RESEARCH ARTICLE

Screening internal controls for expression analyses involving numerous treatments by combining statistical methods with reference gene selection tools

Joyous T. Joseph[1] · Najya Jabeen Poolakkalody[1] · Jasmine M. Shah[1]

Abstract Real-time PCR is always the method of choice for expression analyses involving comparison of a large number of treatments. It is also the favored method for final confirmation of transcript levels followed by high throughput methods such as RNA sequencing and microarray. Our analysis comprised 16 different permutation and combinations of treatments involving four different *Agrobacterium* strains and three time intervals in the model plant *Arabidopsis thaliana*. The routinely used reference genes for biotic stress analyses in plants showed variations in expression across some of our treatments. In this report, we describe how we narrowed down to the best reference gene out of 17 candidate genes. Though we initiated our reference gene selection process using common tools such as geNorm, Normfinder and BestKeeper, we faced situations where these software-selected candidate genes did not completely satisfy all the criteria of a stable reference gene. With our novel approach of combining simple statistical methods such as *t* test, ANOVA and post hoc analyses, along with the routine software-based analyses, we could perform precise evaluation and we identified two genes, *UBQ10* and *PPR* as the best reference genes for normalizing mRNA levels in the context of 16 different conditions of *Agrobacterium* infection. Our study emphasizes the usefulness of applying statistical analyses along with the reference gene selection software for reference gene identification in experiments involving the comparison of a large number of treatments.

Keywords Reference genes · *Arabidopsis* · *Agrobacterium* · Stable expression · Normalization · Real-time PCR

✉ Jasmine M. Shah
jasmine@cukerala.ac.in; jas109@gmail.com

[1] Department of Plant Science, Central University of Kerala, Periye, Kasaragod 671316, India

Introduction

Studies on the response of plants towards pathogen stress are often initiated with expression profiling of host genes using techniques involving real-time PCR analysis, microarray, or transcriptome sequencing. Microarrays and transcriptome sequencing are high throughput techniques that reveal the expression response of almost whole of the genome. Real-time PCR is for specific gene expression analysis and is also used for the reconfirmation of microarray-derived and transcriptome sequencing-derived data. Real-time PCR quantifies the exponential amplification of a specific transcript by monitoring newly synthesized DNA in each PCR cycle (Higuchi et al. 1993).

One of the hindrances in doing real-time PCR is the selection of an internal control/reference gene. Reliability of real-time PCR is highly dependent on the stability of internal control, although other factors such as quality of template, and data evaluation are equally important. The use of inappropriate reference genes can lead to improper results that may draw false conclusions (Gutierrez et al. 2008; Guénin et al. 2009). Selection of an internal control often becomes a tedious task when a large number of conditions/treatments are to be considered because its expression should be stable in all the conditions/treatments under study and several genes may need to be systemically evaluated.

Springer

Our group works on plant-*Agrobacterium* interaction and we were involved in studying the influence of various *Agrobacterium* strains on host genome stability in terms of somatic mutations, using the model plant *Arabidopsis thaliana* (Shah et al. 2015). This work was extended to study the temporal response of DNA repair genes because of various *Agrobacterium*-derived factors such as Vir proteins and oncoproteins. (Joseph et al. unpublished). Since our analysis comprised a large number of treatments involving four different *Agrobacterium* strains (Table 1) at three different post-infection intervals, i.e. 16 different treatments in parallel, we opted for real-time PCR technique over whole transcriptome analysis. Unfortunately, none of the routinely used reference genes showed stable expression in the permutation and combination of various strains and time-frames that we used. Thus, we made an attempt to find out suitable reference genes for our analyses.

In *Arabidopsis,* validation and selection of superior reference genes has been carried out by different workers under different developmental stages (Czechowski et al. 2005), viral infection (Lilly et al. 2011) and abiotic stresses (Gutierrez et al. 2008; Remans et al. 2008). Though there are many reports on gene expression studies in *Arabidopsis* upon *Agrobacterium* infection, there is no report on the detailed evaluation of reference genes under *Agrobacterium* infection. Since our treatment conditions were of many types, we faced situations where the routinely used gene selection tools failed to identify a stable reference gene that satisfies all the desired criteria. In this report, we describe how we narrowed down to the best reference gene, out of 17 candidate genes by our novel approach of combining simple statistical methods along with the routine software-based analyses.

Materials and methods

Arabidopsis culture conditions

Surface sterilized *Arabidopsis thaliana* (ecotype Columbia) seeds were plated equidistantly on MS (Murashige and Skoog) media with 3% sucrose, pH 5.7. Surface sterilization was done by washing the seeds with 70% ethanol followed by sterile water and then treated with 0.5% sodium hypochlorite (bleach) for 1–2 min. Then, the seeds were washed five times with sterile water. Uniform germination of seeds was attained by keeping the MS plates for 48 h vernalization in the dark at 4 °C. Subsequently, the plates were transferred to growth chamber (Percival, USA. Model Ar-3663), having uniform light intensity (8000 lx), temperature 24 °C and 100% humidity with a 16-h light/8-h dark cycle. These conditions were maintained throughout the experiment.

Agrobacterium strains and method of infection

Four *A. tumefaciens* strains were used for infection—the wild type Ach5 and three of its derivatives (Table 1) named LBA4002, LBA4404, and LBA4404 (pCAMBIA2300) (*See* Shah et al. 2015). All the *Agrobacterium* strains were grown on Luria–Bertani (LB) media at room temperature (28 °C). The strain LBA4404 (pCAMBIA2300) was grown on media containing 100 mg/l kanamycin. The bacteria were inoculated into liquid LB media and kept overnight with shaking at 24 °C. Liquid cultures with optical density 0.8–1 at 600 nm were used for infection. Wounds (4–5 pricks) were made at the nodal region of 4 weeks old *Arabidopsis* plants using 2 ml clinical syringe and the culture was injected. Mock-inoculated controls received blank LB medium without any bacteria.

RNA isolation and cDNA synthesis

Total RNA was extracted from uninfected control plant, mock-inoculated control plant, and different *Agrobacterium* infected plants after 4 h, 24 h and 48 h post-infection intervals using Trizol (Invitrogen), according to the manufacturer's instructions. For all the 17 treatments, aerial parts (leaves and stem) of 10 four-week-old infected plants were taken for each RNA extraction. Infection and RNA isolation was done three times independently i.e., three biological replicates were used. Hence, in total, 30 plants were taken for each treatment. Extracted RNA was treated with DNaseI (Genie). The quantity and quality of total RNAs were assessed using Nanodrop 2000c Spectrophotometer (Thermo Scientific) and RNA gel electrophoresis. cDNA was synthesized from 1 μg total RNA

Table 1 *Agrobacterium tumefaciens* strains used

Name of the strain	Description	Bacterial factors		
		T-DNA	Vir proteins	Oncogenes
Ach5	Wild type	+	+	+
LBA4002	Avirulent, non-tumorigenic	–	–	–
LBA4404	Virulent, non-tumorigenic	+	–	+
LBA4404 (pCAMBIA2300)	Virulent, non-tumorigenic	+	–	+

as templates using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV-RT) kit (Invitrogen) according to the manufacturer's instruction. Synthesized cDNA was treated with RNase H (Invitrogen).

Primer designing, PCR standardization, and nucleotide sequence verification

Nucleotide sequences of all the 17 genes (Table 2) were accessed from NCBI (https://www.ncbi.nlm.nih.gov). Primer sequences of *EF1α, PPR, UP, AP2M and UBC28* were obtained from previous reports (Czechowski et al. 2005; Lilly et al. 2011). Primers for the remaining genes were designed from the exon region of genomic DNA, flanking an intron (Fig. 1A), using the online tool Integrated DNA Technology (IDT, https://www.idtdna.com). Gradient PCR was carried out to find out the best working temperature for the primers. For nucleotide sequence verification, PCR products were either sent directly for sequencing or cloned with TOPO TA Cloning Kit (Invitrogen) according to manufacturer's instructions, followed by sequencing. Sequencing was done at SciGenom Labs Private Ltd. Cochin, India.

Real-time PCR

Real-time PCR was used to detect the expression level of reference genes in *Arabidopsis* infected with *Agrobacterium*. cDNA template concentration was kept uniform by normalizing RNA concentration. The reaction was performed using Essential DNA Green Master (Roche). Each reaction mixture consisted of 1 μl cDNA, 10 μl 2X master mix, 1 μl (10 pmol/μl) each of both forward and reverse primers and nuclease-free water making up to a final volume of 20 μl. All the reactions were run on Roche-LightCycler® 480 II system. The thermal cycle consisted of an initial denaturation at 95 °C for 10 min, followed by 45 PCR cycles at 95 °C for 30 s, 47–55 °C (primer annealing Tm) for 15 s and 72 °C for 15 s. Each PCR cycle was followed by fluorescence acquisition at 95 °C for 1 min and 65 °C for 1 min. Subsequently, a melting curve was generated by increasing temperature from 65 to 95 °C, in order to verify primer specificity. Cp (crossing point) values were automatically generated by the LightCycler® 480 SW 1.5.1 software. Three independent experiments (biological replicates) were run for each gene and, triplicates of each treatment condition (technical replicates) were kept in each experiment.

Checking the amplification efficiency

Real-time PCR efficiency was determined for each gene with the slope of a linear regression model (Pfaffl 2001;

Walker 2002). For this, cDNA sample of each treatment was pooled and then used as the PCR template in a range of 100, 50, and 20 ng and run the reaction for 50 cycles. The corresponding real-time PCR efficiencies were calculated according to the equation by Radonic et al. 2004.

$$E = 10^{-1/\text{slope}}$$

Box plot analysis

To measure the statistical dispersion of the Cp values, box plot analysis was done and interquartile range (IQR) was calculated for all candidate reference genes using Rstudio (http://www.r-project.org/).

Reference gene selection tools

Three different types of softwares, geNorm (Vandesompele et al. 2002), NormFinder (Andersen et al. 2004), and BestKeeper (Pfaffl et al. 2004) were used for selecting the best candidate. The optimum number of reference genes required for normalization was determined by calculating pairwise variations using geNorm.

T-test, ANOVA, and Tukey-HSD

To identify the most stable gene under all the tested conditions, the mRNA expression stability of each reference gene was statistically analyzed by t-test. The genes were also statistically validated using single factor ANOVA. Genes with significant variations ($P < 0.05$) were taken for post-hoc analysis using Tukey-HSD in Rstudio and expression stability in each combination of treatments were noted.

Results

Selection of candidate reference genes, primer design and assessment of nucleotide sequence fidelity

Seventeen genes routinely used as reference genes in previous studies (Czechowski et al. 2005; Lilly et al. 2011; Remans et al. 2008; Pecinka et al. 2009; Jain et al. 2006) were selected (Table 2) for the analysis and identification of most stably expressed gene(s) in *Arabidopsis thaliana* infected with *Agrobacterium* for different time intervals. The seventeen genes included conventional reference genes with housekeeping activity such as *ACT II, Tubulin, EF1α, GAPDH* etc. and, genes with some other known/unknown functions such as *UP* (Unknown protein) and *EP* (Expressed protein). Primers were designed in such a way

Table 2 Details of reference genes used for real-time PCR analysis

Gene name	Accession number	Description	Primer sequence	Annealing Tm (°C)	Amplicon length	Real-time PCR efficiency
ACT II	AT3G18780	Actin II	F: 5′ TATGTCGCCATCCAAGC 3′	50	87	2.04
			R: 5′ GTGAGACACACCATCACC 3′			
GAPDH	AT1G42970	Glyceraldehyde-3-phosphate dehydrogenase beta subunit	F: 5′ AGAGTGCCATTCACTCTTTCC 3′	52	91	2.16
			R: 5′ CCACCGATACTACTCATACGC 3′			
β-tubulin 8	AT5G23860	Tubulin beta-8 chain	F: 5′ TCTCGATGTTGTTCGTAAGG 3′	50	95	2.12
			R: 5′ CCATTCCAGATCCAGTTCC 3′			
β-tubulin 6	AT5G12250	Beta-6 tubulin	F: 5′ CTCGATGTTGTACGCAAAGAGG 3′	53	81	2.09
			R: 5′ GGTGCCTCCACCAAGTG 3′			
eIF1	AT5G54760	Eukaryotic initiation factor 1	F: 5′ TAAGTTTCATGTCTGAACTTGACTCC 3′	51	87	2.02
			R: 5′ TTCCGGCACCTGAGTCC 3′			
eEF1α	AT5G60390	Eukaryotic elongation factor 1α	F: 5′ TTCACCCTTGGTGTCAAGCA 3′	54	76	2.15
			R: 5′ TTTCATCGTACCTGGCCTTGCA 3′			
PPR	AT1G62930	Pentatricopeptide repeat	F: 5′ GAGTTGCGGGTTTGTTGGAG 3′	52	61	2.16
			R: 5′ CAAGACAGCATTTCCAGATAGCAT 3′			
EP	AT1G56415	Expressed protein	F: 5′GAGCTGAAGTGGCTTCCATGAC 3′	55	81	2.11
			R: 5′GGTCCGACATACCCATGATCC 3′			
18S rRNA	X52320 (GenBank Accession)	18S ribosomal RNA	F: 5′ GCTTCTTAGAGGGACTATG 3′	47	118	1.98
			R: 5′ GTGTGAACTCGTTGAATAC 3′			
25S rRNA	AT2G01010	25S ribosomal RNA	F: 5′ CTCAACGAGAACAGAAATC 3′	50	110	1.95
			R: 5′ AGTCTAAAGGATCGATAGG 3′			
TUA 4	AT1G04820	Tubulin alpha-4 chain	F: 5′ GAACTCAGTGACATCAAC 3′	51	111	2.15
			R: 5′ ATTGAGAGACCAACCTAC 3′			
TUA 5	AT5G19780	Tubulin alpha-5 chain	F: 5′ CACCATCAAGACAAAGAG 3′	47	125	2.06
			R: 5′ CAGCTCTCTGAACCTTAG 3′			
UP	AT4G26410	Unknown protein	F: 5′ CCTGGAAGGGATGCTATCAA 3′	49	106	2.13
			R: 5′ GTCCGACATACCCATGATCC 3′			
AP2M	AT5G46630	Adaptor protein-2 mu-adaptin	F: 5′ GTGCCAATGTTCACAGCATC 3′	52	118	2.13
			R: 5′ TGATCTCGTAAGATCCCGCT 3′			
UBC28	AT1G64230	Ubiquitin conjugating enzyme 28	F: 5′ TCCAGAAGGATCCTCCAACTTCCTGCAGT 3′	55	240	2.15
			R: 5′ATGGTTACGAGAAAGACACCGCCTGAATA 3′			
UBC9	AT4G27960	Ubiquitin conjugating enzyme 9	F: 5′ TAACCATCCATTTCCCTCCA 3′	49	100	2.09
			R: 5′ TGGAAATTGTGAGAGCAGGA 3′			
UBQ10	AT4G05320	Polyubiquitin 10	F: 5′ CACCGGAAAGACAATCAC 3′	50	161	1.99
			R: 5′ GGATATTGTAATCAGCCAAC 3′			

that, they would bind to the exon region flanking an intron (Fig. 1a) so that the product size would be larger in genomic DNA and smaller in cDNA (Fig. 1b). This helped to confirm that, there was no genomic DNA contamination in the cDNA. The size of all PCR products was verified on a 2% agarose gel. Single fragments of the expected size and no bands were detected in control cDNA and non-template control (NTC), respectively, for all the 17 candidate reference genes (Supplementary Fig. 2). The absence of band in NTC indicate that there is no DNA
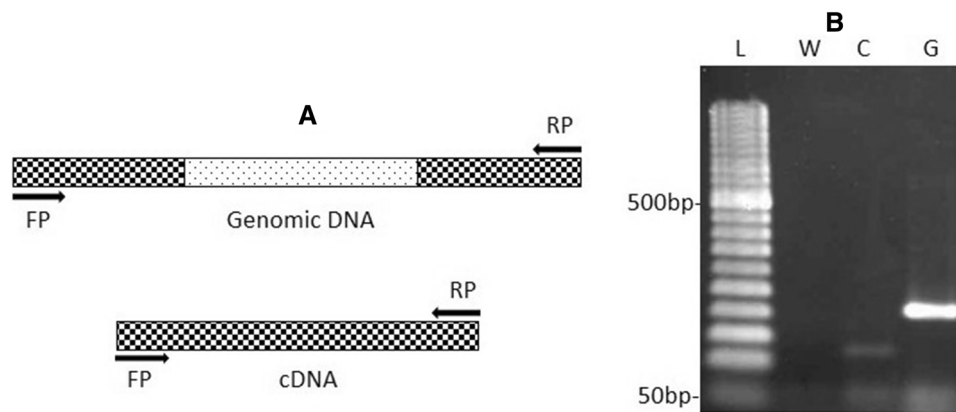
Fig. 1 **a** Location of primer designing. The box with thick squares indicates exon and box with small dots indicate intron. Arrows (FP and RP) indicate the location of forward and reverse primers. **b** Agarose gel showing amplification of a specific PCR product of expected size for four representative reference genes tested in the study. L, W, C and G indicates 50 bp ladder, water control, cDNA and genomic DNA, respectively

contamination in water and primer. Sequencing analyses displayed that all genes were identical to those deposited in the NCBI database.

Assessment of target specificity and real-time PCR efficiencies

All the primers were analyzed for target specificity by melting curve analyses following real-time PCR, and specific amplification was confirmed by a single peak (Fig. 2, Supplementary Fig. 1). All PCRs displayed efficiencies that ranged between 95% and 116% i.e., E = 1.95–2.16 (Table 2), though the acceptable range is 90% – 110% i.e., E = 1.6–2.1 (Pfaffl et al. 2004), there are chances of slight variations in this value due to factors including the instrument used, effect of the volume transferred throughout the dilution series etc. (Svec et al. 2015). Also, the standard curves plotted for calculating real-time PCR efficiency (data not shown) exhibited good linear relationships ($R^2 > 0.980$) (Table 1), confirming noble amplification efficiency of all the primer pairs in the real-time PCR-based quantification.

Evaluation of expression levels of candidate genes by Cp values

The expression levels of 17 candidate reference genes in *Arabidopsis* upon infection by different *Agrobacterium* strains under three different post-infection intervals (4 h, 24 h and 48 h) were assessed by means of Cp values obtained from real-time PCR (Supplementary Table 1). The median Cp values of candidate reference genes in this study ranged from 18.56 to 37.97 (Supplementary Table 1). *18SrRNA* and *25SrRNA* displayed the smallest median Cp values of 18.56 and 21.32, respectively,

indicating the highest accumulation of these two genes in *Arabidopsis*. *β-tubulin 8* showed the largest median Cp value of 37.97, representing the lowest accumulation of the same. Cp values of six genes (*β-tubulin 8, ACT II, GAPDH, β-tubulin 6, EIF1,* and *EF1α*) were high in some treatments (35–38.96) due to decreased expression under the stressed condition. These samples were subjected to agarose gel electrophoresis (Supplementary Fig. 3) and confirmed that the high Cp was not due to primer dimer. The melting curve of these samples also supported this (Supplementary Fig. 1). To measure the statistical dispersion and to find out how spread-out the Cp values are, box plot analysis was done (Fig. 3). The interquartile range (IQR) was calculated for all candidate reference genes. IQR of the Cp values of candidate genes analyzed in this study, varied from 0.773 to 3.7 at the overall level. Three of the 17 genes (*GAPDH, EIF1,* and *AP2M*) showed very high IQR, indicating high variability across Cp values. Least variation in gene expression across all 16 different conditions was exhibited by four genes, *UBQ10, 18SrRNA, 25SrRNA* and *PPR* (with IQR 0.882, 0.886, 0.773 and 0.933, respectively) (Fig. 3).

Analysis of expression stability of candidate genes using selection tools

Expression stability of 17 candidate genes across 16 different conditions of *Agrobacterium* infection was analyzed thrice using three different softwares—geNorm, NormFinder and BestKeeper.

geNorm analysis: The geNorm determines the average expression stability (M) of all the genes in a study. It is based on the geometric averaging of control genes and mean pairwise variation (V) of a gene from all other control genes in a given set of samples (Vandesompele et al. 2002). A low M value indicates more stable expression and
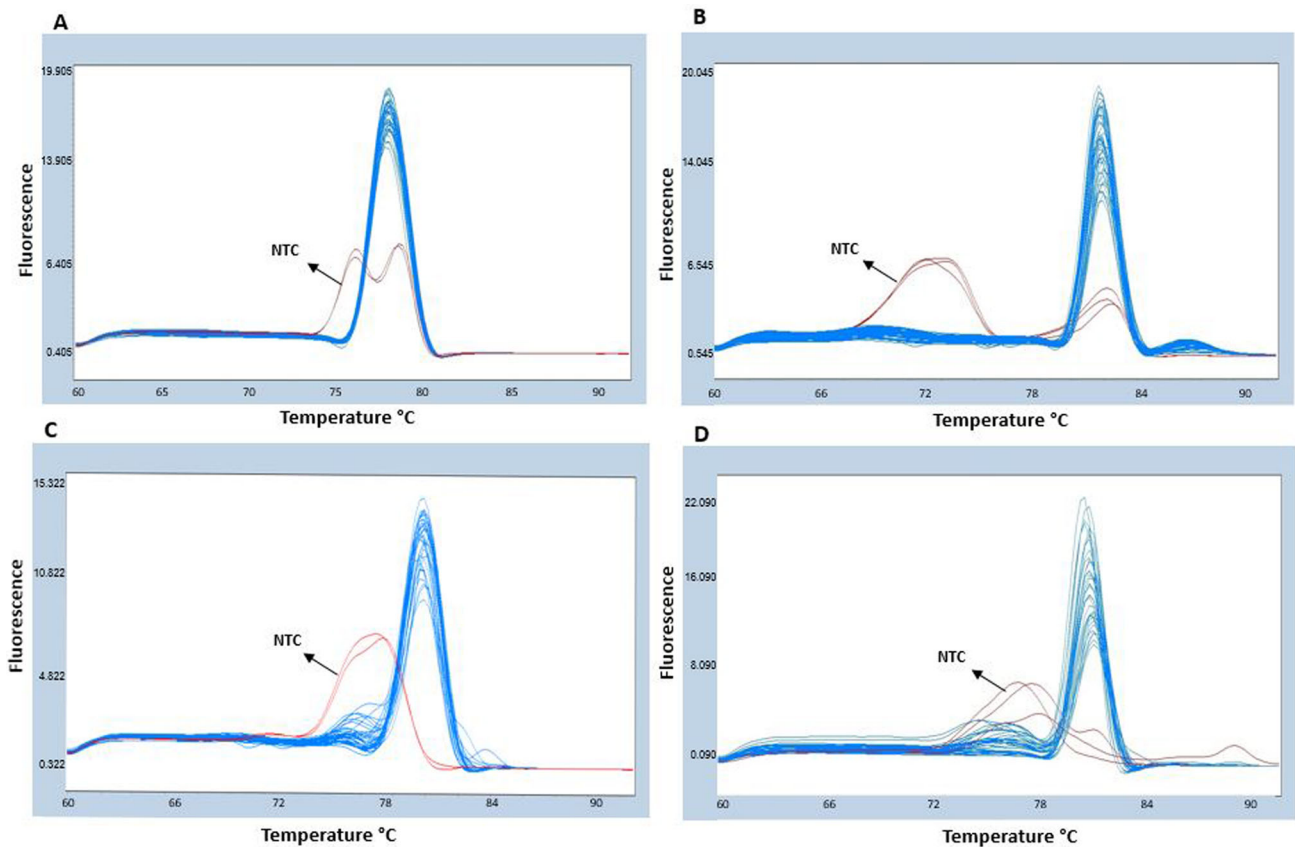
Fig. 2 Melting curves for four representative reference genes with single peak obtained from three replicates of 16 samples along with NTC (non-template control). **a** *25SrRNA*, **b** *18SrRNA*, **c** *eEF1α,* **d** *AP2M*
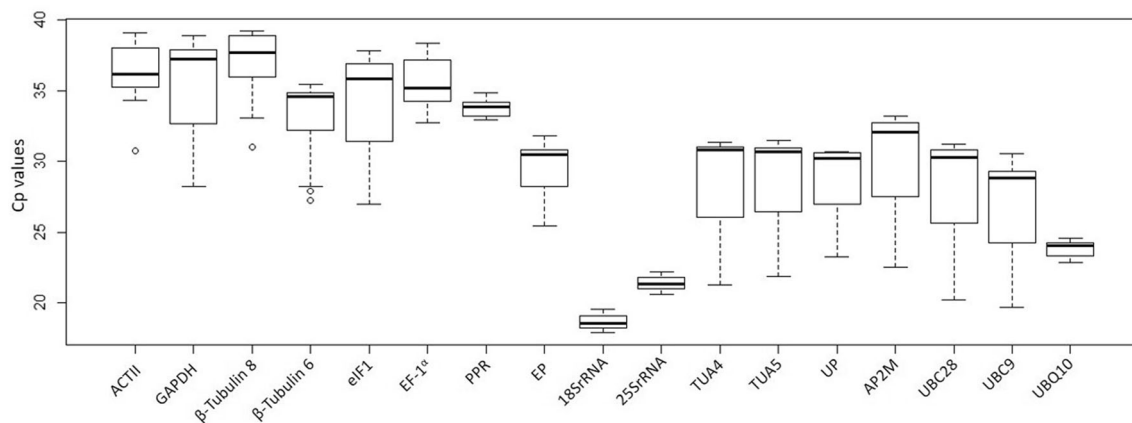

Fig. 3 Real time-PCR crossing point (Cp) values of 17 candidate reference genes in *A. thaliana* plants. Each box represents the values of respective genes in plants under 16 different conditions including infection with various *Agrobacterium* strains at different time intervals. The horizontal lines inside the boxes refer to median Cp values; the upper and lower hinges of the box indicate 75 and 25 percentiles. The whiskers show the largest and smallest Cp values that fall within a distance of 1.5 times the interquartile range from 75 and 25 percentiles. Outliers, values smaller (minimum) or larger (maximum) than 1.5 times the interquartile range (IQR), are indicated

hence, *UP* was picked up by the software as best gene out of the 17 candidates, followed by *β-tubulin 6, EP, TUA5* and *TUA4* (Fig. 5a). *β-tubulin 8, ACT II* and *EF1α*, with high M value, were chosen as the least stable candidates.

NormFinder analysis: NormFinder evaluates expression variation among candidate genes in a set of genes by means of raw Cp values. Stability is estimated by taking intra- and intergroup variation and a stability value is allotted to each gene. The gene(s) with the lowest stability value is the

most stable (Andersen et al. 2004). The ranking order of the genes obtained when our data was subjected to NormFinder analysis (Fig. 5b), was with only slight differences from that of geNorm. *β-tubulin 6, EP* and *UP* occupied the three top positions, whereas *β-tubulin 8, UBC28, UBC9* were found to be least stable genes followed by *ACT II* and *EF1α* (Fig. 5b). The ranking order of genes was highly consistent between the two softwares.

BestKeeper analysis: The BestKeeper tool calculates Cp value variations of candidate reference genes and thus estimates expression variability by performing a comparative analysis based on pairwise correlations of all candidate gene combinations. This tool ranks the genes based on standard deviation (SD), coefficient of correlation (r) and percentage covariance (CV) of all samples for each candidate reference gene. An index (named as 'BestKeeper') is given by combining all reference genes and the correlation between each reference gene and the index is calculated (Pfaffl et al. 2004). Gene expression variation was calculated for all the genes (Supplementary Table 2). *TUA4* was selected as the most stable gene, with the highest correlation coefficient (r = 0.997). This was followed by *AP2M, UP,* and *UBC9* (r = 0.996, 0.995 and 0.994). In BestKeeper analysis, genes with smaller SD and correlation factor closer to one are considered as best genes. Except for *UP,* the best genes selected by BestKeeper, and the ranking order were different from that of geNorm and Normfinder.

It is to be noted that the best genes selected by all the three softwares were with a relatively high IQR (IQR > 1.5), suggesting that there is notable variability in the range of Cp value across the treatments. Also, obvious difference was visualized in the Cp values of the best genes chosen by the three softwares (*UP, β-tubulin 6, EP* and *TUA4*) between control and treated samples (Supplementary Table 1). In order to validate this visual observation, we extended our study to statistical analysis using t-test.

Categorization of candidate genes based on t-test

The significance of variation in gene expression by means of Cp value comparison of each treatment to the uninfected control was evaluated for all reference genes using t-test. All genes analyzed, except *UBQ10, PPR, 18SrRNA* and *25SrRNA,* from the plants infected with *Agrobacterium* strains, showed significant variation in their mean Cp value when compared to uninfected plants. Three different patterns of gene expression could be observed after the Cp value comparison of treated and control samples using t-test (Table 3). Ten of the 17 genes were categorized as Type I where, significantly higher Cp values were obtained in all infected treatments, in comparison to uninfected controls (Table 3, Fig. 4a, Supplementary Fig. 4A to I). The mere presence of any of the *Agrobacterium* strains,

irrespective of infection time, reduced the expression of Type I genes. Expression of three genes was categorized as Type II as the Cp values significantly varied with different conditions such as wounding, strain type and infection time, in comparison to the uninfected control (Table 3, Fig. 4b, Supplementary Fig. 4J and K). Four genes were of Type III category and their expression exhibited uniform Cp value across all conditions including uninfected control (Table 3, Fig. 4c, and Supplementary Fig. 4L to N).

These findings from t-test were controversial to the findings by software-based selection. The software-selected genes showed significant variation in the expression levels between infected and uninfected samples. For instance, *UP,* the stable gene according to all the three softwares showed Type I pattern where gene expression was affected equally by all the bacterial strains (Supplementary Fig. 4I). The M value derived from geNorm, stability value derived from Normfinder and coefficient of correlation (r) derived from BestKeeper solely depended on the transformed/raw Cp value of all given set of samples under study and analyses overall expression stability without distinguishing between control and treated samples. In the case of *UP,* all infected samples gave Cp value in a range of 29.9–30.6 with very minute variations which are not statistically significant and therefore the software has chosen this gene as the best considering that 12 out of 16 conditions are stable. In fact, there is a significant difference in expression between the infected samples and uninfected control. Thus, the software selected genes are unfit and therefore, for our study, these analyses were found to be less dependable. Hence, we extended our analysis to Type III genes alone and reconfirmed their stability using three softwares.

Reconfirmation using reference gene selection tools

Four Type III genes, *18SrRNA, 25SrRNA, PPR* and *UBQ10,* having Cp values with no significant variation across 16 conditions were subjected for reconfirmation using three different softwares, geNorm, Normfinder and BestKeeper.

GeNorm analysis: *PPR* and *18SrRNA* were with least M value (0.707) followed by *UBQ10* and *25SrRNA* (M value 0.718 and 0.723 respectively) (Fig. 5c). M > 1.5 indicated an unacceptable level of expression variability and, all four had an M value below 1.5.

Single internal control gene is sufficient for normalization if the gene is highly stable. However, for some analyses, a single reference gene may not be adequate and may require normalization with two or more stable internal control genes. The optimum number of control genes required for normalization was determined by calculating pairwise variations using geNorm. The pairwise variation, V between two sequential normalization factors (NF)

Table 3 Categorization of reference genes based on pattern of gene expression and probable involvement of bacterial factors, after t-test

	Gene name	Factors influencing gene expression
Type I	AP2M, UBC9, GAPD, EIF1, β-tubulin 6, EP, TUA4, TUA5, UBC28, UP	Mere presence of bacteria
Type II	eEF1α, ACT II, β-tubulin 8	T-DNA or Vir proteins or oncoproteins or post-infection interval or unknown factors
Type III	UBQ10, 25SrRNA, 18SrRNA, PPR	None
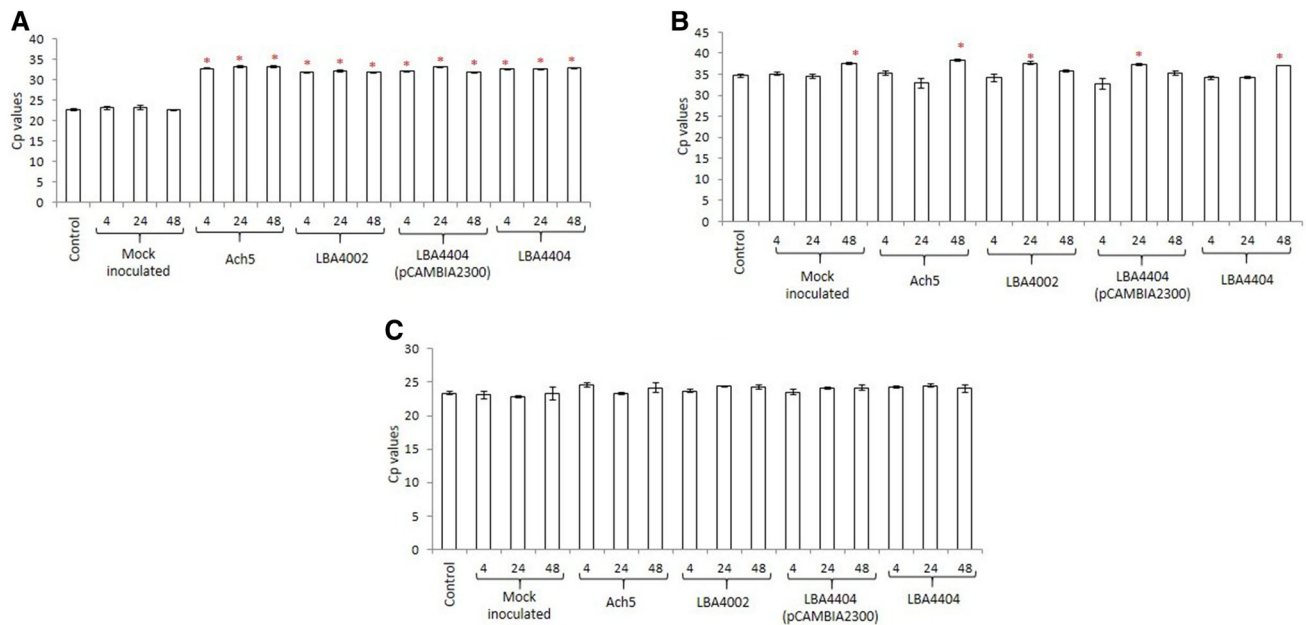
Fig. 4 Patterns of expression (based on Cp values) of three representative candidate reference genes *AP2M, eEF1α* and *UBQ10* belonging to three respective categories—**a** Type I, **b** Type II, **c** Type III, after infection by various *Agrobacterium* stains under different time intervals. The error bars indicate $\pm$ Standard error of the means of three biological replicates. *Indicates significant difference ($P < 0.05$) compared to the uninfected control

containing an increasing number of genes (NF_n and NF_{n+1}) was calculated (Fig. 5e). V value of 0.15 is the cut-off value, below which the inclusion of an additional reference gene is not required. Therefore, an optimal number of reference genes for normalization in this analysis was identified as three, where V was very close to the proposed cut-off value of V = 0.15 (V3/4 = 0.151).

Normfinder analysis: Out of the four genes subjected for this analysis, *PPR* was with least stability value 0.014 followed by, *25SrRNA, UBQ10* and *18SrRNA* (Fig. 5d). The three genes, *UBQ10, 25SrRNA* and *18SrRNA* differed only slightly in their stability values i.e. 0.023, 0.024 and 0.026.

BestKeeper analysis: All the four candidate reference had SD values < 1 (Table 4), indicating that all were good enough to be selected as reference genes. Subsequent assessment using pairwise correlation and regression analysis (Table 4) indicated that *18SrRNA* was with the highest correlation coefficient (r = 0.691). This was followed by *PPR, 25SrRNA,* and *UBQ10* (r = 0.594, 0.583 and 0.467).

The stability order given by three different softwares were different. geNorm detected the stability order *PPR = 18SrRNA > UBQ10 > 25SrRNA,* Normfinder provided the order *PPR > 25SrRNA > UBQ10 > 18SrRNA,* whereas, according to BestKeeper, the order was *18SrRNA > PPR > 25SrRNA > UBQ10.* Moreover, by calculating pairwise variations using geNorm, the optimal number of reference genes required for normalization was identified as three. In order to decide which three genes should be considered, we extended our study using ANOVA and post hoc analysis by Tukey-HSD.
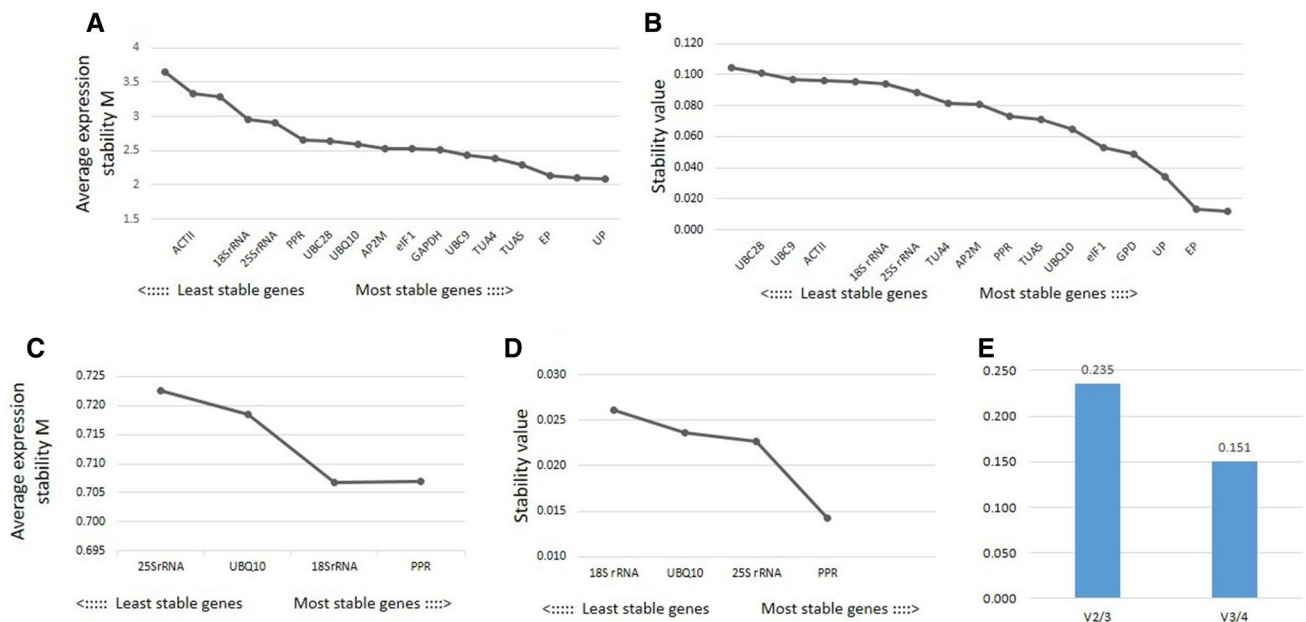
Fig. 5 Expression stability of all the candidate reference genes analyzed by **a** geNorm and **b** Normfinder. **c** Expression stability of Type III reference genes analyzed by geNorm and **d** Normfinder. **e** The optimal number of reference genes required for effective real-time PCR data normalization based on pairwise variation using geNorm

Table 4 Descriptive statistics of the four Type III genes based on their crossing point (Cp) values as calculated by BestKeeper

Gene name	18SrRNA	PPR	25SrRNA	UBQ10
Ranking	1	2	3	4
Geometric mean (Cp)	18.61	33.79	21.41	23.84
Arithemetic mean (Cp)	18.61	33.79	21.41	23.84
Minimum (Cp)	17.91	32.90	20.60	22.88
Maximum (Cp)	19.56	34.85	22.20	24.59
Standard deviation ($\pm$ Cp)	0.45	0.47	0.45	0.47
Covarience (%Cp)	2.41	1.40	2.12	1.99
Coefficient of correlation (r)	0.691	0.594	0.583	0.476
p-value	0.003	0.015	0.018	0.062

Precise evaluation using ANOVA and post hoc analysis by Tukey-HSD

All the four Type III genes were subjected to single factor ANOVA. *UBQ10* and *PPR* were identified without significant variations i.e. with $P > 0.05$ (0.89 and 0.143, respectively). *18SrRNA* and *25SrRNA* showed $P < 0.05$ (6.74e−07 and 0.00143, respectively) indicating that variations do exist in their expression. In order to find out where exactly the variation exists, post hoc analysis by Tukey-HSD was carried out for these latter two genes. For *18SrRNA*, out of 120 comparisons among 16 conditions, 13 pairs of treatments showed significant variation whereas for *25SrRNA*, one pair of treatment showed significant variation (data not shown). Therefore, after performing ANOVA and post hoc analysis by Tukey-HSD, *UBQ10* and *PPR* were chosen as the best genes, without any variation.

Discussion

Normalization is a very important step in gene expression analyses and therefore the choice of the reference gene. A meaningful quantification of target mRNA copy numbers requires accurate and relevant normalization to a stably expressed reference gene. The main aim of using a reference gene in expression analyses is to rectify errors in the whole research (Huggett et al. 2004), which include errors occurring at various stages of sample preparation and problems associated with RNA normalization (Bustin et al. 2009; Mallona et al. 2010). False reference gene selection

leads to false results. If the treatment number is high, finding a stable reference gene under all the treatment conditions can often be difficult. In such situations, several genes need to be analyzed in order to get a stable gene which would satisfy all the treatment conditions under study.

Our group works on plant-*Agrobacterium* interaction and we have studied the influence of various *Agrobacterium* strains on host genome stability in terms of somatic mutations (Shah et al. 2015). This work was extended to study the expression of DNA repair genes due to various *Agrobacterium*-derived factors such as Vir proteins and oncoproteins. (Joseph et al. unpublished). We had to use real-time PCR, to compare gene expression under the influence of various bacterial factors, in different time frames. We had to analyze about 16 treatments in parallel. Unfortunately, we had a delayed take-off in this work because none of the routinely used reference genes gave stable expression in all the 16 conditions (permutation and combination of various strains and time-frames). Initially, we selected six traditional housekeeping genes (*ACT II*, *GAPDH*, *β-tubulin 6*, *β-tubulin 8*, *UBC9* and *eIF1*) which were previously used for normalization (Kim et al. 2007; Lacroix and Citovsky 2014; Tayeh et al. 2014). However, none of these genes were found to be stable under all the tested conditions and therefore, we ended up in screening 17 genes.

The Cp values generated in real-time PCR were used for evaluating the stability of reference genes. The genes were initially analyzed using three different selection tools—geNorm, Normfinder, and BestKeeper. These softwares selected *UP*, *β-tubulin 6*, *EP* and *TUA4* genes as the best genes for normalization. However, we could visualize obvious changes in the Cp values of the genes selected by these softwares. Hence, we validated the visual observation by t-test. Followed by the t-test, 17 genes were categorized into three groups (Type I, Type II and Type III). This grouping was based on the patterns of gene expression. Expression of Type I genes was influenced just by the mere presence of bacteria, irrespective of strain types or infection time. Interestingly, all these genes had Cp values higher than the uninfected controls, reflecting lower expression. Type II genes were influenced by various factors such as bacterial strain, time interval post infection and wounding. These strains varied for the presence or absence of bacterial factors like T-DNA, Vir proteins, and oncogenes. Expression of four genes (*18SrRNA*, *25SrRNA*, *UBQ10* and *PPR*) did not vary with any of the treatment conditions and these were grouped as Type III. This categorization helped for the precise evaluation of gene expression stability under each treatment compared to the uninfected control, rather than an overall evaluation of gene stability. The software-selected genes belonged to

Type I category which showed significantly high Cp value upon infection compared to control.

Software-based reference gene selection is highly useful for experiments involving a few number of treatments or conditions. As the number of conditions to be studied increase, the software selects a gene showing stability across most of the conditions and may sometimes omit one or two conditions which show variations. In our case, this omitted treatment was uninfected control. It is very important that the reference gene expression should be stable between the uninfected control and the infected samples because the ΔΔCp method of gene expression normalization depends not only on the Cp value of reference gene under corresponding treatment but also the Cp value of uninfected control (Livak and Schmittgen 2001). The software is unable to distinguish between control and treatment. This was the reason why we introduced manual analysis using statistical methods in our experiment. Since the software was not completely reliable, we extended the software-based analysis only on Type III genes, i.e. those which showed stability across infected and uninfected samples based on t-test.

The stability of Type III genes was reassessed using geNorm, Normfinder, and BestKeeper. All the three softwares suggested four candidates as best reference genes, *18SrRNA, 25SrRNA, UBQ10* and *PPR,* for our analyses involving different *Agrobacterium* strains and different post-infection intervals. However, the ranking order of four genes by the three selection tools varied considerably. This difference might have resulted because the statistical algorithms used by the three methods were distinct. geNorm detected the stability order *PPR = 18SrRNA > UBQ10 > 25SrRNA,* Normfinder provided the order *PPR > 25SrRNA > UBQ10 > 18SrRNA,* whereas, according to BestKeeper, the order was *18SrRNA > PPR > 25SrRNA > UBQ10.* The particular value which determines gene stability such as M value, stability value and the correlation coefficient of geNorm, Normfinder and BestKeeper respectively, of all the four genes vary from each other only slightly and they fall within the requisite value for reference genes.

More specific analysis of individual treatments using ANOVA and Tukey-HSD gave a more precise picture about which reference gene is better for a particular experiment using any/all of individual *Agrobacterium* strains as well as time frames. For instance, though *18SrRNA* is a stable gene satisfying the stability criteria of all the three selection tools, post hoc analysis of this gene revealed that infection with the strain LBA4404 (pCAMBIA2300) 4 h post infection varied with some other treatments such as LBA4404 24 h, LBA4404 48 h, LBA4002 24 h, Mock 4 h and mock 24 h. This suggests the inability of this gene to be used as a reference for

studies involving this particular combination of strains and time-period. *UBQ10* and *PPR* were without any significant changes among all the 120 combinations of treatments and hence, can be used as reference gene for gene expression analysis involving any combination of the treatments analyzed in our study.

The software such as geNorm, Normfinder and Best-Keeper gave an overall idea about the gene stability whereas, simple statistical analyses such as t-test, ANOVA, and post hoc analyses were useful in precisely analyzing the gene stability across each treatment. Therefore, our study emphasizes the precise evaluation of reference genes by statistical analyses such as t-test and ANOVA along with the routine software based analyses. Though the optimum number of control genes required for normalization determined by calculating pairwise variations using geNorm was three, precise evaluation by statistical analyses such as t-test and ANOVA along with the routine software based analyses suggested only two genes, *PPR* and *UBQ10*, suitable to be used as reference gene for normalising mRNA levels within the context of the different *Agrobacterium* infections and time frames tested in this study.

In previous reports on *Arabidopsis-Agrobacterium* interaction, *ACT II* (Lacroix and Citovsky 2014; Wu et al. 2014), *UBC 21* (Wu et al. 2014), *ACTII/8* (Woo-Lee et al. 2009), *EF 1–α* (Vaghchhipawala et al. 2012; Park et al. 2015), *GAPDH* (Kim et al. 2007), and *18SrRNA* (Iwakawa et al. 2017) were used as reference genes. Most of these genes were not stable in our study because the conditions we used, such as strain, ecotype, infection method and age of plant varied considerably from the previous studies. For instance, the study of Wu et al. 2014, involved different *Agrobacterium* stain (C58 background), different *Arabidopsis* ecotypes and mutants (Col-0, Ws-2, *efr-1*, *fls2* etc.), younger seedling age (4-day-old), different infection method (without wounding) and a longer co-cultivation period (4 days). Similarly, Lacroix and Citovsky 2014, worked with another *Agrobacterium* strain (EHA105), and used root for RNA isolation. Woo-Lee et al. 2009 also used different *Agrobacterium* strains (oncogenic strain C58 and a T-DNA–deficient derivate of C58, GV3101). Park et al. 2015, used non-tumorigenic strains GV3101 containing pBISN1 or pKM1 and the tumorigenic strain A208 and the expression analysis was done on callus from root segment. Similarly, Kim et al. 2007 used *Arabidopsis* suspension culture transformed with *A. tumefaciens* strain At849. In a recent study (Iwakawa et al. 2017), *18SrRNA* proved to be a good internal control in *Arabidopsis* infected with wild type *Agrobacterium*.

Since our experimental set up is different from the above mentioned studies, it is difficult to relate the reference gene expression because it is highly dependent on multiple factors, and reference gene for one experiment may be unfit for another experiment (Joseph et al. 2018). The reference gene expression variaton between six and seven leaved stage in *Solanum lycopersicum*. Similarly, the expression of reference gene differed between bud sizes of 1 to 8mm in *Solanum lycopersicum* (Expósito-rodríguez et al. 2008), thus, indicating that even a very short time interval may also cause variation in reference gene expression. Also, it may vary between different genotypes of the same plant, as in the case of some of the reference genes in *Glycine max* (Rodrigues et al. 2010). It may also show variability in different organs of a plant as well. For instance, in cotton, some genes that are stably expressed in shoot vary when it comes to root (Wang et al. 2013). This could be the reason why we did not get a stable expression in most of the previously used reference genes. Our study suggests two best reference genes, *UBQ10* and *PPR* for analyses involving different *Agrobacterium* strains and different post-infection intervals. *UBQ10* has been previously reported to be stable under various biotic stresses (Denoux et al. 2008) and developmental stages (Czechowski et al. 2005) in *Arabidopsis* and we propose the stability of this gene under *Agrobacterium* stress as well. *PPR* gene showed stable expression in virus-infected *Nicotiana benthamiana* but was omitted due to its low transcript level (Liu et al. 2012).

In this study, *Agrobacterium* strains that varied in the presence of their secreted factors like Vir proteins, T-DNA and oncoproteins (Table 1), induced varied expression of some of the reference genes (e.g. *ACT II, EF1α* and *Tubulin8*) at all three or at least two time intervals. Our study thus reveals that the microbe-associated factors of the closely related strains also alter reference gene expression and thus thorough evaluation of reference gene is compulsory before performing any expression analyses. Previous reports suggested that secreted factors from *Agrobacterium* altered the expression of various genes like defense, cell signalling and growth (Veena et al. 2003) and we report that these factors could alter the expression of some of the housekeeping genes as well. *ACT II* showed significant variation upon infection by all *Agrobacterium* strains 48 h post infection. The downregulation obtained for *ACT II* at 48 h of Ach5 infection is in accordance with the findings of Ditt et al. 2006, where they detected downregulation for different actin genes (*Actin 12, Actin 3, ACT3, ACT II/7*) at 48 h upon infection by wild type *Agrobacterium*. Wild type Ach5 did not alter *ACT II* gene expression at 4 h and 24 h post-infection. Whereas, the other strains LBA4002, LBA4404 and LBA4404 (pCAMBIA2300) caused a significant change (increase/decrease) in Cp value at 24 and 48, 4 and 24 and, 4, 24, and 48 h post-infection, respectively.

Wounding was the method of infection used in our study. Wounding does not influence the expression of Type I and Type III genes since the mock-infected plants did not show any difference in expression levels from that of the controls. The Type II genes, on the other hand, had varied Cp values due to different conditions such as wounding, *Agrobacterium* strains and infection time. *EF1α* and *Tubulin8* were found to be affected by wounding. *EF1α* showed a significantly high Cp value at 48 h post mock inoculation whereas, *Tubulin8* displayed a decrease in the same at 4 h and 24 h post mock inoculation. Reymond et al. 2000, reported that wounding may alter the expression of control genes such as *tubulin*, *actin*, and *translation elongation factors*. However, in our study wounding does not alter the expression of *ACT II* whereas the other two genes (*EF1α* and *Tubulin8*) showed significant variation. *EF-1α* is found to be altered at 48 h post-inoculation by wild type Ach5 as well. It could be possibly due to the effect of wounding since, wounding influence the gene expression at 48 h, and there is no significant variation between 48 h post-mock inoculation and 48 h post-Ach5 infection (Fig. 4b). Vaghchhipawala et al. 2012, used *EF1α* for normalizing the expression of integrated DNA in host plant after transformation. So, it is suspected that the gene would be stable under Ach5 infection and the variation might be drawn solely due to wounding. *EF-1α* expression levels are associated with the extent of survival of the cells with the highest expression conferring the best survival (Talapatra et al. 2002) and therefore, the increase in Cp value, i.e., low level of expression, could be due to cell death after 48 h post-mock inoculation which may be probably due to wounding.

Conclusions

This study is first of its kind, aimed to validate candidate reference genes in *A. thaliana* infected with various *Agrobacterium* strains at different time intervals. The number of treatment conditions was high (16) and we hereby report the analysis of 17 candidate genes to be selected as reference genes for the normalization of real-time PCR data. We have used a novel approach of carefully combining gene selection softwares with statistical analysis like t-test, ANOVA and Tukey-HSD post-hoc test. The use of these statistical methods gave a better understanding of the expression behavior and, this helped us narrow down to two best reference genes, *PPR* and *UBQ10*, stable under all the permutation and combination of treatments studied. If the number of treatments is high, we recommend not to solely rely on the software-based gene selection tools since these are based on algorithms and assumptions and, a better understanding of the underlying principles of using a

stable reference gene would help in reliable validation. Even when using closely related strains belonging to the same species, we recommend thorough validation of reference gene before performing any expression analyses.

References

Andersen CL, Jensen JL, 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. https://doi.org/10.1158/0008-5472.CAN-04-0496

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M et al (2009) The MIQE guidelines: minimum information for publications of quantitative real-time PCR experiments. Clin Chem 55:611–622. https://doi.org/10.1373/clinchem.2008.112797

Czechowski T, Stitt M, Altmann T, Udvardi MK (2005) Genome-wide identification and testing of superior reference genes for transcript normalization. Plant Physiol 139:5–17. https://doi.org/10.1104/pp.105.063743.1

Denoux C, Galletti R, Mammarella N, Gopalan S, Werck D, De Lorenzo G, Ferrari S, Ausubel FM, Dewdney J (2008) Activation of defense response pathways by OGs and Flg22 elicitors in *Arabidopsis* seedlings. Mol Plant 1:423–445. https://doi.org/10.1093/mp/ssn019

Ditt RF, Kerr KF, De Figueiredo P, Delrow J, Comai L, Nester EW (2006) The *Arabidopsis thaliana* transcriptome in response to *Agrobacterium tumefaciens*. Mol Plant Microbe Interact 19:665–681. https://doi.org/10.1094/MPMI-19-0665

Expósito-rodríguez M, Borges AA, Borges-pérez A, Pérez JA (2008) Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biol 8:1–12

Guénin S, Mauriat M, Pelloux J, Van Wuytswinkel O, Bellini C, 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. https://doi.org/10.1186/1471-2229-8-131

Gutierrez L, Mauriat M, Guénin S, Pelloux J, Lefebvre J, Louvet R, Rusterucci C et al (2008) The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnol J 6:609–618. https://doi.org/10.1111/j.1467-7652.2008.00346.x

Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology 11:1026–1030. https://doi.org/10.1038/nbt0993-1026

Huggett J, Dheda K, Bustin S, Zumla A (2004) Real-time RT-PCR normalization; strategies and considerations. Genes Immun 6:279–284. https://doi.org/10.1038/sj.gene.6364190

Iwakawa H, Carter BC, Bishop BC, Ogas J, Gelvin SB (2017) Perturbation of H3K27me3-associated epigenetic processes

increases *Agrobacterium*-mediated transformation. Mol Plant Microbe Interact 30:35–44. https://doi.org/10.1094/MPMI-12-16-0250-R

Jain M, Nijhawan A, Tyagi AK, Khurana JP (2006) Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. Biochem Biophys Res Commun 345:646–651. https://doi.org/10.1016/j.bbrc.2006.04.140

Joseph JT, Poolakkalody NJ, Shah JM (2018) Plant reference genes for development and stress response studies. J Biosci 43:173–187. https://doi.org/10.1007/s12038-017-9728-z

Kim SI, Veena, Gelvin SB (2007) Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. Plant J 51:779–791. https://doi.org/10.1111/j.1365-313X.2007.03183.x

Lacroix B, Citovsky V (2014) A mutation in negative regulator of basal resistance *WRKY17* of *Arabidopsis* increases susceptibility to *Agrobacterium*-mediated transient genetic transformation. F1000 Res 2:1–8. https://doi.org/10.12688/f1000research.2-33.v1

Lilly ST, Drummond RSM, Pearson MN, MacDiarmid RM (2011) Identification and validation of reference genes for normalization of transcripts from virus-infected *Arabidopsis thaliana*. Mol Plant Microbe Interact 24:294–304. https://doi.org/10.1094/MPMI-10-10-0236

Liu D, Shi L, Han C, Yu J, Li D, Zhang Y (2012) Validation of reference genes for gene expression studies in virus-infected *Nicotiana benthamiana* using quantitative real-time PCR. PLoS ONE 7:e46451. https://doi.org/10.1371/journal.pone.0046451

Livak KJ, 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. https://doi.org/10.1006/meth.2001.1262

Mallona I, Lischewski S, Weiss J, Hause B, Egea-Cortines M (2010) Validation of reference genes for quantitative real-time PCR during leaf and flower development in *Petunia hybrida*. BMC Plant Biol 10:1–11. https://doi.org/10.1186/1471-2229-10-4

Park SY, Vaghchhipawala Z, Vasudevan B, Lee LY, Shen Y, Singer K, Waterworth WM et al (2015) *Agrobacterium* T-DNA integration into the plant genome can occur without the activity of key non-homologous end-joining proteins. Plant J 81:934–946. https://doi.org/10.1111/tpj.12779

Pecinka A, Rosa M, Schikora A, Berlinger M, Hirt H, Luschnig C, Scheid OM (2009) Transgenerational stress memory is not a general response in *Arabidopsis*. PLoS ONE 4:e5202. https://doi.org/10.1371/journal.pone.0005202

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucl Acids Res 29:2002–2007. https://doi.org/10.1093/nar/29.9.e45

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. Biotechnol Lett 26:509–515. https://doi.org/10.1023/B:BILE.0000019559.84305.47

Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A (2004) Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun 313:856–862. https://doi.org/10.1016/j.bbrc.2003.11.177

Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A (2008) Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. Planta 227:1343–1349. https://doi.org/10.1007/s00425-008-0706-4

Reymond P, Weber H, Damond M, Farmer EE (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. Plant Cell 12:707–719. https://doi.org/10.1105/tpc.12.5.707

Rodrigues F, Marcelino-Guimaraes FC, Lima A, Vilela R, Margis R (2010) The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. Anal Biochem 406:185–192. https://doi.org/10.1016/j.ab.2010.07.020

Shah JM, Ramakrishnan AM, Singh AK, Ramachandran S, Unniyampurath U, Jayshankar A et al (2015) Suppression of different classes of somatic mutations in *Arabidopsis* by *vir* gene-expressing *Agrobacterium* strains. BMC Plant Biol 15:210–223. https://doi.org/10.1186/s12870-015-0595-1

Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M (2015) How good is a PCR efficiency estimate: recommendations for precise and robust qPCR efficiency assessments. Biomol Detect Quantif 3:9–16. https://doi.org/10.1016/j.bdq.2015.01.005

Talapatra S, Wagner JDO, Thompson CB (2002) Elongation factor-1 alpha is a selective regulator of growth factor withdrawal and ER stress-induced apoptosis. Cell Death Differ 9:856–861. https://doi.org/10.1038/sj.cdd.4401078

Tayeh C, Randoux B, Vincent D, Bourdon N, Reignault P (2014) Exogenous trehalose induces defenses in wheat before and during abiotic stress caused by powdery mildew. Phytopathology 104:293–305. https://doi.org/10.1094/PHYTO-07-13-0191-R

Vaghchhipawala ZE, Vasudevan B, Lee S, Morsy MR, Mysore KS (2012) *Agrobacterium* may delay plant nonhomologous end-joining DNA repair via XRCC4 to favor T-DNA integration. Plant Cell 24:4110–4123. https://doi.org/10.1105/tpc.112.100495

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N et al (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:0034.1–0034.11. https://doi.org/10.1186/gb-2002-3-7-research0034

Veena Jiang H, Doerge RW, Gelvin SB (2003) Transfer of T-DNA and Vir proteins to plant cells by *Agrobacterium tumefaciens* induces expression of host genes involved in mediating transformation and suppresses host defense gene expression. Plant J 35:219–236. https://doi.org/10.1046/j.1365-313X.2003.01796.x

Walker NJ (2002) A technique whose time has come. Science 296:557–559. https://doi.org/10.1126/science.296.5567.557

Wang M, Wang Q, Zhang B (2013) Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.). Gene 530:44–50. https://doi.org/10.1016/j.gene.2013.07.084

Woo-Lee C, Efetova M, Engelmann JC, Kramell R, Wasternack C, Ludwig-Muller J et al (2009) *Agrobacterium tumefaciens* promotes tumor induction by modulating pathogen defense in *Arabidopsis thaliana*. Plant Cell 21:2948–2962. https://doi.org/10.1105/tpc.108.064576

Wu H, Liu K, Wang Y, Wu J, Chiu W, Chen C (2014) AGROBEST: an efficient *Agrobacterium*-mediated transient expression method for versatile gene function analyses in *Arabidopsis* seedlings. Plant Methods 10:1–16. https://doi.org/10.1186/1746-4811-10-19