



Genome-Wide Identification and Evaluation of Reference Genes for Quantitative RT-PCR Analysis during Tomato Fruit Development

Yuan Cheng^{1†}, Wuying Bian^{2†}, Xin Pang³, Jiahong Yu¹, Golam J. Ahammed⁴, Guozhi Zhou¹, Rongqing Wang¹, Meiyong Ruan¹, Zhimiao Li¹, Qingjing Ye¹, Zhuping Yao¹, Yuejian Yang¹ and Hongjian Wan^{1*}

¹ State Key Laboratory Breeding Base for Zhejiang Sustainable Pest and Disease Control, Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, ² Zhejiang Agricultural Technology Extension Center, Hangzhou, China, ³ Suzhou Polytechnic Institute of Agriculture, Suzhou, China, ⁴ Department of Horticulture, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:

Hinanit Koltai,
Agricultural Research Organization,
Israel

Reviewed by:

Panagiotis Kalaitzis,
Mediterranean Agronomic Institute of
Chania, Greece
Antonio Ferrante,
Università degli Studi di Milano, Italy

*Correspondence:

Hongjian Wan
wanhongjian@sina.com

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 19 April 2017

Accepted: 03 August 2017

Published: 29 August 2017

Citation:

Cheng Y, Bian W, Pang X, Yu J, Ahammed GJ, Zhou G, Wang R, Ruan M, Li Z, Ye Q, Yao Z, Yang Y and Wan H (2017) Genome-Wide Identification and Evaluation of Reference Genes for Quantitative RT-PCR Analysis during Tomato Fruit Development.
Front. Plant Sci. 8:1440.
doi: 10.3389/fpls.2017.01440

Gene expression analysis in tomato fruit has drawn increasing attention nowadays. Quantitative real-time PCR (qPCR) is a routine technique for gene expression analysis. In qPCR operation, reliability of results largely depends on the choice of appropriate reference genes (RGs). Although tomato is a model for fruit biology study, few RGs for qPCR analysis in tomato fruit had yet been developed. In this study, we initially identified 38 most stably expressed genes based on tomato transcriptome data set, and their expression stabilities were further determined in a set of tomato fruit samples of four different fruit developmental stages (Immature, mature green, breaker, mature red) using qPCR analysis. Two statistical algorithms, geNorm and Normfinder, concordantly determined the superiority of these identified putative RGs. Notably, SIFRG05 (Solyc01g104170), SIFRG12 (Solyc04g009770), SIFRG16 (Solyc10g081190), SIFRG27 (Solyc06g007510), and SIFRG37 (Solyc11g005330) were proved to be suitable RGs for tomato fruit development study. Further analysis using geNorm indicate that the combined use of SIFRG03 (Solyc02g063070) and SIFRG27 would provide more reliable normalization results in qPCR experiments. The identified RGs in this study will be beneficial for future qPCR analysis of tomato fruit developmental study, as well as for the potential identification of optimal normalization controls in other plant species.

Keywords: qPCR analysis, normalization, reference gene (RG), tomato, fruit development

INTRODUCTION

Tomato (*Solanum lycopersicum*) is an economically important horticultural crop in terms of production, flavor, and nutritional value of fruits. During the course of development and ripening, tomato fruits undergo a number of physiological and biochemical processes that bring forth the overall changes in fruit size, color, texture, and aroma (Klee and Giovannoni, 2011; Ruiz-May and Rose, 2012). Moreover, tomato is considered as an important model for genetic and molecular studies, partly due to its typical climacteric fruit property (Colombiet et al., 2016). A number of studies had been carried out to improve agronomic traits of tomato fruits, including size, pigment content, and flavor substances focusing on the metabolic and regulatory networks (Klee and Giovannoni, 2011; Ruiz-May and Rose, 2012; Tieman et al., 2017). Recent developments in genomic

resources and bioinformatics tools (e.g., Genome-wide association study, GWAS) have enabled rapid elucidation of the complicated biological processes that occur during fruit development. Moreover, relative gene expression profiles during fruit development provide valuable clues for understanding the biological functions of the corresponding genes. So far, quantitative real-time PCR (qPCR) is considered as one of the most efficient tools for the measurement of transcript abundance of a gene due to its high accuracy, sensitivity, and reproducibility (Ginzinger, 2002; Bustin and Nolan, 2004; Gachon et al., 2004; Bustin et al., 2005).

In qPCR experiments, the reliability of the results predominantly depends on the appropriateness of RGs used for normalization, which should be stably expressed under the given experimental conditions (Pfaffl, 2004; Huggett et al., 2005). Highly stable expression of RGs could effectively remove non-biological variations including the difference in amounts, variability in enzymatic efficiency of reverse transcriptase, and sample differences in the overall transcriptional activity (Suzuki et al., 2000; Bustin et al., 2005; Huggett et al., 2005; Exposito-Rodriguez et al., 2008; Gutierrez et al., 2008). Generally speaking, an ideal RG should be a gene that is stably expressed under various experimental conditions or among different tissues (Czechowski et al., 2005; Huggett et al., 2005; Exposito-Rodriguez et al., 2008; Dekkers et al., 2012; Wang et al., 2012). Housekeeping genes (HKGs) encoding, e.g., GAPDH, Actin, UBI, and 18 sRNAs, are usually regarded as suitable normalization controls (Stürzenbaum and Kille, 2001). However, some previous studies reported that the transcription of several HKGs can be fluctuated considerably under certain conditions (Czechowski et al., 2005; Jain et al., 2006; Gutierrez et al., 2008; Jian et al., 2008; Jarosova and Kundu, 2010), which illustrates the importance of systematic identification or validation of optimal RGs in order to avoid inaccurate results (Gutierrez et al., 2008; Guenin et al., 2009). In practice, the expression levels of most RGs are proved to be dependent on the specific conditions, including experimental treatments, tissue types, or developmental stages (Czechowski et al., 2005; Jian et al., 2008; Jarosova and Kundu, 2010). Hence, no single RG is widely applicable for different experimental conditions. Systematic evaluation of RGs must be conducted on each qPCR experiment before their use (Bustin et al., 2009; Jacob et al., 2013). Furthermore, it has been well-recognized that in some cases, one single RG may not be adequate for reliable normalization in gene expression analysis (Yoo et al., 2009; Cassan-Wang et al., 2012). To date, some common statistical algorithms, including geNorm (Vandesompele et al., 2002), NormFinder (Andersen et al., 2004), and Bestkeeper (Pfaffl, 2004), have been developed to determine the expression stabilities of RGs, which effectively simplify the selection of appropriate RGs for qPCR analysis.

Over the decades, a good number of stably expressed RGs have specifically been identified for normalization in the fruits of several fruit crops, such as papaya (Zhu et al., 2012), blueberry (Die and Rowland, 2013), and watermelon (Kong et al., 2015). For tomato, although suitable RGs have been identified under different experimental conditions including biotic/abiotic

stresses (Løvdaal and Lillo, 2009; Mascia et al., 2010) and various tissues of different developmental stages (Suzuki et al., 2000; Dekkers et al., 2012), very limited number of RGs in tomato fruit have been characterized so far (Coker and Davies, 2003; Baldassarre et al., 2015). Moreover, we noticed that most studies involving RG identification, including those relevant to tomato fruit developmental studies, were mainly based on the evaluation or validation of some already known candidate RGs (Mostly HKGs), which are convenient for implementation but also greatly limit the choice of best RGs. With the availability of tomato genome sequence and subsequent transcriptome data (SGN:Sol genomics network, <https://solgenomics.net/>; TFGD: Tomato functional genomics database, <http://ted.bti.cornell.edu/>), our study was aimed to identify some novel RGs for qPCR analysis of tomato fruit development within the entire genome level.

In this study, we initially evaluated the expression stabilities of all the tomato (*S. lycopersicum* L.) genes during various fruit developmental stages based on the RNA-seq data. A total of 38 novel genes stably expressed were identified as putative RGs and were further evaluated through qPCR analysis. Moreover, using two different statistical algorithms (geNorm and Normfinder), five optimal RGs were identified as optimal RGs for normalization during different stages of tomato fruit development. Furthermore, we also found that the combined use of two top-ranked RGs (*SIFRG03* and *SIFRG27*) would potentially improve the accuracy of the qPCR results. Thus, based on the analysis of the entire tomato genome database, we comprehensively identified and evaluated the optimal RGs through large-scale biological information mining and qPCR techniques. These results not only provide useful RG resources for accurate gene expression studies in tomato fruit, but also shed light on the RG identification in fruit developmental study of other plant species.

MATERIALS AND METHODS

Collection and Evaluation of the Previously Reported RGs

Firstly, the potential tomato RGs reported in previous studies (Coker and Davies, 2003; Exposito-Rodriguez et al., 2008; Løvdaal and Lillo, 2009; Mascia et al., 2010; Müller et al., 2015) were selected to evaluate the expression stabilities during different stages of tomato (*S. lycopersicum* L.) fruit development based on the RNA-seq data downloaded from the TFGD (<http://ted.bti.cornell.edu/>). Furthermore, the orthologous genes of 11 potential RGs for watermelon fruit development (Kong et al., 2015) were identified. All the potential RGs selected from RNA-seq data were evaluated for expression stability. Details including accession number, gene locus, gene description, and RNA-seq values were listed in **Table 1** and **Supplementary Table 1**. The corresponding gene sequences of these candidate RGs were collected from NCBI (National Center for Biotechnology Information: <https://www.ncbi.nlm.nih.gov/>), SGN (<https://solgenomics.net/>), TFGD (<http://ted.bti.cornell.edu/>) (Fraser et al., 1994), and CuGenDB (<http://www.icugi.org/>). Through Blastn search, the orthologous genes (E-value was set at $1e^{-5}$) were collected in tomato genome

TABLE 1 | Previously reported RGs in tomato or watermelon.

Gene symbol	Accession number	Gene locus	Gene description	Sequence identities (%)
GAPDH	U97257	Solyc05g014470	Glyceraldehyde 3-phosphate dehydrogenase	99
EF1 α	X53043	Solyc06g005060	Elongation factor 1-alpha	98
RPL8	X64562	Solyc10g006580	Ribosomal protein L2	99
DNAJ	AF124139	Solyc11g006460	DNAJ-like protein	96
TUA	AC122540	Solyc09g074220	Musmusculus BAC clone	99
Actin	Q96483	Solyc11g005330	Actin-51	99
TAPG4	U70481.1	Solyc12g096750	Abscission polygalacturonase	99
CHI9	NM001247474	Solyc10g055810	Chitinase (CHI9)	99
ACT	BT013707	Solyc03g078400	Actin	100
CYP	AK326854	Solyc01g108340	Cyclophilin Elongation factor 1- α	100
GAPDH	U93208	Solyc03g111010	Glyceraldehyde 3-phosphate dehydrogenase	98
UBI	X58253	Solyc01g056940	Ubiquitin 3	99
UK	AK322232	Solyc01g088480	Uridylate kinase	100
18S rRNA	X51576	Solyc11g051210	18S rRNA	97
TBP	SGN-U329249	Solyc01g028930	TATA binding protein 2	100
TIP41	SGN-U321250	Solyc10g049850	TIP41-like family protein	99
SAND	SGN-U316474	Solyc03g115810	SAND family protein	99
CAC	SGN-U314153	Solyc08g006960	Clathrin adaptor complexes medium subunit	100
Expressed	SGN-U346908	Solyc07g025390	Expressed sequence	98
PK	TC123837	Solyc07g066610	Phosphoglycerate kinase	100
CBCP	TC115713	Solyc01g028810	Chaperonin-60 beta chain prec	100
DNAJ	TC123959	Solyc04g009770	DNAJ-like protein	100
AT5G	TC123964	Solyc06g072120	AT5g	100
CICAC	Cla016178	Solyc06g061150	Clathrin adaptor complex subunit	95
CIPP2A	Cla021905	Solyc01g011340	Protein phosphatase 2A regulatory subunit A	82
CIRAN	Cla012277	Solyc05g023800	A member of RAN GTPase gene family	82
CIRPS15	Cla021565	Solyc06g053820	Cytosolic ribosomal protein S15	84
CISAND	Cla001870	Solyc03g115810	SAND family protein	80
CITBP2	Cla011119	Solyc01g028930	TATA binding protein 2	82
CITIP41	Cla016074	Solyc10g049850	TIP41-like family protein	86
CITUA5	Cla003129	Solyc02g087880	Alpha tubulin 5	80
CITUB	Cla022418	Solyc04g081490	β -tubulin	84
CIUPL7	Cla017746	Solyc09g008700	Ubiquitin-protein ligase 7	80
CI18SrRNA	Cla010159	Solyc05g007050	18SrRNA	82
		Solyc10g006100	Transcription initiation factor TFIID subunit 6	100
		Solyc07g062920	Genomic DNA chromosome 5 TAC clone K19P17	100
		Solyc01g111780	Importin beta-2 subunit	100
		Solyc06g051420	PHD finger family protein	100
		Solyc12g057120	Subunit VIb of cytochrome c oxidase	100
		Solyc01g009290	Polyribonucleotide 5'-hydroxyl-kinase Clp1	100
		Solyc09g018730	Ubiquitin carboxyl-terminal hydrolase family 1	100
		Solyc02g088110	Polypyrimidine tract-binding protein-like	100
		Solyc08g060860	P1 clone MSJ11	100
		Solyc09g009640	U6 snRNA-associated Sm-like protein LSm7	100
		Solyc04g015370	Acyl carrier protein	100
		Solyc08g005140	Serine/threonine-protein kinase BUD32	100
		Solyc02g062920	Splicing factor U2AF large subunit	100
		Solyc10g076910	ATP-dependent RNA helicase-like protein	100
		Solyc03g121980	Developmentally-regulated GTP-binding protein 2	100
		Solyc01g097140	Dual-specificity tyrosine-phosphatase CDC25	100
		Solyc07g007040	Zinc finger CCCH-type protein	100

(Continued)

TABLE 1 | Continued

Gene symbol	Accession number	Gene locus	Gene description	Sequence identities (%)
		Solyc06g069310	Nuclear transcription factor Y subunit B-6	100
		Solyc03g078020	Peptide chain release factor 1	100
		Solyc10g078180	cyclin gene family	100
		Solyc02g089230	DSBA oxidoreductase family protein	100
		Solyc06g036720	HLA-B associated transcript 3	100
		Solyc01g109620	Unkown	100
		Solyc07g064510	Polyadenylate-binding protein 2	100
		Solyc11g071930	DnaJ homolog subfamily C member 8	100
		Solyc06g084000	Heterogeneous nuclear ribonucleoprotein K	100
		Solyc04g009230	Mitosis protein dim1	100
		Solyc06g073870	DNA-directed RNA polymerase II subunit RPB4	100
		Solyc09g055760	T-snare	100
		Solyc12g005780	TraB family protein	100
		Solyc04g008610	Histone acetyltransferase	100
		Solyc04g015300	Alpha/beta hydrolase	100
		Solyc10g005800	CWC15 homolog	100
		Solyc12g021130	3-beta-hydroxysteroid dehydrogenase-like	100
		Solyc01g079330	RNA helicase DEAD3	100
		Solyc07g041550	RNA polymerase-associated protein Ctr9 homolog	100
		Solyc03g059420	Sister chromatid cohesion 2	100
		Solyc11g071950	Unknown	100
		Solyc12g099570	Heat shock factor binding protein 2	100
		Solyc10g044900	CASTOR protein	100
		Solyc10g084270	Importin alpha-2 subunit	100
		Solyc06g016750	Transcription factor	100
		Solyc02g092380	Peptidyl-prolylcis-trans isomerase	100
		Solyc05g052960	BTB/POZ domain containing protein	100
		Solyc06g009860	Mercaptopyruvate sulfurtransferase-like protein	100
		Solyc10g008950	Guanylate-binding protein 10	100
		Solyc10g055450	Ubiquitin-protein ligase 4	100
		Solyc05g006580	Unknown	100
		Solyc03g121310	RWD domain-containing protein	100
		Solyc09g010180	Unknwon	100

Detailed information was derived from NCBI (<https://www.ncbi.nlm.nih.gov/>), TFGD (<https://www.ncbi.nlm.nih.gov/>), CuGenDB (<http://www.icugi.org/>), and SGN (<https://solgenomics.net/>). The three redundant genes were labeled in bold.

databases (SGN) (Table 1). The Reads Per Kilobase Million (RPKM) value of each reported RG was collected from RNA-seq data, and the average expression (AVE) and standard deviation (SD) values in different fruit developmental stages of fruit were calculated (Supplementary Table 1; Table 1). Relative expression level per gene was calculated through dividing the expression value of each fruit development stage by AVE (Figure 1A). CV (co-efficient variation) value of each gene during fruit development was calculated as the ratio of the SD to the AVE (Supplementary Table 3).

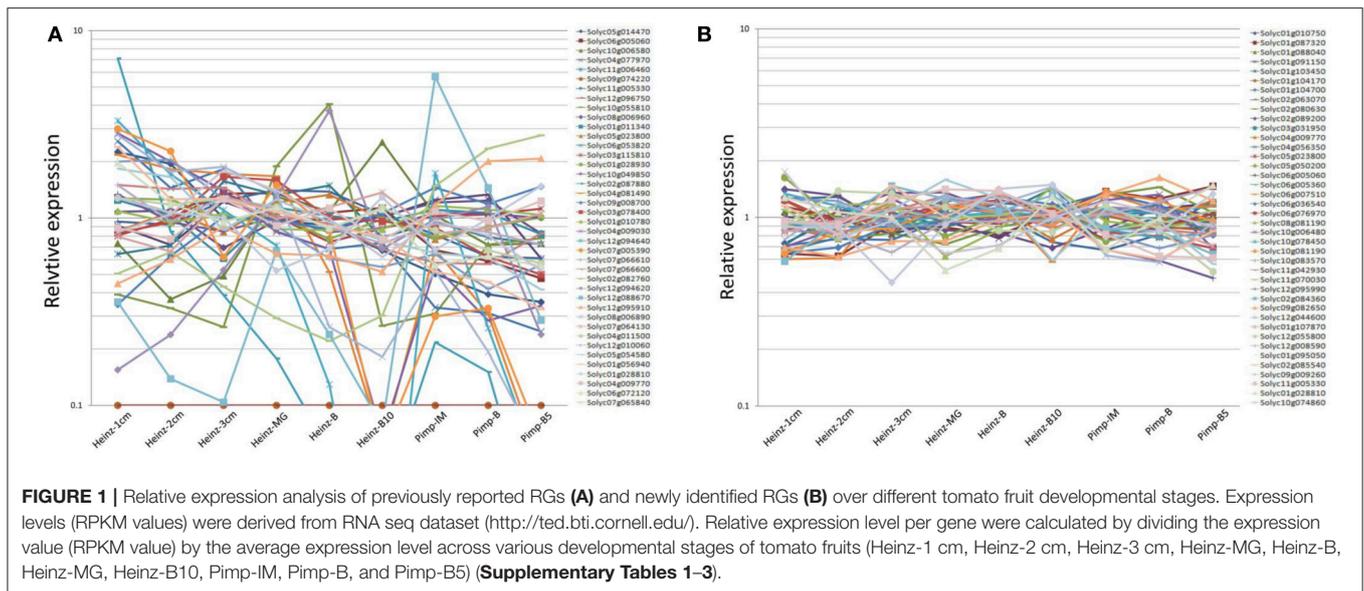
Identification of Stably Expressed Genes through Mining Fruit RNA-Seq Data Set

Through mining the entire set of RNA-seq data file (TFGD), genes with medium expression ($200 < \text{RPKM value} <$

$2,000$) at all stages of fruit development were selected (Supplementary Table 2). The CV value of each gene was calculated (Supplementary Table 2) and the genes with $\text{CV} < 0.35$ were chosen as putative RGs in the following qPCR analysis (Table 2; Supplementary Table 3).

Plant Materials

The tomato (*S. lycopersicum* L.) inbred line “S14” was used in the current experiments. Generally, the fruits of this genotype become completely mature (red flash) 45 days after pollination. One month old tomato seedlings were transplanted to the greenhouse at Zhejiang Academy of Agricultural Sciences, Hangzhou, China (east longitude $120^{\circ}2''$, north latitude $30^{\circ}27''$) on August 5th (late summer), 2016. Field management was implemented following the standard commercial practices. Tomato fruits were harvested at four developmental stages



(Immature: IM, Mature Green: MG, Breaker: B, and Mature Red: MR) from October to November (autumn), 2016. Three fruits were randomly collected at each sampling point, with each of them representing a biological replicate. All the samples were frozen in liquid nitrogen and stored at -80°C for subsequent experiments.

Total RNA Isolation and cDNA Synthesis

Total RNA from collected samples was isolated using TRIZOL reagent according to the manufacturer's protocol (Tiangen, Beijing, China) as described previously (Cheng et al., 2016). The concentration and purity of extracted RNA were measured using a BioDropULite spectrophotometer (Biochrom, England), and RNA samples with $A260/A280 > 1.8$ and $A260/A230 > 2.0$ (indicating good RNA quality) were used for following experiments. All RNA samples were adjusted to the same concentration in order to ensure that the RNA input was homogenized for subsequent reverse transcription reactions using a mix of random primers. Then, according to the manufacturer's instruction (TIANGEN, Beijing, China), genomic DNAs (gDNA) were eliminated from the RNA samples and single-stranded cDNAs were synthesized.

Primer Design and qPCR Analysis

Gene-specific primers were designed using Real-time PCR (TaqMan) primer design (<https://www.genscript.com/>) as listed in Table 2. The qPCR experiments was performed in a 96-well plate using SYBR Green-based PCR assay. A 20 μL reaction mixture [6 μL diluted cDNA (10 ng), 10 μL SYBR Green PCR Master Mix (Invitrogen, USA), 250 nM of each primer and, 0.1 μL ROX] was subjected to the following procedure: 10 min at 94°C , 30 cycles of 45 s at 94°C , 45 s at 55°C , and 1 min at 72°C following a 7 min extension at 72°C (ABI real time PCR system, USA). Three technical duplications performed for all of the RGs. The amplified products were resolved by 1.5% agarose gel

electrophoresis (Supplementary Figure 1). The melting curves were created and exhibited for all the investigated qPCR products in the qPCR experiments (Supplementary Figure 2).

Evaluation of RG Expression Stability Using geNorm and Normfinder

The expression levels of the detected genes were obtained through the qPCR analysis and the results were demonstrated as Ct values (Supplementary Table 4). The amplification efficiency (E) and correlation coefficient (R^2) for each gene were calculated using the standard curve method by amplifying the 10-fold serial dilution of cDNA samples. The amplification efficiency (E) was calculated with the formula: $E = (10^{-1/\text{slope}} - 1)$. The geNorm and NormFinder software packages were used to evaluate the gene expression stability as described before in this study (Vandesompele et al., 2002; Andersen et al., 2004). The geNorm applet not only provides a measure of gene expression stability value (M), but also creates pairwise variation values (V) to determine the minimum number of RGs required for reliable normalization, no additional genes were required when the pairwise variation ($V_{n/n+1}$) was lower than 0.15 (Vandesompele et al., 2002). The NormFinder measured the variations across groups and determine the expression stability of each tested gene (Andersen et al., 2004). Lower stability values (M) in both geNorm and Normfinder implied the higher expression stability of the genes. The stability values (M) obtained from geNorm and NormFinder were listed in Supplementary Table 5.

RESULTS

Evaluation of Previously Reported RGs during Tomato Fruit Development

In our study, tomato (*S. lycopersicum* L.) genome database in the SGN (<https://solgenomics.net/>) and transcriptome data derived from the TFGD (<http://ted.bti.cornell.edu/>) were used for

TABLE 2 | Description of the newly identified RGs, primer sequences, and amplification characteristics.

Gene ID	Gene locus ^a	Gene description	Primer sequence (5'–3')	Product size (bp)	E	R ²
<i>SIFRG01</i>	Solyc06g005360	Actin depolymerizing factor 3	F: TGACAAAGGAAAGTGTCCCA R: ATCTGAATTCGGTCCAGCTC	133	0.76	0.994
<i>SIFRG02</i>	Solyc01g088040	Unknown	F: GTTGATGAAGGGAGAGCCAT R: GAGAGTAGATCACGGCAGCA	95	0.78	0.998
<i>SIFRG03</i>	Solyc02g063070	Unknown	F: GGCTGAAGTGGCTCCTACTC R: TTTTCGCAAGGTTACAAGCAC	109	0.77	0.999
<i>SIFRG04</i>	Solyc12g095990	RNA helicase DEAD40	F: ACTCACGGAGACATGGATCA R: TCAATACCACGAGCAAGGAG	107	0.78	0.999
<i>SIFRG05</i>	Solyc01g104170	Ankyrin repeat domain-containing protein	F: CAAAGGTTGATGCATTGGAC R: AGTTACAGCGGCTCCATTCT	110	1.05	0.958
<i>SIFRG06</i>	Solyc11g070030	Unknown	F: CCACACGACACTTCCATCTC R: TCAGCCGGTTTAGATATCCC	111	0.90	1
<i>SIFRG07</i>	Solyc01g091150	Golgi SNAP receptor complex member	F: CAATCAAGAAGGATCCAGCA R: GCCTTGCCAGTTTCTGTGTA	149		
<i>SIFRG08</i>	Solyc11g042930	E3 ubiquitin ligase	F: CTGGTGCCAACTACTTGAA R: TTAATCATGTCTGCCACGGT	71	0.82	0.983
<i>SIFRG09</i>	Solyc03g031950	unknown	F: CCCAATTTCTTCTCCGTAA R: TGGCCTAAGAAGTCCGAATC	108	0.98	0.954
<i>SIFRG10</i>	Solyc10g078450	U6 snRNA-associated Sm-like protein	F: TATAGCAATGGAGCAAACCG R: TGTCTCTTTGATGTGCTGA	112	0.75	0.997
<i>SIFRG11</i>	Solyc05g023800	Uncharacterized protein	F: ACAGGCACTGCATGAAGAAG R: GTGAAACTGATCCGCTCTGA	109	0.91	0.993
<i>SIFRG12</i>	Solyc04g009770	Ran protein/TC4 protein 1	F: AATCAAGTCCCAACCTGGAG R: TTCATGAATGGCCTTTGGTA	93	0.90	0.989
<i>SIFRG13</i>	Solyc01g103450	DNAJ chaperone	F: AATGACAAGGTTTCCAAGGC R: CTTCTAATCCAGCAATGCGA	100		
<i>SIFRG14</i>	Solyc06g076970	Peptidyl-prolylcis-trans isomerase	F: CTCTGCAGTTTGGTCGTGAT R: AATACGACCGGCAGGTTTAC	126	0.84	0.983
<i>SIFRG15</i>	Solyc01g010750	Stress responsive protein	F: GGATGAAGAGGATTTCCAA R: TCAAGAACGTCGGGATTGTA	120	0.96	0.995
<i>SIFRG16</i>	Solyc10g081190	Leucine-rich repeat receptor-like kinase	F: AACAGGACCAATCCCAAGAG R: TCAAGTCGAGGATTGTGCTC	147	0.99	0.995
<i>SIFRG17</i>	Solyc04g056350	Zinc finger family protein	F: ATGCAGCAACGACTAACAGG R: TGCAAAGAGGACATTCAAGC	99	0.84	0.993
<i>SIFRG18</i>	Solyc01g104700	Ran protein/TC4 protein ran2b	F: ACTTAGCTGCACAGGCATTG R: ATGGAAAGGTGCAACAGTGA	125	0.91	0.989
<i>SIFRG19</i>	Solyc10g006480	Ubiquitin	F: CAACCCTTCACTTGGTCCTT R: CTTTCCACCTCAAGGGTGAT	91	0.86	0.988
<i>SIFRG20</i>	Solyc10g083570	Fructose-bisphosphatealdolase	F: TGTTGAGCCTGAGATCCTTG R: GGGCTTCAACAATGTACCCT	136	1.01	0.992
<i>SIFRG21</i>	Solyc02g089200	SEP subfamily of the MADS-box gene f	F: GGTGGTGAGCAAAGTCTCAA R: GAGGTTGGAAGAAAACCTGA	67	1.10	0.988
<i>SIFRG22</i>	Solyc08g081190	plasma membrane intrinsic protein	F: GACCACTGGATCTTCTGGGT R: TTAAGACCTGTGGAATGGCA	99	1.39	0.999
<i>SIFRG23</i>	Solyc06g036540	Unknown	F: GGGCTTATCAAGGAAGGGAT R: GGCACCTCTTGTACCTTT	138	0.86	0.992
<i>SIFRG24</i>	Solyc02g080630	Lactoylglutathionylase	F: GGATGCAAGTGGTGAAGAAA R: CGTATGCATTTCCCTTTGTG	144	1.03	0.996
<i>SIFRG25</i>	Solyc05g050200	Eukaryotic translation initiation factor 1A	F: CGAAGAACAAGGGTAAGGGA	137	0.90	0.993

(Continued)

TABLE 2 | Continued

Gene ID	Gene locus ^a	Gene description	Primer sequence (5'–3')	Product size (bp)	E	R ²
SIFRG26	Solyc01g087320	tRNA-dihydrouridine synthase A	R: ACATCGTCCATTACCAAGCA F: TGGTGGTCCCATGATGAAAT	97	0.94	0.994
SIFRG27	Solyc06g007510	Ubiquitin-conjugating enzyme E2	R: AGCAGCAAGCATTCTGTGT F: GCTCTCTGTTGACAGACCCA	108	1.10	0.998
SIFRG28	Solyc06g005060	elongation factor 1-alpha	R: GAGTCCAGCTACGAGCAGTG F: TTGGTGGTATTGACAAGCGT	124	0.84	0.991
SIFRG29	Solyc02g084360	V-type proton ATPase 16 kDaproteo lipid	R: GTGATACCACGCTCACGTTTCC F: TCGCCTTTCTTCGGTTTCC	88	0.94	0.993
SIFRG30	Solyc09g082650	Acireductonedioxygenase	R: CACACCGCTCTTCGCTGTTC F:GGCAAGAATATGTTGAGACGTTTGTCA	147	0.80	0.983
SIFRG31	Solyc12g044600	NADP-malic enzyme	R:CATCAAGAACAGCAAAGCCACCA F: TCAGCGCATATTGCTGCCAA	125	0.90	0.996
SIFRG32	Solyc01g107870	Poly(A) RNA binding protein	R: CTGCGGTATGCTGGGCTGTA F: ACCTGGAATGAGGCTGGTG	150	0.91	0.99
SIFRG33	Solyc12g055800	Unknown	R: TGGTGGTATGCTGGGCTGTA F: GCTTCAGACTTGGCCTGTACG	127	0.91	1
SIFRG34	Solyc12g008590	Acireductonedioxygenase	R: GGCACAAGTCCACCAAGCA F: TCGACGATCACCTCCTGTGTG	81	1.12	0.962
SIFRG35	Solyc02g085540	NADH ubiquinone oxidoreductase	R: CCCAACAGCGCCGTCAAG For: CGCCTGGACAGTCCCGTAAA	127	0.87	0.998
SIFRG36	Solyc09g009260	fructose-1,6-bisphosphate aldolase	R: TGGGCCCAAGTTTCAAGGGT F: CTGAGTACACCGTCCGTGCT	102	0.90	0.996
SIFRG37	Solyc11g005330	Actin	R: GGTTGACGGTGGCTTCCTCT F: GGAGATTGAAACTGCCAGGAGCA	143	1.00	0.915
SIFRG38	Solyc01g028810	chaperonin	R: CTGCAGCTTCCATACCAATCATGG F:GTTGTGTGGTGGTTGAGATCAAGG	74	0.99	0.915
	Solyc01g095050	Negatively light-regulated protein	No primer			
	Solyc10g074860	Unknown	No primer			

^aTomato gene ID in SGN (<https://solgenomics.net/>). E: Amplification efficiency; R²: Correlation coefficient.

analysis. Based on the previously reported RGs in tomato (Coker and Davies, 2003; Exposito-Rodriguez et al., 2008; Løvdal and Lillo, 2009; Mascia et al., 2010; Müller et al., 2015), 73 reported potential RGs of tomato were identified (Table 1). Moreover, by collecting the cDNA sequences of 11 previously reported candidate RGs for watermelon fruit study (From CuGenDB: Cucurbit genomics database, <http://www.icugi.org/>) (Kong et al., 2015), we subsequently collected their corresponding orthologs in tomato through blastN in the SGN (Table 1). Thus, a total of 84 tomato potential RGs were collected in the present study. Among all the reported RGs and their corresponding orthologs, three genes (Solyc01g028930, Solyc10g049850, and Solyc03g115810) were found to be redundant. Thus, a total of 81 previously reported RGs were collected and listed in Table 1.

According to the RNA-seq data derived from TFGD, RPKM values of the 81 previously reported RGs in five different fruit developmental stages (1, 2, 3 cm, MG [Mature green], B [Breaker], B10 [10 days after breaker]) of accession “Heinz” and in three developmental stages (IM [Immature], B [Breaker], B5 [5 days after breaker]) of accession “Pimp” were used to evaluate their expression stabilities (Supplementary Table 1). The CV

values were calculated as described in Materials and Methods (Supplementary Table 3). Based on the RPKM values shown in Supplementary Table 1, we analyzed the relative expression of the 81 reported RGs among different developmental stages of tomato fruit. As shown in Supplementary Table 3, nearly 30% (24/81) of the 81 reported RGs demonstrated high expression variations (CV>0.35) during tomato fruit development (Figure 1A; Supplementary Table 3). Some CV values were even higher than 2.0 (Solyc12g096750/Solyc06g009860 [3.00], Solyc02g087880 [2.31]), suggesting their poor expression stabilities during tomato fruit development. Further analysis demonstrated that although the remaining 57 stably expressed reported RGs have CV values lower than 0.35, most of them (52/57) had the average expression levels (RPKM values) lower than 200, and some RPKM values of them (Solyc09g074220, Solyc12g096750, Solyc11g051210, and Solyc06g009860) were even close to 0 (low transcription level or undetected signal) (Supplementary Table 3), which indicated that these 52 reported RGs may not be qualified for normalization due to their low expression levels. According to Supplementary Table 3, only the PPKM values of five previously reported RGs, Solyc06g005060

(1088.0), Solyc11g005330 (581.1), Solyc01g028810 (509.5), Solyc04g009770 (357.6), and Solyc05g023800 (359.9), were high enough (>200) to be considered as candidate RGs in tomato fruit. Hence, we came to the conclusion that most of the previously reported candidate RGs identified so far were not well-qualified for normalization during tomato fruit development.

Identification of Putative RGs Based on RNA-Seq Data Mining

To comprehensively identify qualified RGs for tomato fruit, we searched the entire set of RNA-seq data derived from TFGD, and a total of 56 genes with RPKM values ranged between 200 and 2,000 among all developmental stages of tomato fruit were identified, by searching the derived RNA-seq data (**Supplementary Table 2**). The CV values of these genes were calculated, and more than 70% (40/56) of them were shown to be lower than 0.35 (**Supplementary Table 3**). The 40 genes were listed in details in **Table 2**. Further analysis revealed that the genes identified from the entire genome level (**Figure 1B**) were generally more stably expressed than those previously reported as candidate RGs (**Figure 1A**) during tomato fruit development.

qPCR Analysis of the Putative RGs in Tomato Fruit Development Process

Next, we intended to validate the expression stabilities of the selected 40 putative RGs by qPCR analysis. When designing the primers using genscript online tool (<https://www.genscript.com/>), we found that proper primers of two candidate RGs (Solyc01g095050 [321 bp] and Solyc10g074860 [72 bp]) for qPCR analysis cannot be designed due to their short cDNA sequences or high homologies with other genes in tomato. Thus, a total of 38 candidate RGs, which are designated as *SIFRG01* to *SIFRG38*, were eventually chosen for further expression validation in the qPCR experiments (**Table 2**).

qPCR amplification of the 38 candidate RGs were carried out using specific primers listed in **Table 2**, and the amplicon lengths ranged from 67 bp (*SIFRG21*) to 149 bp (*SIFRG07*). PCR-amplification specificity of each primer pair was verified by 1.5% agarose gel electrophoresis using cDNA templates (**Supplementary Figure 1**), and the melting curve analysis also showed single product peak (**Supplementary Figure 2**), which both confirm the specificities of the primer pairs. The amplification efficiencies (E) of these candidate RGs were found to vary from 0.76 (*SIFRG01*) to 1.39 (*SIFRG22*), and E values of more than 50% primer pairs (20/38) were ranged from 0.9 to 1.1, indicating their qualifications as primer pairs (Tichopad et al., 2002; Chung et al., 2004). Notably, the amplification efficiencies (E) of *SIFRG07* and *SIFRG13* could not be calculated due to their low transcript level in tomato fruit (**Supplementary Table 4**). The correlation coefficients (R^2) of the 38 candidate RGs ranged from 0.915 (*SIFRG37/SIFRG38*) to 1 (*SIFRG06/SIFRG33*) (**Table 2**).

The Ct values of the each putative RG derived from different tomato fruit developmental stages (Immature-IM; Mature Green-MG; Breaker-B; Mature Red-MR) (**Figure 2**;

Supplementary Table 4) were used here to evaluate the expression levels. The average Ct values of most candidate RGs (36/38) in various fruit developmental samples ranged between 20 and 30 (Except for *SIFRG07* [33.91] and *SIFRG13* [32.29]) (**Supplementary Table 4**).

Validation of Putative RGs Using geNorm and Normfinder

Of all the 38 tested RGs, the geNorm analysis determined that the 12 most stable candidate RGs with average expression stabilities (M-value) less than 0.6, which were: *SIFRG03*, *SIFRG27*, *SIFRG04*, *SIFRG23*, *SIFRG30*, *SIFRG24*, *SIFRG35*, *SIFRG37*, *SIFRG05*, *SIFRG16*, *SIFRG12*, and *SIFRG31* (**Figure 3A**; **Table 3**). To ensure the evaluation results, Normfinder was also conducted for evaluation, and 12 RGs with expression stabilities (M-value) less than 0.5 were identified, which were: *SIFRG25*, *SIFRG35*, *SIFRG05*, *SIFRG14*, *SIFRG38*, *SIFRG17*, *SIFRG04*, *SIFRG37*, *SIFRG16*, *SIFRG27*, *SIFRG12*, and *SIFRG29* (**Figure 3B**; **Table 3**).

Although the two different assessing systems (geNorm and Normfinder) came up with different results, there were still seven putative RGs (*SIFRG27*, *SIFRG04*, *SIFRG35*, *SIFRG05*, *SIFRG37*, *SIFRG16*, *SIFRG12*) that were found to be commonly top-ranked in both statistical algorithms (**Table 3**). Generally speaking, the primer pairs with amplification efficiency (E) between 0.9 and 1.1 (Tiangen, China) possess the lowest variability in qPCR analysis (Tichopad et al., 2002; Chung et al., 2004). However, we found that among the seven top-ranked RGs, the amplification efficiencies (E) of two primer pairs (*SIFRG04* [0.78] and *SIFRG35* [0.87]) were lower than 0.9, suggesting that the primer pairs of these two genes were not recommended for subsequent RG application in this study. Ct values of the remaining five RGs were all between 20 and 25 (qualified as RG). Therefore, *SIFRG27*, *SIFRG05*, *SIFRG37*, *SIFRG16*, and *SIFRG12* were finally identified as qualified and optimal RGs for normalization in the tomato fruit developmental process.

Previously, some researchers have reported that the use of more than one internal control genes in normalization could effectively improve the reliability of qPCR results (Reid et al., 2006; Exposito-Rodriguez et al., 2008; Gutierrez et al., 2008). Thus, we applied the geNorm software to calculate the pairwise variation values (V) of the 38 putative RGs (described in detailed in the Materials and Methods section). The pairwise variation revealed that the V_{2/3} value was 0.06 (significantly < 1.5) (**Figure 4**), which indicated that the combined use of two most stably expressed RGs as reflected in geNorm, *SIFRG03* and *SIFRG27*, was potentially sufficient for better normalization in qPCR experiments of tomato fruit developmental studies (**Figure 4**).

DISCUSSION

The advent of qPCR technology has brought a new revolution in the gene expression analysis area. Accurate interpretation of qPCR results mainly depends on the use of stable RGs for normalization, which can potentially minimize non-biological variations of different samples. Hence, the systematic

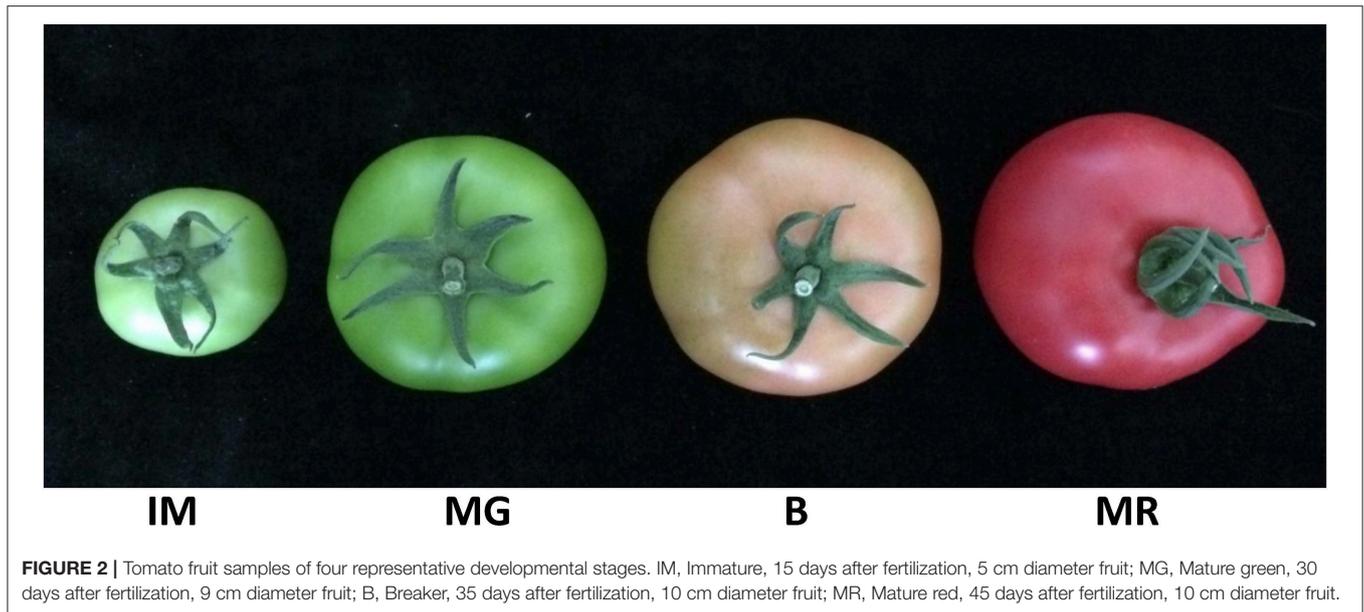


FIGURE 2 | Tomato fruit samples of four representative developmental stages. IM, Immature, 15 days after fertilization, 5 cm diameter fruit; MG, Mature green, 30 days after fertilization, 9 cm diameter fruit; B, Breaker, 35 days after fertilization, 10 cm diameter fruit; MR, Mature red, 45 days after fertilization, 10 cm diameter fruit.

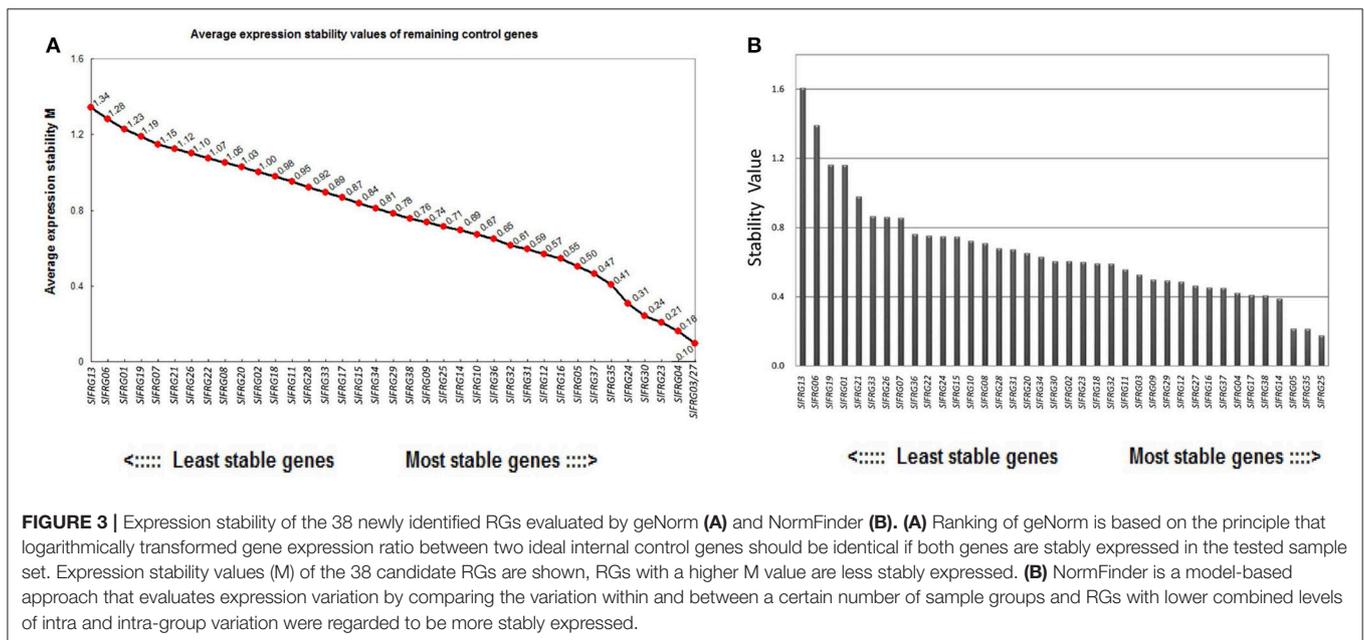


FIGURE 3 | Expression stability of the 38 newly identified RGs evaluated by geNorm (A) and NormFinder (B). (A) Ranking of geNorm is based on the principle that logarithmically transformed gene expression ratio between two ideal internal control genes should be identical if both genes are stably expressed in the tested sample set. Expression stability values (M) of the 38 candidate RGs are shown, RGs with a higher M value are less stably expressed. (B) NormFinder is a model-based approach that evaluates expression variation by comparing the variation within and between a certain number of sample groups and RGs with lower combined levels of intra and intra-group variation were regarded to be more stably expressed.

identification of appropriate RGs is essential for obtaining reliable results in qPCR experiments (Udvardi et al., 2008; Bustin et al., 2009; Guenin et al., 2009). Nowadays, some HKGs (e.g., Actin, Ubiquitin, and 18s rRNA), are usually used as RGs under various experimental conditions, or across a broad range of tissue samples (Bustin, 2002; Kong et al., 2014). However, an increasing number of evidence showed that optimal RGs varied depending on the experimental conditions or organs/tissues assayed, and it seems to be impossible to acquire a list of RGs universally practicable across a wide range of experimental conditions (Guenin et al., 2009; Warzybok and Migocka, 2013; Kong et al., 2014). Therefore, the identification of suitable RGs

for specific experimental conditions is essential for avoiding unnecessary error in the qPCR experimental results. So far, many studies involving the identification or evaluation of tomato RGs under various experimental conditions, including biotic/abiotic stresses (*Cucumber mosaic virus*, *tobacco mosaic virus*, *bacterium Xanthomonas campestris*, nitrogen stress, cold, light stress) (Coker and Davies, 2003; Løvdal and Lillo, 2009; Mascia et al., 2010; Wicczorek et al., 2013; Müller et al., 2015) and different organs/tissues (Leaf, fruit, flower, and seed) (Exposito-Rodriguez et al., 2008; Dekkers et al., 2012; Gao et al., 2012; Baldassarre et al., 2015), had been conducted. For example, *SIAC1*, *SICAC*, and *SIEF1α* were validated to be suitable RGs in

TABLE 3 | Top-ranked RGs evaluated by geNorm and NormFinder.

geNorm	M-value	NormFinder	M-value	Consensus
<i>SIFRG03</i>	0.095	<i>SIFRG25</i>	0.177	<i>SIFRG27</i>
<i>SIFRG27</i>	0.095	<i>SIFRG35</i>	0.215	<i>SIFRG04</i>
<i>SIFRG04</i>	0.156	<i>SIFRG05</i>	0.216	<i>SIFRG35</i>
<i>SIFRG23</i>	0.206	<i>SIFRG14</i>	0.390	<i>SIFRG05</i>
<i>SIFRG30</i>	0.242	<i>SIFRG38</i>	0.408	<i>SIFRG37</i>
<i>SIFRG24</i>	0.305	<i>SIFRG17</i>	0.411	<i>SIFRG16</i>
<i>SIFRG35</i>	0.409	<i>SIFRG04</i>	0.423	<i>SIFRG12</i>
<i>SIFRG37</i>	0.467	<i>SIFRG37</i>	0.452	
<i>SIFRG05</i>	0.504	<i>SIFRG16</i>	0.454	
<i>SIFRG16</i>	0.547	<i>SIFRG27</i>	0.464	
<i>SIFRG12</i>	0.569	<i>SIFRG12</i>	0.487	
<i>SIFRG31</i>	0.594	<i>SIFRG29</i>	0.494	

The seven RGs identified commonly in geNorm and Normfinder were labeled in bold.

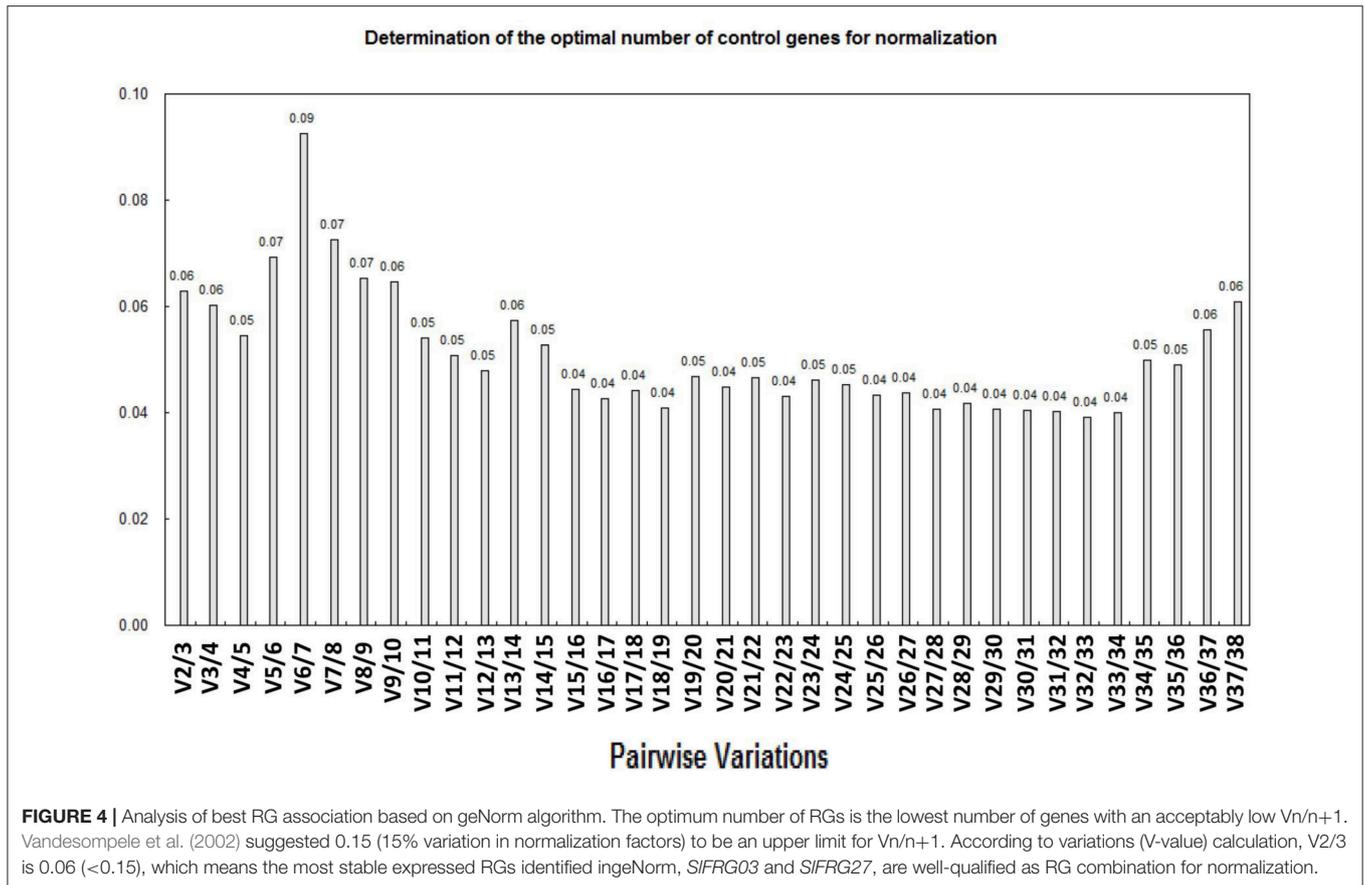
studies of host-virus interactions in tomato (Wieczorek et al., 2013). Exposito-Rodriguez et al. (2008) found that the widely used RGs, such as *SICAC*, *SITIP41*, *Expressed*, and *SISAND*, provide superior transcript normalization in various tissues of tomato (Exposito-Rodriguez et al., 2008). When studying the changes of gene expression in the wounded ripening-stage tomato fruit, Baldassarre et al. (2015) selected two most stably expressed RGs (*EF1- α* and *GADPH*) from seven routine used HKGs for normalization in the subsequent qPCR analysis. Thus, it occurred to us that although tomato is regarded as model plant for fruit development study, little attention has yet been paid to screen the best RGs specifically for normalization during the development of tomato fruit.

So far, most studies involving RG identification were based on the evaluation or validation of the expression stabilities of traditional or novel RGs under corresponding conditions (Czechowski et al., 2005; Løvdal and Lillo, 2009; Schmidt and Delaney, 2010; Dekkers et al., 2012; Baldassarre et al., 2015). In the present study, we collected 70 putative RGs that had been previously reported in tomato (Coker and Davies, 2003; Exposito-Rodriguez et al., 2008; Løvdal and Lillo, 2009; Mascia et al., 2010; Müller et al., 2015) and 11 orthologs of reported RGs in watermelon fruit study (Kong et al., 2015), and subsequently validated their expression stabilities during different stages of fruit development according to the RPKM values derived from RNA-seq data sets. Out of expectation, the majority of these putative RGs identified previously were not well-qualified for normalization as internal control genes due to either their unstable expressions or low transcript levels (Table 1; Figure 1A; Supplementary Table 1). Therefore, we next intended to identify some novel RGs that are stably expressed during the whole developmental process of tomato fruit.

Nowadays, the open sources of the SGN and TFGD allowed us to search for the most stably expressed genes on a comprehensive evaluation system. In the present study, we conducted a data mining based on the RNA-seq data set of different tomato fruit developmental stages, and initially collected 38 most stably

expressed genes as putative RGs for normalization in tomato fruit developmental study (Table 2). Further evaluation of these 38 putative RGs was conducted using qPCR analysis in four developmental stages of tomato fruit (IM, MG, B, and MG) (Figure 2). Next, we used two popular statistical algorithms for RG ranking, geNorm, and Normfinder (Vandesompele et al., 2002; Andersen et al., 2004), for the RG evaluation based on the qPCR results (Supplementary Table 4). We found different ranking results from the evaluation results of geNorm and Normfinder (Table 3), which were explicable as these two algorithms are based on different models and assumptions (Schmidt and Delaney, 2010). The geNorm algorithm is based on the principle that logarithmically transformed expression ratio of two genes should be constant if both of them are stably expressed in the tested sample set. The relative stability of each gene (M) is defined as the mean pairwise variation (reflected by standard deviation of the expression ratios of two genes) of the gene in the sample set. Furthermore, as normalization with single RG can cause inevitable errors, geNorm is also used to determine the minimum number of RGs required for more reliable normalization (Vandesompele et al., 2002). Normfinder measures gene expression stability by comparing the variation within and between a certain number of sample groups. The genes with the lowest combined levels of intra and inter-group variation were regarded as most stably expressed (Andersen et al., 2004; Schmidt and Delaney, 2010). Taken together, Normfinder is based on analyzing the variation level of each tested gene rather than pairwise analysis of gene stability relative to a set of potential RGs (Schmidt and Delaney, 2010). So far, numerous ranking differences of RGs derived by these two algorithms had been found in many previous studies (Schmidt and Delaney, 2010; Cassan-Wang et al., 2012; Kong et al., 2015). Nevertheless, we identified 7 putative RGs (*SIFRG04*, *SIFRG35*, *SIFRG27*, *SIFRG05*, *SIFRG37*, *SIFRG16*, and *SIFRG12*) that were common in geNorm and Normfinder. Considering the unqualified primer amplification efficiencies (E) of *SIFRG04* and *SIFRG35*, these two genes were excluded from our recommendation list, and the remaining five genes (*SIFRG05*, *SIFRG12*, *SIFRG16*, *SIFRG27*, and *SIFRG37*) were finally identified as suitable internal controls for normalization in tomato fruit development. Notably, we believe that some alternative primer pairs of *SIFRG04* and *SIFRG35* with improved amplification efficiencies might be redesigned for RG use in tomato fruit developmental study.

In practice, it is believed that the use of more than one RG in the normalization can efficiently improve the reliability of qPCR results (Alba and Giovannoni, 2005; Exposito-Rodriguez et al., 2008; Gutierrez et al., 2008). Thus, in order to explore the minimum number of RGs needed, the pairwise variation (V) values were calculated in geNorm (Figure 4). According to the evaluation, the combined application of two RGs, *SIFRG03*, and *SIFRG27*, would be a better choice than the use of only one RG for normalization when more reliable qPCR results are needed. It is also worth noting that due to the multiple sections of tomato fruit and the complex biological processes of fruit development, gene expression analysis has been extended to more precise tissue parts (e.g., pericarp, flesh, and even seeds) or longer developmental stages of fruits (Fraser et al., 1994; Carrari and Fernie, 2006;



Fei et al., 2011; Cheng et al., 2016). Therefore, we propose that the RGs identified in this study should be further validated in different tissue sections or earlier developmental stages (e.g., 1, 2 cm green fruits) of tomato fruit in the future according to specific experimental requirements.

CONCLUSION

To our knowledge, this study is the first systematic identification and evaluation of putative RGs as internal controls for normalization of qPCR analysis in tomato fruit developmental process. According to our extensive evaluation, five identified RGs-*SIFRG05*, *SIFRG12*, *SIFRG16*, *SIFRG27*, and *SIFRG37* could be recommended for normalization of qPCR experiments in tomato fruits. Furthermore, according to geNorm analysis, a combination of two most stably expressed RGs, *SIFRG03* and *SIFRG27*, were recommended when more reliable qPCR results were needed. Moreover, by comparative analysis of the previously published materials involving RG identification for fruit developmental study in other plants, we found that two RGs identified in this study were also chosen as optimal RGs for fruit developmental study in other plants (Zhu et al., 2012; Die and Rowland, 2013; Kong et al., 2015), which are ubiquitin conjugating enzyme (UBI) encoding genes (*SIFRG27* in tomato, *PEX4* and *UBC28* in blueberry, *UBCE* in

papaya) and actin encoding genes (*SIFRG37* in tomato, *CIACT* in watermelon, *ACTIN* in papaya). Thus, *SIFRG27/SIFRG37* and their corresponding orthologs seem to be universally applicable as RGs among plants of different families including Cucurbitaceae, Rosaceae, Vacciniaceae, and Solanaceae. Taken together, the results presented here not only unveil optimal RGs for qPCR analysis in tomato fruit development, but also provide referable guidelines for identification of RGs in other plant species.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: HW and YC; Performed the experiments: YC, WB, XP, JY, MR, QY, RW, ZL, GZ, and ZY; Analyzed the data: JY, WB, XP and GZ; Wrote the paper: YC, WB, XP, GA, and HW. All authors have read and approved the manuscript.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (31501749, 31301774, 31272156, 31772294), the Zhejiang Provincial Natural Science Foundation of China (LQ15C150002), Jiangsu Provincial Natural Science Foundation of China (BK20140277), Suzhou Science and

Technology Project (SYN201418), the Zhejiang Provincial Agricultural (Vegetable) New Variety Breeding Project (2016C02051-2-3, 2016C02051-1-1, 2016C02051-1-3), the Public Agricultural Technology Research in Zhejiang (2016C32101, 2015C32049), and the Technological System of Ordinary Vegetable Industry (CARS-25-G-16).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01440/full#supplementary-material>

Supplementary Figure 1 | PCR amplification products of the newly identified RGs on cDNA templates.

Supplementary Figure 2 | Specificity of the newly identified RG primers demonstrated in qPCR analysis. Melting curves generated for all amplicons.

Supplementary Table 1 | Expression levels (RPKM-value) of previously reported RGs against RNA-seq data of different developmental stages of tomato fruit.

Supplementary Table 2 | Expression levels (RPKM-value) of newly identified RGs against RNA-seq data of different developmental stages of tomato fruit.

Supplementary Table 3 | CV (Co-efficient variation), average expression (AVE) and standard deviation (SD) values of previous reported RGs and newly identified RGs in tomato.

Supplementary Table 4 | Ct-values of 38 newly identified RGs during tomato fruit development (IM, MG, B, MR).

Supplementary Table 5 | Expression stability (M) values of 38 newly identified RGs evaluated by geNorm and NormFinder.

REFERENCES

- Alba, R., and Giovannoni, J. J. (2005). Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. *Plant Cell* 17, 2954–2956. doi: 10.1105/tpc.105.036053
- Andersen, C. L., Jensen, T. F., and Orntoft, T. F. (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. doi: 10.1158/0008-5472.CAN-04-0496
- Baldassarre, V., Cabassi, G., Spadafora, N. D., Aprile, A., Müller, C. T., Rogers, H. J., et al. (2015). Wounding tomato fruit elicits ripening-stage specific changes in gene expression and production of volatile compounds. *J. Exp. Bot.* 66, 1511–1526. doi: 10.1093/jxb/eru516
- Bustin, S. A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* 29, 23–29. doi: 10.1677/jme.0.0290023
- Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., and Kubista, M. (2009). The MIQE Guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622. doi: 10.1373/clinchem.2008.112797
- Bustin, S. A., Benes, V., Nolan, T., and Pfaffl, M. W. (2005). Quantitative realtime RT-PCR—a perspective. *J. Mol. Endocrinol.* 34, 597–601. doi: 10.1677/jme.1.01755
- Bustin, S. A., and Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* 15, 155–166.
- Carrari, F., and Fernie, A. R. (2006). Metabolic regulation underlying tomato fruit development. *J. Exp. Bot.* 57, 1883–1897. doi: 10.1093/jxb/erj020
- Cassan-Wang, H., Soler, M., Camargo, E. L., Carocha, V., Ladouce, N., and Savelli, B. (2012). Reference genes for high-throughput quantitative reverse transcription-PCR analysis of gene expression in organs and tissues of Eucalyptus grown in various environmental conditions. *Plant Cell Physiol.* 53, 101–116. doi: 10.1093/pcp/pcs152
- Cheng, Y., Jalal, A. G., Yu, J., Yao, Z., Ruan, M., and Ye, Q. (2016). Putative WRKYs associated with regulation of fruit ripening revealed by detailed expression analysis of the WRKY gene family in pepper. *Sci. Rep.* 6:39000. doi: 10.1038/srep39000
- Chung, D. T., Drabek, J., Opel, K. L., Butler, J. M., and Mccord, B. R. (2004). A study on the effects of degradation and template concentration on the amplification efficiency of the STR miniplex primer sets. *J. Forensic Sci.* 49, 733–740. doi: 10.1520/JFS2003269
- Coker, J. S., and Davies, D. (2003). Selection of candidate housekeeping controls in tomato plants using EST data. *Biotechniques* 35, 740–742.
- Colombiet, S., Beauvoit, B., Nazaret, C., Benard, C., and Vercambre, G. (2016). Respiration climacteric in tomato fruits elucidated by constraint-based modelling. *New Phytol.* 213, 1726–1739. doi: 10.1111/nph.14301
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W. R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in arabidopsis. *Plant Physiol.* 139, 5–17. doi: 10.1104/pp.105.063743
- Dekkers, B. J., Willems, L., Bassel, G. W., van Bolderen-Veldkamp, R. P., Ligterink, W., and Hilhorst, H. W. (2012). Identification of reference genes for RT-qPCR expression analysis in arabidopsis and tomato seeds. *Plant Cell Physiol.* 53, 28–37. doi: 10.1093/pcp/pcr113
- Die, J. V., and Rowland, L. J. (2013). Superior cross-species reference genes: a blueberry case study. *PLoS ONE* 8:e73354. doi: 10.1371/journal.pone.0073354
- Exposito-Rodriguez, M., Borges, A. A., Borges-Perez, A., and Perez, J. A. (2008). Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biol.* 8:131. doi: 10.1186/1471-2229-8-131
- Fei, Z., Joung, J. G., Tang, X., Zheng, Y., Huang, M., and Lee, J. M. (2011). Tomato functional genomics database: a comprehensive resource and analysis package for tomato functional genomics. *Nucleic Acids Res.* 39, D1156–D116. doi: 10.1093/nar/gkq991
- Fraser, P. D., Truesdale, M. R., Bird, C. R., Schuch, W., and Bramley, P. M. (1994). Carotenoid biosynthesis during tomato fruit development (evidence for tissue-specific gene expression). *Plant Physiol.* 105, 405–413. doi: 10.1104/pp.105.1.405
- Gachon, C., Mingam, A., and Charrier, B. (2004). Real-time PCR: what relevance to plant studies. *J. Exp. Bot.* 55, 1445–1454. doi: 10.1093/jxb/erh181
- Gao, S., Xu, T., Qi, M., Liu, Y., Li, H., and Lv, S. (2012). Evaluation of the expression of internal control transcripts by real-time RT-PCR analysis during tomato flower abscission. *Afr. J. Biotechnol.* 11, 12983–12989. doi: 10.5897/AJB12.931
- Ginzinger, D. G. (2002). Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp. Hematol.* 30, 503–512. doi: 10.1016/S0301-472X(02)00806-8
- Guenin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C., and Gutierrez, L. (2009). Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *J. Exp. Bot.* 60, 487–493. doi: 10.1093/jxb/ern305
- Gutierrez, L., Mauriat, M., Pelloux, J., Bellini, C., and Van Wuytswinkel, O. (2008). Towards a systematic validation of references in real-time RT-PCR. *Plant Cell* 20, 1734–1735. doi: 10.1105/tpc.108.059774
- Huggett, J., Dheda, K., Bustin, S., and Zumla, A. (2005). Real-time RT-PCR normalisation: strategies and considerations. *Genes Immun.* 6, 279–284. doi: 10.1038/sj.gene.6364190
- Jacob, F., Guertler, R., Naim, S., Nixdorf, S., Fedier, A., and Hacker, N. F. (2013). Careful selection of reference genes is required for reliable performance of RT-qPCR in human normal and cancer cell lines. *PLoS ONE* 8:e59180. doi: 10.1371/journal.pone.0059180
- Jain, M., Nijhawan, A., Tyagi, A. K., and Khurana, J. P. (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. doi: 10.1016/j.bbrc.2006.04.140

- Jarosova, J., and Kundu, J. (2010). Validation of reference genes as internal control for studying viral infections in cereals by quantitative real-time RT-PCR. *BMC Plant Biol.* 10:146. doi: 10.1186/1471-2229-10-146
- Jian, B., Liu, B., Hou, W., Wu, C., and Han, T. (2008). Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol. Biol.* 9:59. doi: 10.1186/1471-2199-9-59
- Klee, H. J., and Giovannoni, J. J. (2011). Genetics and control of tomato fruit ripening and quality attributes. *Annu. Rev. Genet.* 45, 41–59. doi: 10.1146/annurev-genet-110410-132507
- Kong, Q. S., Yuan, J., Gao, L., Zhao, L., Cheng, F., and Huang, Y. (2015). Evaluation of appropriate reference genes for gene expression normalization during watermelon fruit development. *PLoS ONE* 10:e0130865. doi: 10.1371/journal.pone.0130865
- Kong, Q., Yuan, J., Niu, P., Xie, J., Jiang, W., and Huang, Y. (2014). Screening suitable reference genes for normalization in reverse transcription quantitative real-time PCR analysis in melon. *PLoS ONE* 9:e87197. doi: 10.1371/journal.pone.0087197
- Løvdaal, T., and 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. doi: 10.1016/j.ab.2009.01.024
- Mascia, T., Santovito, E., Gallitelli, D., and Cillo, F. (2010). Evaluation of reference genes for quantitative reverse-transcription polymerase chain reaction normalization in infected tomato plants. *Mol. Plant Pathol.* 11, 805–816. doi: 10.1111/j.1364-3703.2010.00646.x
- Müller, O. A., Grau, J., Thieme, S., Prochaska, H., Adlung, N., and Sorgatz, A. (2015). Genome-wide identification and validation of reference genes in infected tomato leaves for quantitative RT-PCR analyses. *PLoS ONE* 10:e0136499. doi: 10.1371/journal.pone.0136499
- Pfaffl, M. W. (2004). “Quantification strategies in real-time PCR,” in *AZ of quantitative PCR*, ed S. A. Bustin (La Jolla, CA: International University Line), 87–112.
- Reid, K. E., James, S., Niclas, O., Reid, K. E., and Lund, S. T. (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. doi: 10.1186/1471-2229-6-27
- Ruiz-May, E., and Rose, J. K. (2012). Progress toward the tomato fruit cell wall proteome. *Front. Plant Sci.* 4:159. doi: 10.3389/fpls.2013.00159
- Schmidt, G. W., and Delaney, S. K. (2010). Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol. Genet. Genomics* 283, 233–241. doi: 10.1007/s00438-010-0511-1
- Stürzenbaum, S., and Kille, P. (2001). Control genes in quantitative molecular biological techniques: the variability of invariance. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 130, 281–289. doi: 10.1016/S1096-4959(01)00440-7
- Suzuki, T., Higgins, P. J., and Crawford, D. R. (2000). Control selection for RNA quantitation. *Biotechniques* 29, 332–337.
- Tichopad, A., Dzidic, A., and Pfaffl, M. W. (2002). Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency. *Biotechnol. Lett.* 24, 2053–2056. doi: 10.1023/A:1021319421153
- Tieman, D., Zhu, G., Resende, M. F., Lin, T., Nguyen, C., and Bies, D. (2017). A chemical genetic roadmap to improved tomato flavor. *Science* 27, 391–394. doi: 10.1126/science.aal1556
- Udvardi, M. K., Czechowski, T., and Scheible, W. R. (2008). Eleven golden rules of quantitative RT-PCR. *Plant Cell* 20, 1736–1737. doi: 10.1105/tpc.108.061143
- Vandesompele, J., de Preter, K., Pattyn, F., Poppe, B., van Roy, N., dePaepe, A., 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. doi: 10.1186/gb-2002-3-7-research0034
- Wang, S. B., Liu, K. W., Diao, W. P., Zhi, L., Ge, W., and Liu, J. B. (2012). Evaluation of appropriate reference genes for gene expression studies in pepper by quantitative real-time PCR. *Mol. Breed.* 30, 1393–1400. doi: 10.1007/s11032-012-9726-7
- Warzybok, A., and Migocka, M. (2013). Reliable reference genes for normalization of gene expression in cucumber grown under different nitrogen nutrition. *PLoS ONE* 8:e72887. doi: 10.1371/journal.pone.0072887
- Wieczorek, P., Wrzesińska, B., and Obrepalska-Stepłowska, A. (2013). Assessment of reference gene stability influenced by extremely divergent disease symptoms in *Solanum lycopersicum* L. *J. Virol. Methods* 194, 161–168. doi: 10.1016/j.jviromet.2013.08.010
- Yoo, W. G., Kim, T. I., Li, S., Kwon, O. S., Cho, P. Y., and Kim, T. S. (2009). Reference genes for quantitative analysis on *Clonorchis sinensis* gene expression by real-time PCR. *Parasitol. Res.* 104, 321–328. doi: 10.1007/s00436-008-1195-x
- Zhu, X., Li, X. P., Chen, W. X., Chen, J. Y., Lu, W. J., and Chen, L. (2012). Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. *PLoS ONE* 7:e44405. doi: 10.1371/journal.pone.0044405

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Cheng, Bian, Pang, Yu, Ahammed, Zhou, Wang, Ruan, Li, Ye, Yao, Yang and Wan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.