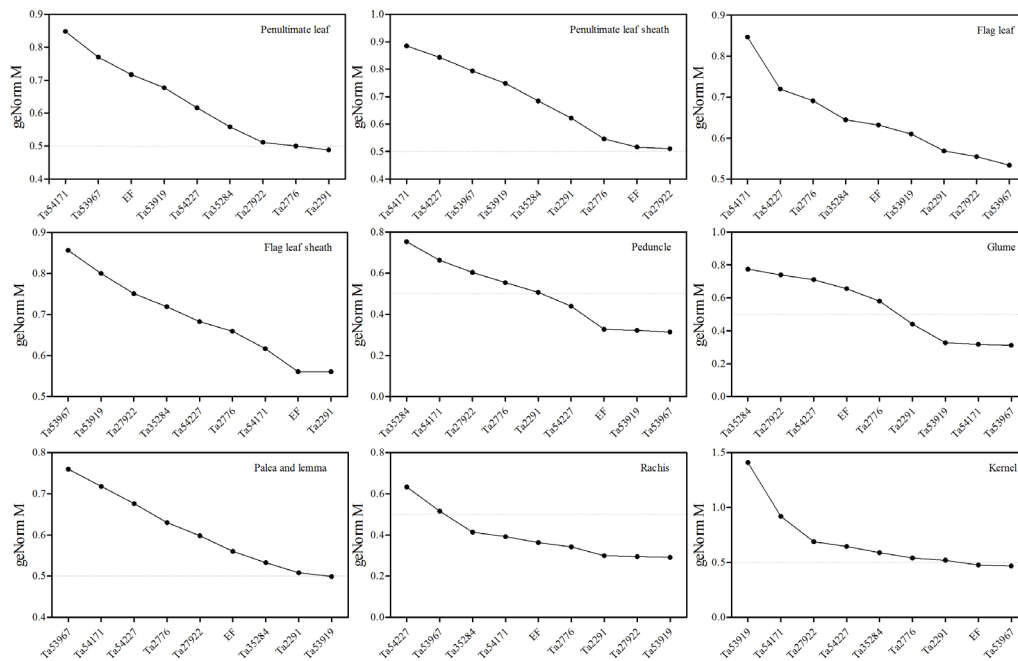


Figure 2. Stability rankings of reference candidate genes based on the M value calculated by geNormPlus. A strict threshold M value for homologous tissues is 0.5. The lower the value, the more stable the gene expression.

BCTG was the combination of the top n genes on the ranking list in terms of expression stability (Figure 3), and n was the optimal gene numbers determined by the V value. The geometric average of relative expression levels of *Ta53967*, *EF*, and *Ta2291* can be applied to normalization of quantitative analysis in wheat kernel, and that of *Ta2291*, *Ta2776*, *EF*, and

Ta27922 can be used in the penultimate leaf sheath. Furthermore, only two genes were necessary for normalization-based quantification in the peduncle (*Ta53919* and *Ta53967*), glume (*Ta54171* and *Ta53967*), and the rachis (*Ta27922* and *Ta53919*), while three genes were required for the flag leaf (*Ta2291*, *Ta27922*, and *Ta53967*), its sheath (*EF*, *Ta2291* and *Ta54171*), penultimate leaf (*Ta27922*, *Ta2776*, and *Ta2291*), as well as the palea and lemma (*Ta35284*, *Ta2291*, and *Ta53919*).

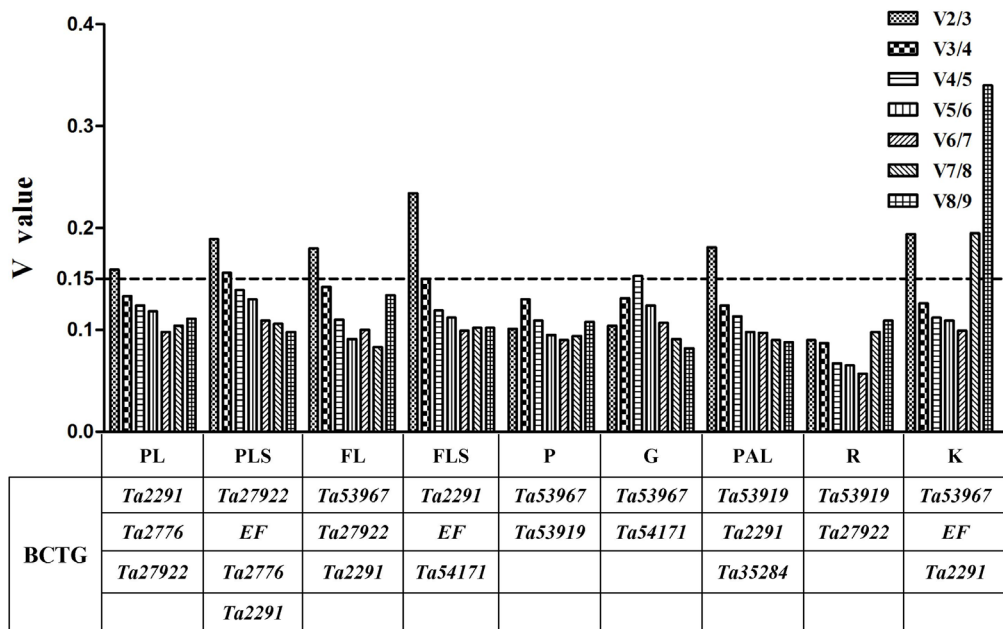


Figure 3. Optimal gene numbers of multiple reference genes identified by geNormPlus. An additional gene is unnecessary to attain robust and reliable multiple reference genes when the V value drops below 0.15. The genes in the table below the histogram were the optimal genes to be combined. PL: Penultimate leaf; PLS: Penultimate leaf sheath; FL: Flag leaf; FLS: Flag leaf sheath; P: Peduncle; PAL: Palea and lemma; G: Glume; R: Rachis; K: Kernel.

NormFinder analysis

The results from NormFinder also changed among different data sets to some extent (Table 2). Specifically, *EF* showed the highest stability in the complete set of samples. *Ta2291* was the most constantly expressed gene in the penultimate leaf, flag leaf sheath, and the kernel, while in the penultimate leaf sheath, glume, palea, and lemma it was *Ta2776*. *Ta53967*, *Ta53919*, and *Ta27922* ranked first in the flag leaf, peduncle, and the rachis respectively. The two most stable genes were usually selected as the optimal BCTG for each tissue; however, *Ta2291* (2nd) and *Ta2776* (4th) were indicated as the best pair of genes for the complete set of samples. *Ta2291* and *EF* were the optimal BCTG shared by the flag leaf sheath and the kernel. Furthermore, unique BCTGs were identified for the penultimate leaf (*Ta2291* and *Ta27922*), its sheath (*Ta2776* and *Ta35284*), flag leaf (*Ta53967* and *Ta27922*), peduncle (*Ta2291* and *Ta53919*), and the rachis (*Ta2776* and *Ta27922*).

Table 2. Stability rankings of reference candidate genes and the optimal multiple reference genes identified by NormFinder. The genes are listed in ascending order based on the expression stability value. AT: All samples.

Ranking	AT	PL	PLS	FL	FLS	P	G	PAL	R	K
1	<i>EF</i> 0.0983	<i>Ta2291</i> 0.0103	<i>Ta2776</i> 0.0160	<i>Ta53967</i> 0.0272	<i>Ta2291</i> 0.0417	<i>Ta53919</i> 0.0286	<i>Ta2776</i> 0.0422	<i>Ta2776</i> 0.0244	<i>Ta27922</i> 0.0109	<i>Ta2291</i> 0.0182
2	<i>Ta2291</i> 0.0985	<i>Ta27922</i> 0.0267	<i>Ta35284</i> 0.0231	<i>Ta27922</i> 0.0322	<i>EF</i> 0.0448	<i>Ta2291</i> 0.0326	<i>Ta2291</i> 0.0523	<i>Ta2291</i> 0.0279	<i>Ta2776</i> 0.0264	<i>EF</i> 0.0416
3	<i>Ta27922</i> 0.1011	<i>Ta35284</i> 0.0300	<i>Ta27922</i> 0.0232	<i>Ta2291</i> 0.0349	<i>Ta27922</i> 0.0468	<i>Ta54227</i> 0.0361	<i>Ta53967</i> 0.0530	<i>Ta35284</i> 0.0288	<i>Ta2291</i> 0.0269	<i>Ta2776</i> 0.0548
4	<i>Ta2776</i> 0.1011	<i>Ta2776</i> 0.0344	<i>Ta53919</i> 0.0360	<i>Ta35284</i> 0.0362	<i>Ta2776</i> 0.0480	<i>Ta53967</i> 0.0390	<i>Ta54171</i> 0.0599	<i>EF</i> 0.0316	<i>Ta53919</i> 0.0315	<i>Ta35284</i> 0.0558
5	<i>Ta53919</i> 0.1033	<i>Ta54227</i> 0.0669	<i>EF</i> 0.0397	<i>Ta2776</i> 0.0366	<i>Ta54227</i> 0.0500	<i>EF</i> 0.0419	<i>EF</i> 0.0609	<i>Ta53919</i> 0.0320	<i>Ta35284</i> 0.0367	<i>Ta53919</i> 0.0562
6	<i>Ta54171</i> 0.1033	<i>EF</i> 0.0704	<i>Ta54227</i> 0.0426	<i>EF</i> 0.0391	<i>Ta54171</i> 0.0507	<i>Ta27922</i> 0.0689	<i>Ta27922</i> 0.0675	<i>Ta27922</i> 0.0336	<i>EF</i> 0.0437	<i>Ta54227</i> 0.0696
7	<i>Ta35284</i> 0.1143	<i>Ta54171</i> 0.0748	<i>Ta53967</i> 0.0451	<i>Ta54227</i> 0.0393	<i>Ta53919</i> 0.0579	<i>Ta2776</i> 0.0707	<i>Ta54227</i> 0.0741	<i>Ta54227</i> 0.0552	<i>Ta54171</i> 0.0618	<i>Ta53967</i> 0.0879
8	<i>Ta54227</i> 0.1205	<i>Ta53967</i> 0.0767	<i>Ta2291</i> 0.0595	<i>Ta53919</i> 0.0700	<i>Ta35284</i> 0.0695	<i>Ta54171</i> 0.0750	<i>Ta35284</i> 0.0795	<i>Ta54171</i> 0.0742	<i>Ta53967</i> 0.0793	<i>Ta54171</i> 0.0984
9	<i>Ta53967</i> 0.1205	<i>Ta53919</i> 0.0772	<i>Ta54171</i> 0.0835	<i>Ta54171</i> 0.0736	<i>Ta53967</i> 0.0749	<i>Ta35284</i> 0.1123	<i>Ta53919</i> 0.0863	<i>Ta53967</i> 0.0789	<i>Ta54227</i> 0.0868	<i>Ta27922</i> 0.1164
BCTG ¹	<i>Ta2291</i> <i>Ta2776</i>	<i>Ta2291</i> <i>Ta27922</i>	<i>Ta2776</i> <i>Ta35284</i>	<i>Ta53967</i> <i>Ta27922</i>	<i>Ta2291</i> <i>EF</i>	<i>Ta53919</i> <i>Ta2291</i>	<i>Ta2776</i> <i>Ta2291</i>	<i>Ta2776</i> <i>Ta2291</i>	<i>Ta27922</i> <i>Ta2776</i>	<i>Ta2291</i> <i>EF</i>

¹BCTG: best combination of genes.

BestKeeper analysis

The stability of candidate gene expression varied among different data sets and indexes (Table S3). Based on the SD value, *EF* was the gene most frequently ranked first on the lists of stability, followed by *Ta2291*, which was expressed stably in the complete set of samples, flag leaf sheath, rachis, paleo, and the lemma. Other than that, the abundance of *Ta27922*, *Ta53919*, and *Ta54227* transcripts remained more constant in the penultimate leaf, peduncle, and the glume respectively. According to the r value, *Ta2776* (penultimate leaf sheath, flag leaf sheath, palea and lemma, complete set of samples) and *Ta2291* (peduncle, glume, penultimate leaf, and the kernel) all showed the least variation in four subsets of samples. *Ta27922* and *Ta53967* were expressed at a constant level in two different tissues, namely the rachis and the flag leaf. Furthermore, significant correlations of the most stable gene with other genes in each data set were observed except for *Ta54227*, which showed no good correlation ($r = 0.463$) at a lower level of significance ($P = 0.295$) in the glume.

Evaluations of identified reference genes

The expression dynamics of three *NAM* genes were analyzed to show the effect of different reference genes on the calculated relative expression. As shown in Figure 4, similar patterns of *TaNAM-B1* expression (Figure 4B) were obtained after normalization with either single or multiple reference genes (Figure 5), which largely agreed with the former results but with a minor difference (e.g., a decrease during the first five days after anthesis). Inconsistencies were frequently observed when different single reference genes were utilized to analyze the expression dynamics of *TaNAM-A1* (Figure 4A) and *TaNAM-B2* (Figure 4C). Specifically, abundance of the *TaNAM-B2* transcript did not continuously rise but decreased after 15 DAA when *Ta53967* (NormFinder, and BestKeeper based on r, flag leaf) and *Ta2291* (BestKeeper

based on SD, flag leaf) were used as control genes. In contrast to results obtained in a previous study, expression of *TaNAM-A1* and *TaNAM-B2* declined earlier (since 10 DAA) when *Ta2776* (BestKeeper based on r, all samples) was applied. Furthermore, a slight decrease in the level of the *TaNAM-A1* transcript was detected at 15 DAA using *EF* (NormFinder and BestKeeper based on SD, all samples) to calculate the normalization factor. Notably, patterns of *TaNAM-A1* and *TaNAM-B2* expression were consistent with those obtained previously when normalized against *Ta27922* (geNormPlus, all samples).

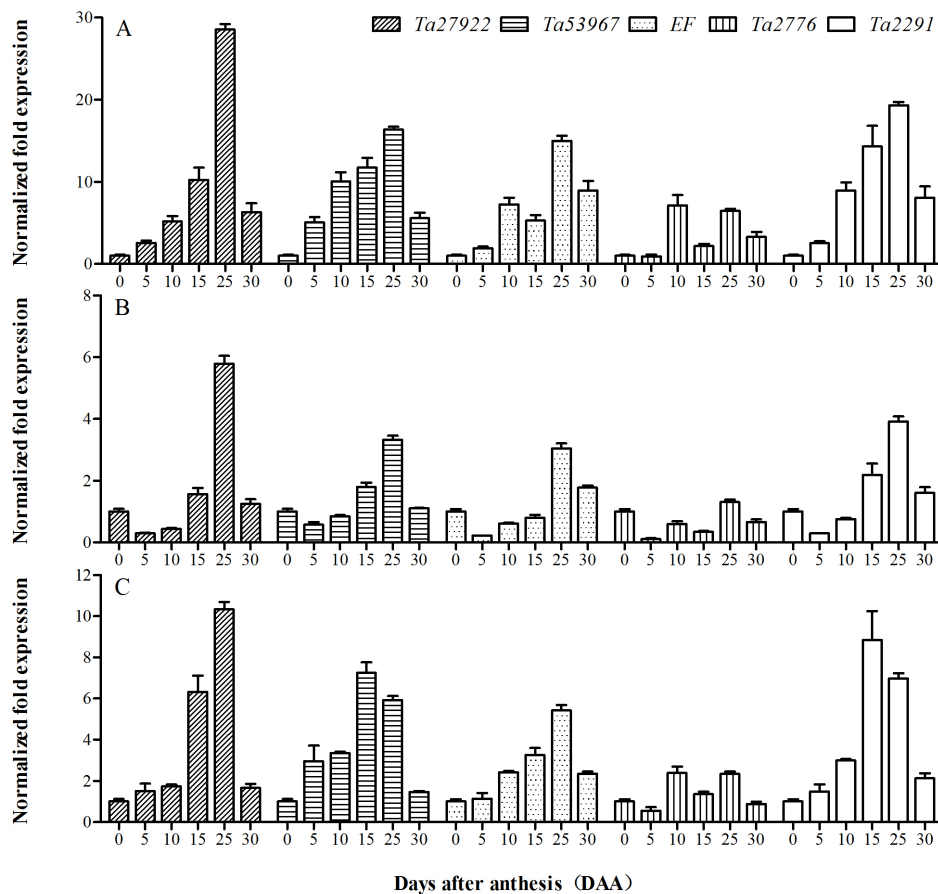


Figure 4. Expression dynamics of three *NAM* genes in the flag leaf using different single reference genes. **A.** *TaNAM-A1*; **B.** *TaNAM-B1*; **C.** *TaNAM-B2*.

Multiple reference genes exhibited higher reliabilities than did single ones since the results were universally consistent with those obtained previously (Figure 5), in which those identified by geNormPlus and NormFinder for the flag leaf (Figure 5B and D) outperformed the results for the complete set of samples (Figure 5A and C).

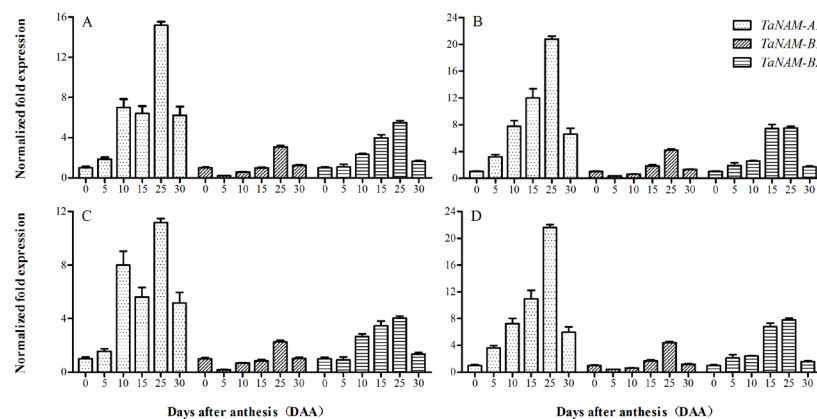


Figure 5. Expression dynamics of three *NAM* genes in the flag leaf using different multiple reference genes. **A.** *Ta27922+Ta2291+Ta2776 + EF* (geNormPlus, all samples); **B.** *Ta53967+Ta27922+Ta2291* (geNormPlus, flag leaf); **C.** *Ta2291+Ta2776* (NormFinder, all samples); **D.** *Ta53967+Ta27922* (NormFinder, flag leaf).

DISCUSSION

The identification of diagnostic genes is required to elucidate the signaling pathways that underlie kernel development, and the dynamics of gene expression are a valuable resource to achieve this goal. However, the accurate and reliable monitoring of gene expression relies on the correct use of single or multiple reference genes to eliminate variation. To our knowledge, this study represents the first effort to comprehensively identify reference genes with highly uniform expression during wheat grain filling. The results showed that identical results were not produced by all programs, and moreover, either the selected single or multiple reference genes differed among data sets. In practice, gene expression is heterogeneous and complex, and to some extent depends on the tissue, developmental stage, and environmental condition studied (Tenea et al., 2011). In addition to variations in experimental conditions, the diverse results were most likely caused by the different mathematical algorithms employed or distinct development processes underway in tested tissues during the sampling period in which late developmental stages were included for some tissues or cell types.

Several algorithms and freely available software, from simple to advanced, have been developed to assist in identifying stably expressed reference genes; however, no universally and univocally accepted methods exist. Furthermore, considering the different results that each algorithm may deliver, the choice of a suitable method is a crucial yet difficult task (Mac-coux et al., 2007; Hibbeler et al., 2008).

geNormPlus calculates variation in the expression of a candidate gene in a pair-wise manner; therefore, it is sensitive to the presence of co-regulated genes and tends to select the genes with the highest similar but possibly unstable expression (Hibbeler et al., 2008). Although the absence of co-regulations among some candidate genes has been previously confirmed (Paolacci et al., 2009), particularly noteworthy was the possibility that *Ta27922* (Cyclin-T1 protein) and *Ta54227* [Cell division control protein 48 homolog C (*cdc48*)] might be involved in related biological processes. It has been speculated that *Ta54227* is a cyclin-dependant kinase in *Saccharomyces cerevisiae*; however, it was reported that *cdc48* was specifically associated with G1 cyclin2 but not Cyclin-T1 (Archambault et al., 2004). Since

co-regulated genes might show similar expression patterns (Yeung et al., 2004), the expression dynamics of their homologs (AT4G19600 for *Ta27922* and AT5G03340 for *Ta54227*) in *Arabidopsis thaliana* were assessed by exploiting the EST expression profiles in the GEO datasets (<http://www.ncbi.nlm.nih.gov/geoprofiles>). In most cases, no obvious similarity was observed. Considering that the relationships between co-regulated genes are complex (Yu et al., 2003), each data set underwent further verification. It has been shown that the remaining candidate genes were placed in a different order after the high-ranking stable gene defined by geNorm was eliminated, if there was no co-regulation (Ransbotyn and Reusch, 2006). The same changes were also detected in all subsets of samples (data not shown); therefore, no evident relationship among candidate genes existed.

The most stable reference gene defined by BestKeeper should exhibit high similarity in expression with all other candidate genes; hence the *r* value was the default standard for the evaluation of expression stability in most studies. However, the SD value was also utilized for the preliminary assessment of each candidate gene, since the *r* value depends on sets of chosen candidate genes and is invalidated by heterogeneous variance between genes (Pfaffl et al., 2004). In the present study, both SD and *r* were applied. The results showed that uniform outcomes were not obtained by either of two parameters in all subsets of samples except for the kernel. Furthermore, less variable genes did not always show higher correlations with other genes. It is therefore proposed that if any of these values was used alone, it would probably result in biases or even errors.

Normalization against the geometric mean of the relative expression of multiple reference genes has been preferred as it outperforms the use of a single gene in terms of accuracy and having a low error rate (Vandesompele et al., 2002; Andersen et al., 2004; Gu et al., 2011). Compared with multiple reference genes, 25% of studies using a single reference gene resulted in a 3.0-fold error and 10% had a 6.4-fold error (Vandesompele et al., 2002). The multiple reference genes (identified by geNormPlus and NormFinder) were also estimated by the outcomes of BestKeeper. In the complete set of samples, SD values of *Ta54227* (1.05), *Ta53919* (1.12), *Ta53967* (1.18), and *Ta54171* (1.03) were beyond the threshold ($SD < 1$), in which *Ta54171* ($0.560 < r < 0.711$) also showed lower pair-wise correlations with other genes. Therefore, these candidate genes were all excluded, and *r* values for the remaining genes increased (from $0.883 < r < 0.941$ to $0.914 < r < 0.945$), and the significant correlations between each gene of multiple reference genes (as defined by geNormPlus and NormFinder) and BI proved that *Ta27922* (0.930), *Ta2291* (0.945), *Ta2776* (0.930), and *EF* (0.917) were suitable multiple reference genes for the complete set of samples. In this way, multiple reference genes were reassessed, and all showed higher correlations with BI (data not shown). As expected, multiple reference genes outperformed the single reference genes when they were used to calculate normalization factors to assess the expression of three *NAM* genes.

To summarize, *Ta27922* can be utilized as the single reference gene for gene expression analysis conducted during wheat grain filling. Given the limit of NormFinder in gene numbers of multiple reference genes, the multiple genes determined by geNormPlus in individual tissues can also be used for further quantification analysis.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by grants from the “twelfth five-year plan” of the national science and technology support project for the rural development in China (#2011AA100501) and the special project of national modern wheat industry technology system construction (#CARS-3-2-47).

[Supplementary material](#)

REFERENCES

- Andersen CL, Jensen JL and Ørntoft TF (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64: 5245-5250.
- Archambault V, Chang EJ, Drapkin BJ, Cross FR, et al. (2004). Targeted proteomic study of the cyclin-Cdk Module. *Mol. cell* 14: 699-711.
- Argyropoulos D, Psallida C and Spyropoulos CG (2005). Generic normalization method for real-time PCR. Application for the analysis of the mannanase gene expressed in germinating tomato seed. *FEBS J.* 273: 770-777.
- Barsalobres-Cavallari CF, Severino FE, Maluf MP and Maia IG (2009). Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. *BMC Mol. Biol.* 10: 1.
- Bustin SA (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* 29: 23-39.
- Czechowski T, Stitt M, Altmann T, Udvardi MK, et al. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* 139: 5-17.
- Dheda K, Huggett JF, Chang JS, Kim LU, et al. (2005). The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal. Biochem.* 344: 141-143.
- Giménez MJ, Pistón F and Atienza SG (2011). Identification of suitable reference genes for normalization of qPCR data in comparative transcriptomics analyses in the Triticeae. *Planta* 233: 163-173.
- Ginzinger DG (2002). Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Exp. Hematol.* 30: 503-512.
- Gu YR, Li MZ, Zhang K, Chen L, et al. (2011). Evaluation of endogenous control genes for gene expression studies across multiple tissues and in the specific sets of fat- and muscle-type samples of the pig. *J. Anim. Breed. Genet.* 128: 319-325.
- Hellemans J, Mortier G, Paepe AD, Speleman F, et al. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 8: R19.
- Hibbeler S, Scharsack JP and Becker S (2008). Housekeeping genes for quantitative expression studies in the three-spined stickleback *Gasterosteus aculeatus*. *BMC Mol. Biol.* 9: 18.
- Johnson ML, Redmer DA and Reynolds LP (1995). Quantification of lane-to-lane loading of poly(A) RNA using a biotinylated oligo(dT) probe and chemiluminescent detection. *Biotechniques* 19: 712-715.
- Kriegova E, Arakelyan A, Fillerova R, Zatloukal J, et al. (2008). PSMB2 and RPL32 are suitable denominators to normalize gene expression profiles in bronchoalveolar cells. *BMC Mol. Biol.* 9: 69.
- Kumar V, Sharma R, Trivedi PC, Vyas GK, et al. (2011). Traditional and novel references towards systematic normalization of qRT-PCR data in plants. *Aust. J. Crop Sci.* 5: 1455-1468.
- Liu W, Zhao X, Zhang P, Mar TT, et al. (2013). A one step real-time RT-PCR assay for the quantitation of *Wheat yellow mosaic virus* (WYMV). *Virol. J.* 10: 173.
- Livak KJ and Schmittgen TD (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25: 402-408.
- Long XY, Wang JR, Ouellet T, Rocheleau H, et al. (2010). Genome-wide identification and evaluation of novel internal control genes for Q-PCR based transcript normalization in wheat. *Plant Mol. Biol.* 74: 307-311.
- Maccoux LJ, Clements DN, Salway F and Day PJ (2007). Identification of new reference genes for the normalisation of canine osteoarthritic joint tissue transcripts from microarray data. *BMC Mol. Biol.* 8: 62.
- Maroufi A, Bockstaele EV and Loose MD (2010). Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. *BMC Mol. Biol.* 11: 15.
- Paolacci AR, Tanzarella OA, Porceddu E and Ciaffi M (2009). Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol. Biol.* 10: 11.
- Pfaffl MW, Tichopad A, Prgomet C and 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.
- Ransbotyn V and Reusch TBH (2006). Housekeeping gene selection for quantitative real-time PCR assays in the seagrass *Zostera marina* subjected to heat stress. *Limnol. Oceanogr-Meth.* 4: 367-373.
- Scholtz JJ and Visser B (2013). Reference gene selection for qPCR gene expression analysis of rust-infected wheat. *Physiol. Mol. Plant Pathol.* 81: 22-25.
- Solanas M, Moral R and Escrich E (2001). Unsuitability of using ribosomal RNA as loading control for Northern blot analyses related to the imbalance between messenger and ribosomal RNA content in rat mammary tumors. *Anal. Biochem.* 288: 99-102.
- Tenea GN, Bota AP, Raposo FC and Maquet A (2011). Reference genes for gene expression studies in wheat flag leaves grown under different farming conditions. *BMC Res. Notes* 4: 373.
- Thellin O, Zorzi W, Lakaye B, Borman BD, et al. (1999). Housekeeping genes as internal standards: use and limits. *J. Biotechnol.* 75: 291-295.
- Uauy C, Distelfeld A, Fahima T, Blechl A, et al. (2006). A NAC gene regulation senescence improves grain protein, zinc, and iron content in wheat. *Science* 314: 1298-1301.
- Vandesompele J, Preter KD, Pattyn F, Poppe B, et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3: research0034.
- Wang E, Wang J, Zhu X, Hao W, et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nature Genet.* 40: 1370-1374.
- Warner JR (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* 24: 437-440.
- Yeung KY, Medvedovic M and Bumgarner RE (2004). From co-expression to co-regulation: how many microarray experiments do we need? *Genome Biol.* 5: R48.
- Yu H, Luscombe NM, Qian J and Gerstein M (2003). Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet.* 19: 422-427.