A spiking strategy facilitates housekeeping selection for RT-qPCR analysis under different biotic stresses in eggplant

A Spiking strategy facilitates Housekeeping Selection for RT-qPCR Analysis under different biotic stresses in Eggplant.

Valeria BARBIERATO[1], Tea SALA[1], Paola RINALDI[1], Laura BASSOLINO[1], Lorenzo BARCHI[2], Giuseppe Leonardo ROTINO[1] and Laura TOPPINO[1]*

*Corresponding author: laura.toppino@crea.gov.it

[1] CREA- Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Unità di Ricerca per l'Orticoltura, Via Paullese, 28 26836 Montanaso Lombardo, Lodi, Italy

[2] DISAFA Plant Genetics and Breeding, University of Torino, 10095 Grugliasco, Torino, Italy

ABSTRACT

Endogenous housekeeping genes are traditionally employed to normalize the expression of target genes in RT-qPCR studies. Assuming that a perfect housekeeping suitable for every condition does not exist, expression stability of the chosen reference gene should be evaluated at every new experiment. The housekeeping selection process reveals furthermore complicated and time-consuming when different conditions have to be compared in the same experimental dataset. As alternative strategy we spiked an External Reference Transcript (*ERT*) into all RNA samples of our dataset (eggplant roots subjected to different biotic stresses), and used it to normalize the expression levels of native candidate housekeeping. *ERT* expression resulted highly stable across all samples and enabled to indicate Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) as the most stable endogenous housekeeping. This result was confirmed by the use of GeNorm, Normfinder and BestKeeper algorithms. This method might be generally applied to expedite the selection process of the best reference gene.

1. INTRODUCTION

RT-qPCR (Real-Time Quantitative Reverse Transcription PCR) is the most commonly used technique for gene expression profiling due to its high sensitivity and precision (Bustin 2002). The expression level of a target gene is evaluated through the relative quantification of its mRNA against that of a reference gene called housekeeping (Livak and, Thellin et al. 1999; Schmittgen 2001; Brunner et al. 2004). The accuracy of this method strongly depends on the selection of a stably expressed reference, which ideally should exhibit little variation under a wide range of conditions such as tissue types, plant developmental stages and experimental treatments. Genes involved in basic cellular functions are the most common reference employed because they are assumed to have a constant expression pattern. However, there is a growing recognition that expression of the commonly employed housekeeping may also vary considerably (Bustin and Nolan 2004; Dheda et al. 2004; Kozera and Rapacz 2013) depending on the conditions, thus indicating that their choice as stable references may sometimes be inappropriate. When unrecognized, unexpected changes in their expression may result in biased target gene expression profiling (Dheda et al. 2005; Gutierrez et al. 2008a and 2008b) and erroneous conclusion about their actual biological action (Mascia et al. 2010). In addition, these changes often remains hidden because most experiments include only a single reference gene, which, obviously, cannot be checked itself for its stability. For these reasons, it highly recommended to carefully assess the reference expression at any new experimental condition under study

(Schmittgen and Zakrajsek. 2000; Turabelidze et al. 2010) better if with different statistical algorithms; moreover it is suggested to normalize experimental data to more reference genes (Huggett et al. 2005; Bustin et al. 2009). The identification of appropriate housekeeping genes becomes particularly difficult and time-consuming when different experimental conditions (e.g. biotic stresses) are assessed, due to the raised number of samples and genes to be compared. The number of publications describing evaluation of reference genes in model and non-model plants has markedly increased in the last decade and a set of reference genes has been proposed for almost every living species under study including tobacco (Corteleven et al. 2009, Schmidt and Delaney 2010), potato (Nicot et al. 2005; Lopez-Pardo et al. 2013), tomato (Dekkers et al. 2011; Wieczorek et al. 2013) and eggplant (Gantasala et al. 2013; Zhou et al. 2014). Many works recently focused on identification of the optimal reference gene(s) for expression profiling of molecular mechanisms related to plant responses to abiotic (Løvdal and Lillo 2009; Ma et al. 2013; Maksup et al. 2013; Park et al. 2012) and biotic stresses (Nicot et al. 2005; Libault et al. 2008; Barsalobres-Cavallari et al. 2009), in particular when caused by viruses (Mascia et al. 2010; Lilly et al. 2011) and fungi (Sestili et al. 2014), often pointing out that biotic stress, which cause severe effects on the plant metabolism, could also interfere with the expression of the most common reference genes.

The debates on the criteria for selecting the best reference are still a hot spot in the scientific community (Guénin et al. 2009), as demonstrated by the raising of work-groups focused on this argument (e.g. ERCC-External RNA controls consortium), or the development of web resources devoted to discussion about gene expression like the qPCR Forum, the portal www.Gene-Quantification.info and the database OMICTools (Henry et al., 2014). Several statistical algorithms have been set up for normalization of experimental data from gene expression analysis to assist the evaluation of housekeepings' expression stability. The algorithm and software developed for evaluation of candidate reference have their own advantages and pitfalls, but to date there is no consensus on the one that should be used (Goulao et al., 2012). One of the most commonly used is GeNorm (Vandesompele et al. 2002 and 2009), an Excel-based tool consisting in a looped pair wise comparison and geometric averaging across a matrix of candidate genes; at every cycle of comparisons, the gene corresponding to the highest M-value (less expression stability) is eliminated until the couple of the two most stable expressed genes remains. The GeNorm analysis should be ideally performed on 6 to 12 candidate genes, the reliability of the results dropping when a slighter number of candidates are considered. Moreover, the method is based on the principle that in all samples the expression ratio of two housekeeping genes remains constant and invariable, therefore is highly dependent on the sometimes-erroneous assumption that none of the analyzed genes are co-regulated. In order to reduce the risk of artificial selection of co-regulated reference genes, the GeNorm outputs need to be compared with those of other computational programs whose algorithms are less sensitive to co-regulation (Gutierrez et al. 2008b) such as Normfinder (Andersen et al. 2004) and BestKeeper (Pfaffl 2001; Pfaffl et al. 2004). The former is a model-based variance estimation that provides a direct measure, named *stability value*, of the expression variation for each housekeeping, enabling the user to evaluate the systematic error introduced when using that gene for normalization. The latter, is a free Excel-based tool that determines the optimal housekeeping gene through pair-wise correlation analysis of all pairs of candidates trough the calculation of the geometric mean of the best suited ones. When two or more computational methods are compared, often the ranking orders show slight variation especially in the middle positions (Mascia et al. 2010; Liu et al. 2012; Mafra et al. 2012; Zhou et al. 2014), rising

the necessity of a reasoned comparison of the different output to gain a consensus ranking order. All this translates once again in a more and more time-consuming preliminary evaluation of the panel of chosen candidates that must be performed at any new experiment. Smith et al. (2003), proposed an alternative method for normalization of RT Quantitative PCR data across different experimental conditions, which uses an exogenous sequence (RuBisCo transcript) as an External Reference Transcript (ERT) in human. Contrary to what often happens when new technologies developed in one field is rapidly transferred also to other fields of studies, the spiking technique, which was broadly employed both in humans and animals, is still in a preliminary stage of spreading in experiments dealing with plants. In fact, according to our knowledge, bibliography dealing with spiking strategy in plant is very poor as only two papers are reported (McMaugh and Lyon, 2003, Czechowski et al., 2005) while the most common pipeline for the selection strategy of reference genes remains the traditional statistical analysis through one or more algorithms. On the other side, in animals and humans the ERT spiking strategy is widely employed but mainly as external normalizer itself (and the only one utilized). This, although leading to reliable results, may be time and costs consuming depending on the dimension of the panel to be considered because the ERT must be added to all the samples.

In this paper we propose a different strategy. Indeed, we will use the *ERT* as a normalizer for a panel of native candidate reference genes chosen from literature, with the aim to speed up the selection of the most suitable housekeeping to be used in an experiment dealing with the expression profiling of genes involved in the response to different fungal pathogens in eggplant (Barbierato et al., 2016). We confirmed the reliability of this approach by testing the stability of gene expression with the three algorithms GeNorm, Normfinder and BestKeeper.

2. MATERIAL AND METHODS

2.1 Plant materials

Seed-derived plantlets of an advanced introgressed line of eggplant (Toppino et al., 2008), were grown in greenhouse and individually inoculated at the 3-4th true leaf stage, according to the root-dip method described in Cappelli et al., (1995) and Barbierato et al., (2016) with a conidia suspension of *Fusarium* (sample set named "F"), *Verticillium* ("V"), or both fungi together ("M"), while root dipping in water ("C") was used as mock-inoculation. Both inoculated and mock-inoculated eggplant roots were harvested at 0, 4, 8 and 24 hours after artificial inoculation, frozen in liquid N_2 and stored at -80 °C.

2.2 RNA isolation, ERT addition and reverse transcription

Total RNA was purified from 100 mg of powdered tissue using the RNeasy® plant RNA extraction kit (Qiagen, Clifton Hill, Victoria, Australia) according to the manufacturer's instructions. RNA purity and quantification was determined with Nanodrop (Thermo Scientific Wilmington, USA). The Kanamycin 1.2 kb Control RNA (KANAr) was found as positive control RNA in the Promega kit Improm-II reverse transcription system, therefore we used this gene as ERT because it was available, cheap and of good quality (considered that it is used as positive control in a kit it guarantees good transcription quality). Moreover, is of average length with respect to most part of transcripts therefore should be transcribed with the same quality as most part of the genes in the sample. Thus, 30 ng of Kanamycin 1.2 kb Control RNA (Kan 1.2; Promega, Madison, WI, USA, from here called *KANAr*) was then added to 3000 ng of total RNA (ERT/total RNA =1/100) of each sample before DNase Treatment using 1U/µl of

RQ1 RNase-Free DNase (Promega). Reverse transcription was then performed with the ImProm-II™ Reverse Transcription System (Promega, Madison, WI, USA), in a total volume of 20 µL.

2.3 Primer design

Sequences of *GAPDH* (*glyceraldehyde-3-phosphate dehydrogenase*, AB110609.1), *EF1* (*elongation factor α1*, X14449.1), *TUB* (*beta-tubulin*, DQ205342.1), *PP2Acs* (*catalytic subunit of protein phosphatase 2A*, AY325817.1), *18S* (*18S rRNA*, AJ421474) and *UBI* (*ubiquitin*, BT012698.1), were retrieved from the DFCI-TGI (Tomato Gene Index) EST database (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=tomato). Primers to amplify each gene in eggplant (Table 1) on the basis of the homologous tomato sequences plus the *KANAr* ERT were designed using PRIMER3 software (170 bp maximum length, optimal Tm at 59°C, GC% between 40% and 60%). Primer's specificity was confirmed by checking the correct PCR product sizes on 1% agarose gel and then by sequencing of the amplicons.

2.4 Two step real-time quantitative PCR

Real-time amplifications using SYBR Green (IQTM Supermix Master Bio-Rad) were performed in a Rotor-Gene RG-6000 thermal cycler (Corbett Research) according to the manufacturers. The efficiencies (E) of the primer specific amplifications and the correlation coefficient (R^2) of linear regression models were derived from qPCR standard curves generated using a 4-fold serial dilution of pooled cDNA (obtained mixing equal proportion of all 16 cDNA samples) with the Rotor gene software according to the equation: $E= 10^{[-1/slope]}$. Primer conditions were optimized by determining the correspondent best annealing temperature and primer concentration. The cycling conditions were set as follows: initial denaturation step of 95°C for 3 min, 50 cycles of 15s at 95°C + 40s at 59°C, followed by melting analysis to insure the detection of one gene-specific peak and the absence of primer-dimer peaks. The sets of samples V, F, M, and C, at the four timings after inoculation (T0, T4, T8, T24) were used for qRT-PCR expression analyses of each candidate gene and the ERT. Each reaction was run in duplicate in two biological replicates, so that each gene was analyzed in a total of 64 samples.

2.5 Data processing

Expression levels were determined by Ct values. The threshold was set at 0.004 fluorescent units, and the threshold cycle (Ct) values were plotted against the starting template concentration. The average Ct-values obtained from each duplicate qRT-PCR reaction were converted in relative expression levels for subsequent analysis with geNorm software version 3.5 (http://allserv.ugent.be/;jvdesomp/genorm/index.html) according to the manufacturer's instructions; the Ct data of all the samples were used as input in Normfinder (moma.dk/normfinder-software) Version 20 and BestKeeper (www.gene-quantification.de/bestkeeper.html) Version 1 all according to the manufacturer's instructions. All other statistics were performed with Microsoft Excel and Past software.

3. RESULTS

3.1 Pre-analytical assessment of the panel of candidate genes

Primer pairs were designed on the basis of the consensus sequences retrieved from six tomato housekeeping candidates *TUB*, *EF1*, *UBI*, *PP2Asc*, *18S*, and *GAPDH* frequently used in experiments of pathogen-mediated stress induction in plant (Løvdal and Lillo, 2009). Despite the use of several alternative primer pairs, *UBI* sequence still revealed un-specific amplification and was excluded from the study (data not shown). PCR-based screening of the

five genes on eggplant cDNA samples confirmed their expression in roots (data not shown). All amplicons were sequenced for verification and all shared more than 96% of identity with the tomato consensus sequences (Table 1). The linear R^2 for the five pairs of primers had a range of 0.989-0.999, and all the genes displayed good values of PCR efficiency (E) varying from 0.88 to 1.01. The identity of each qPCR product was confirmed by the presence of a single peaks in melting curve analysis.

3.2 Evaluation of the expression stability of the External Reference Transcript

A fixed amount of *KANAr* mRNA transcript was added in a constant ratio (1:100) to each sample of RNA before DNAse treatment and retrotranscription, thus it was incorporated together with all the endogenous eggplant transcripts in qRT-PCR. qRT-PCR standard curve generated for *KANAr* showed a E = 1.05. A qRT-PCR analysis was performed to assess the expression stability of *KANAr* among all the sets of samples. The *KANAr* amplification (Figure 1) gave a mean Ct-value of 5.96, with values varying from 5.16 to 7.33 (sd ±0.62), thus revealing a very stable expression which remained nearly unaltered among all the samples derived from different experimental conditions. As the expression of *KANAr* was highly reliable, this transcript was considered suitable to be used as ERT to normalize the expression data of the eggplant native candidate housekeeping genes. Considering also the RT-qPCR analysis in which it was used as reference with respect to the other candidates, ERT transcript was amplified in 228 replicates, ensuing an even better stability value (Ct 6.02, se ±0.06).

3.3 Evaluation of expression levels of the native candidates.

Raw qRT-PCR expression levels of the five candidates showed very marked variation (Figure 1), ranging from Ct values of 2.14 to 24.62, with *18S* having the highest transcriptional levels (Ct values from 2.14 to 4.68, with a mean of 3.50 and sd ±0.68) and *PP2Acs* the lowest ones (20.31-24.62, with average of 22.23 and sd ±1.18). *TUB* showed the widest expression variability across all samples, with Ct values ranging from 15.1 to 20.61 (average 17.02, sd ±1.32), while *GAPDH* had the smallest variation (range: 15.30-17.30; average 16.19; sd ±0.66). *EF1* ranged from 12.61 to 18.02, with a mean of 14.35 and sd ±1.18.

3.4 Evaluation of the candidate reference genes stability with respect to the *ERT*

For each gene, ΔCt comparisons with respect to *KANAr* are shown in figure 2. The five genes displayed a wide range of relative expression levels with the mean ΔCt values ranging from -3.6 (*18S*) to +16.8 (*PP2Acs*). ERT normalization enabled to indicate *GAPDH* as the gene that displayed the slighter variability among all the samples (ΔCt-value ranging between 9.7 and 10.6). This gene is often reported in literature as the best housekeeping in several tested conditions and species (Barsalobres-Cavallari et al. 2009; Iskandar et al. 2004; Wan et al. 2011) including eggplant (Zhou et al. 2014; Gantasala et al. 2013), although sometimes was reported as highly reliant on the experimental conditions (Expòsito-Rodriguez et al. 2008; Kozera & Rapacz 2013; Løvdal and Lillo 2009; Wieczorek et al 2013). Slighter variation with respect to *GAPDH* was evidenced for *18S* (ΔCt-value ranging from -1.3 and 3.6). Thus, considering that similar expression levels of reference and target genes are an important issue in RT-qRCR normalization (Bustin et al. 2009), its high abundance make it a good choice when comparing highly expressed target genes. However, its inappropriateness as reference was reported especially under stress condition (Lopez-Pardo et al. 2013), probably due to its specific role in ribosomal activity and translation, which are both severely affected by these conditions. Often reported as a stable gene (Løvdal and Lillo, 2009, Schmidt and Delaney 2010) *PP2Acs* showed in our dataset expression variation after *Fusarium* and mixed inoculation (ΔCt-

value between 14.7 and 16.7), while no differences were detected after *Verticillium* and mock inoculations. Our results confirm that different pathogens may induce divergent alteration in the expression of reference genes probably as a consequence of differences in the molecular and biological features of the pathogenetic processes (Liu et al. 2012, Whitham et al. 2003). *TUB* showed a wide variability across all samples (ΔCt-value from 9.9 and 12.3). In eggplant, it was already stated as poorly stable under cold/heat abiotic stress (Gantasala et al. 2013) and therefore may be definitely considered poorly recommendable as reference in experiments involving stress-induced eggplant tissues. The highest and widest ΔCt variability was detected for *EF1*, which proved to be severely affected by the four types of inoculations (ΔCt between 7.02 and 10.4): although often indicated as good reference, discrepancies on its evaluation have been already documented in tomato, potato, tobacco and melon and also eggplant (Exposito-Rodriguez et al. 2008; Løvdal and Lillo 2009; Liu et al. 2012; Sestili et al. 2014; Zhou et al. 2014) depending on the different conditions and also on the algorithms utilized. (Mascia et al. 2012; Wieczorek et al. 2013).

3.5 Evaluation of the candidate reference genes stability through different algorithms.

The 64 Ct-values obtained from the qRT-PCR analysis of both *KANAr* and the panel of native candidates were therefore analyzed with GeNorm, BestKeeper and NormFinder algorithms.

3.5.1 GeNorm analysis

The panel of chosen candidates included genes involved in basal metabolism, but distantly related with respect to their metabolic function; therefore they were suitable for the GeNorm algorithm test. Stepwise exclusion of the gene with the highest M-value allowed ranking of the candidate genes according to their expression stability thus enabling the selection of the couple of genes that showed the most stable expression with respect to each other. The outcome from the most stable (lowest M-value) to the least stable (highest M-value) was: *GAPDH/KANAr* < *18S* < *PP2Acs* < *TUB* < *EF1* (Table 2 and Figure 3). The GeNorm algorithm also stated at 4 (*KANAr*, *GAPDH*, *18S* and *PP2Acs*) the optimal number of genes (Vn) required for accurate normalization, based on the pair wise variation between two sequential normalization factors containing an increasing number of genes (Vn/Vn+1), as the V4/5 value of 0.165 obtained was the closest to the cut-off threshold of 0.15.

3.5.2 Normfinder analysis

By using Normfinder, the reference genes were ranked according to their expression stability with the lower *Stability Value* corresponding to the more stable reference gene. The measure of expression variation indicated *KANAr* as the best reference gene, followed by *GAPDH*, *PP2Acs*, *18S*, *TUB1*, and *EF1* as reported in table 2 and in Figure 4.

3.5.3 BestKeeper analysis

The BestKeeper output confirmed *KANAr* as the most stable housekeeping gene, while *18S* was the second best gene followed by *GAPDH* (Table 2). The ranking order of the reference genes, based on the standard deviation of the absolute regulation coefficients (std dev. ± x fold value), from the most stable to the less one, was: *KANAr* and *18S*, *GAPDH*, *TUB1*, *EF1*, *PP2Acs*. The three different algorithms (GeNorm, Normfinder and BestKeeper) confirmed the reliability our strategy for choosing a reference gene through the spiking *KANAr* as ERT because they gave coherent ranking outcomes with only slight variation in its middle part. GeNorm and Normfinder assigned the best expression stability in eggplant roots affected by fungal inoculations to *KANAr*, *GAPDH* and

18S genes, and discarded *TUB* and *EF1* as the worst choice. Bestkeeper preferred *18S* gene to *GAPDH* in the ranking order, although their stability values are very close, and put *TUB* in the last position. After accurate comparison of the three software's outputs, a consensus ranking order may be ventured, which indicates, from the most to the least affordable *KANAr, GAPDH , 18S, PP2Acs, EF1* and *TUB* reference gene.

4. DISCUSSION

Experimental conditions could have strong effects on the basal plant metabolism and, consequently, interfere with expression of the so-called housekeeping genes (Schmittgen and Zakrajsek, 2000), therefore the choice of a suitable reference genes at every new experiment represents the major critical bottleneck for an accurate evaluation of gene expression. The fact that in our previous work the gene *actin*, one of the most commonly used housekeeping reported in literature, was identified among a panel of differentially expressed genes in response to fungal wilt (Barbierato et al. 2016) and therefore unreliable as reference, prompted us to explore an alternative strategy to ease and speed up the selection process of the most suitable housekeeping. This strategy is based on the assumption that a synthetic alien RNA (also called RNA spike-ins), added to the extracted plant RNA prior to reverse transcription can act as valuable standardization tool in real-time PCR experiments as it is completely independent from the plant biological process adopted (Gilsbach et al. 2006), and therefore it can be used not only as reference gene itself but also as a normalizer to evaluate the suitability of the native reference genes. This strategy has been successfully used in human (Smith et al. 2003; O'Shaughnessy et al. 2011) and animals (Lanes et al. 2012) gene expression analysis and recently a suit of 96 spike-in unique RNA control sequences has been set up (ERCC, https://www.thermofisher.com/order/catalog/product/4456740) to better analyze the RNAseq experiments; while, up to now, spike-in RNA has been very sporadically applied in plants (McMaugh and Lyon, 2003, Czechowski et al., 2005). The possibility of using an ERT as external control to check the native candidate housekeeping has been successfully explored in our experimental dataset (cDNA of eggplant roots samples inoculated with different fungal pathogens), enabling the identification of *GAPDH* as the most suitable reference gene across the different conditions. The reliability and goodness of the spiking strategy was confirmed through comparison of the ranking order retrieved with the *KANAr* spiking strategy with the outputs of three different algorithms. Moreover, according to GeNorm output, the use of 4 housekeeping genes was stated as the minimum number for a sufficiently robust validation of the expression data, while the ERT strategy enabled the selection of *GAPDH* as the single native housekeeping gene suitable for accurate qRT-PCR analysis in our experimental conditions. The spiking strategy is simple, fast and may find a general application in any qPCR-based study; in our laboratory, it is routinely utilized as preliminary step at any new experiment in which one or more conditions have to be compared. Recently, this approach allowed the selection of *18S* as the best housekeeping gene to compare the expression levels of genes involved in glycoalkaloids biosynthesis among fruit tissues of eggplant at different developmental and ripening stages (unpublished). Moreover, both *GAPDH* and *18S* have been already used in our lab as references in several RT-qPCR experiments (Docimo et al, 2015; e.g. candidate genes study on biotic stress-unpublished) and their stability across the different samples and experimental conditions have been confirmed.

Figure legends

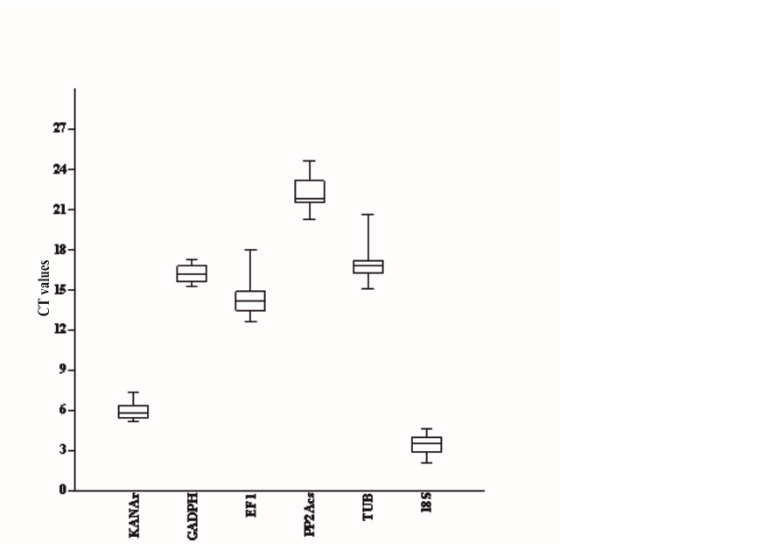

Figure 1 Box-plot representing the raw expression profile of the candidate reference genes and of the External Reference Transcript (ERT) across all samples set: for each gene, Median Ct values are represented as lines, 25-75 percentiles as boxes and range of Ct values as whiskers

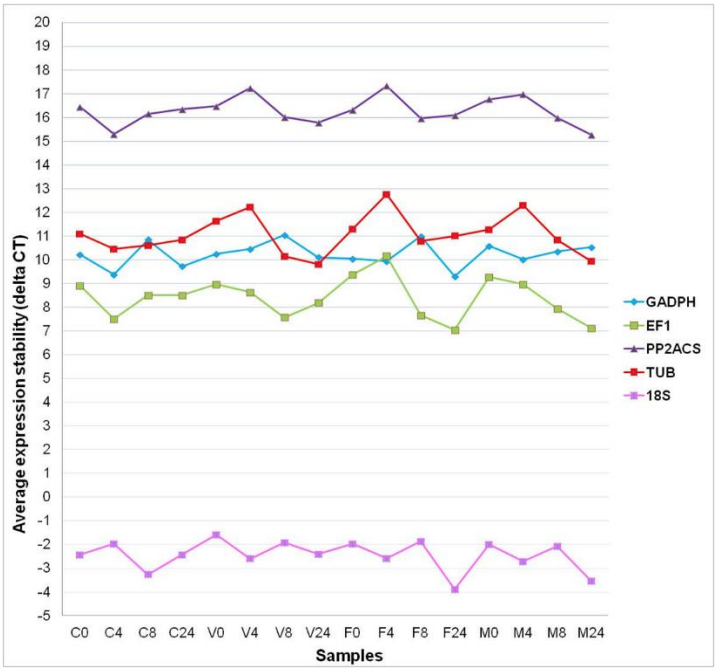

Figure 2: Expression profiles of the candidate housekeeping genes after different types of fungal inoculation (C: control, V: *Verticillium*, F: *Fusarium*, M: Mixed) and timings (0, 4, 8, 24 hours) using the ERT as reference gene. Each data point represents the average of two experiments (performed in duplicate) and the error bars indicate the standard error of the mean of four replicates. *EF1* (green-rimmed squares), *TUB* (red-circles), *PP2Acs* (violet-triangles) *18S* (pink-squares), *GAPDH* (blue-diamonds)

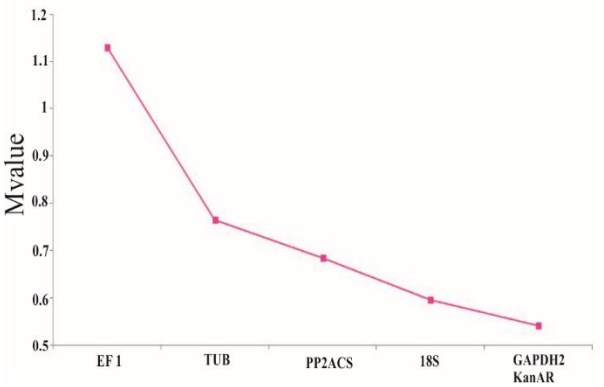

Figure 3: GeNorm output for average expression stability values of reference genes from the most stable (lowest M-value) to the least stable (highest M-value): elongation factor 1- α (EF1), β tubulin (TUB) , catalytic subunit of phosphatase 2A (PP2Acs), 18S rRNA (18S), glyceraldeyde-3-phosphate dehydrogenase (GAPDH) and Kanamycin resistance (KANAr).

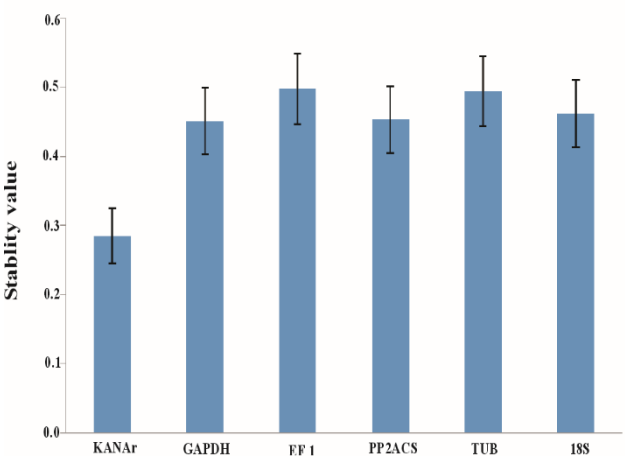

Figure 4: Histogram obtained from the results of Normfinder program; Stability value and Standard error are calculated by the algorithm and indicates *KANAr* and *GAPDH* (lowest stability values) as the most stable housekeeping genes. Abbreviations: *elongation factor 1- α* (*EF1*), *β tubulin* (*TUB*), *catalytic subunit of phosphatase 2A* (*PP2Acs*), *18S rRNA* (*18S*), *glyceraldeyde-3-phosphate dehydrogenase* (*GAPDH*) and *Kanamycin resistance* (*KANAr*).

Gene	Genbank accession number	Primer sequence	Amplicon lenght (bp)	Ta (°C) (1)	% Identity (2)	E (%) (3)	R^2 (4)
KANAr	FJ621586.1	5' GATGTTGGACGAGTCGGAAT 3' 3' CGAGCATCAAATGAAACTGC 5'	157	59	100%	105	0.979
GAPDH	AB110609.1	5' GGTGCCAAGAAGGTTGTGAT 3' 3' CGTTGTGCAACTAGCATTGG 5'	120	59	96%	101	0.995
18S	AJ421474	5' ATGATAACTCGACGGATCGC 3' 3' CTTGGATGTGGTAGCCGTTT 5'	169	59	98%	100	0.995
PP2Acs	AY325817.1	5' GGACTCTCACCATCCCTTGA 3' 3' GAGGTGATATTCCCCAACCA 5'	136	59	99%	96	0.991
TUB	DQ205342.1	5' CCAGACAGGATGATGCTCAC 3' 3' GCTTCGTTGTCAAGGACCAT 5'	140	59	96%	99	0.989
EF1	X14449.1	5' ACCAAGATTGACAGGCGTTC 3' 3' TGGAGGGTATTCAGCAAAGG 5'	132	59	100%	88	0.999
UBI	BT012698.1	5' GGACGGACGTACTCTAGCTGAT 3' 3' AGCTTTCGACCTCAAGGGTA 5'	134	59	n.d.	n.d.	n.d.

Table 1. *Solanum melongena* candidate reference gene description. The gene bank accession number, the code, primer sequences**,** amplified fragment length, (1) Optimal annealing temperature. (2) Percentage sequence identity between the amplicon and the corresponding-homolog-tomato sequence from Genbank. (3) Measure of the real-time PCR reaction efficiency E (calculated by standard curve method). (4) Reproducibility of the real-time PCR reaction. n.d. = no data because the gene was excluded from the study are reported.

Gene	Ranking order vs KANAr (ΔCt-range) [5]	Ranking order Genorm (M-value) [6]	Ranking order Normfinder (stability value) [7]	Ranking order Bestkeeper (sd [± x-fold]) [8]
KANAr	-	-	1 (0,285± 0,04)	1 (1,421955)
GAPDH	1 (9.7-10.6)	1-2 (0.54)	2 (0,451±0,05)	3 (1,499745)
18S	2 (-3.6 -1.3)	3 (0.59)	4 (0,461±0,05)	2 (1,478101)
PP2Acs	3 (14.7-16.7)	4 (0.68)	3 (0,453±0,05)	5 (1,93559)
TUB	4 (9.9-12.3)	5 (0.76)	5 (0,494±0,05)	6 (1,954305)
EF1	5 (7.0-10.4)	6 (1.13)	6 (0,498±0,05)	4 (1,831924)
UBI	n.d.	n.d.	n.d.	n.d.

Table 2 The rank order of each reference gene for the tested algorithms. Candidate reference genes for normalization ranked according to their expression stability with respect to the ERT (5), to GeNorm (6) (calculated as the average M-Value after stepwise exclusion of the worst scoring gene), Normfinder (7) (the most stable housekeeping genes indicated by its lowest Stability value), and Bestkeeper (8) (ranking order based on the standard deviation of the absolute regulation coefficients) algorithms.

References

Andersen CL, Jensen JL and Ørntoft TF (2004). Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. Cancer Research 64:5245–5250

Barbierato V, Toppino L, Rinaldi P, Sala T, Bassolino L, Valè G, Ferrarini A, Delledonne M, Bagnaresi P and Rotino GL. (2016). Phenotype and gene expression analyses of the Rfo-sa1 resistant aubergine interaction with Fusarium oxysporum f. sp. melongenae and Verticillium dahliae. Plant Pathology. Version of Record online DOI: 10.1111/ppa.12518

Barsalobres-Cavallari CF, Severino FE, Maluf MP and Maia IG. (2009). Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. BMC Molecular Biology 10,1. doi:10.1186/1471-2199-10-1.

Brunner AM, Yakovlev IA and Strauss SH. (2004). Validating internal controls for quantitative plant gene expression studies. BMC Plant Biology 4, 14. DOI: 10.1186/1471-2229-4-14

Bustin SA. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. Journal of Molecular Endocrinology 29: 23–39.

Bustin SA and Nolan T. (2004). Pitfalls of Quantitative Real-time reverse transcription polymerase chain reaction. Journal of Biomolecular Techniques 15: 155–166.

Bustin SA Benes V, Garson JA, Hellemans J, Huggett JF, Kubista M and Wittwer CT (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry 55(4): 611–622.

Cappelli C, Stravato VM., Rotino GL and Buonaurio R. (1995). Sources of resistance among Solanum spp. to an Italian isolate of *Fusarium oxysporum* f. sp. *melongenae*. In: Andra`Sfalvi, A; Moo` r, A; Zatyko` (eds) Proceedings on the EUCARPIA 9th Meeting on Genetics and Breeding of Capsicum and Eggplant. SINCOP, Budapest, p. 221–224.

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R. Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. Plant Physiology. 2005;139(1):5-17. doi:10.1104/pp.105.063743

Cortleven A, Remans T, Brenner WG and Valcke R. (2009). Selection of plastid- and nuclear-encoded reference genes to study the effect of altered endogenous cytokinin content on photosynthesis genes in *Nicotiana tabacum*. Photosynthesis Research 102(1): 21–9.

Dekkers BJ, Willems L, Bassel GW, van Bolderen-Veldkamp RP, Ligterink W, Hilhorst HW and Bentsink L. (2011). Identification of reference genes for RT–qPCR expression analysis in Arabidopsis and tomato seeds. Plant & Cell Physiology 53: 28–37.

Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A (2004). Validation of housekeeping genes for normalizing RNA expression in real-time PCR. Biotechniques 37(1):112–119.

Dheda K, Huggett JF, Chang JS, Kim LU, Bustin SA, Johnson MA, Rook G and Zumla A. (2005). The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Analalytical Biochemestry 344(1): 141-143.

Docimo T, Francese G, Ruggiero A, Batelli G, De Palma M, Bassolino L, Toppino, L, Rotino GL, Mennella G. (2015) Phenylpropanoids Accumulation In Eggplant Fruit: Characterization Of Biosynthetic Genes And Regulation By A Myb Transcription Factor. Frontiers In Plant Science, 6:1233 . Doi: 10.3389/Fpls.2015.01233

Expósito-Rodríguez M, Borges AA, Borges-Pérez A and Pérez JA. (2008). Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biology 8, 131. doi:10.1186/1471-2229-8-131.

Gantasala NP, Papolu PK., Thakur PK, Kamaraju D, Sreevathsa R and Rao U. (2013). Selection and validation of reference genes for quantitative gene expression studies by real-time PCR in eggplant (*Solanum melongena* L). BMC Research Notes 6(1): 312. doi:10.1186/1756-0500-6-312.

Gilsbach R, Kouta M, Bönisch H, Brüss M. (2006). Comparison of in vitro and in vivo reference genes for internal standardization of real-time PCR data. BioTechniques 40(2): 173–177.

Goulao LF., Fortunato AS and Ramalho J. (2011). Selection of Reference Genes for Normalizing Quantitative Real-Time PCR Gene Expression Data with Multiple Variables in Coffea spp. Plant Molecular Biology Reporter 30(3): 741–759.

Guénin 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. Journal of Experimental Botany 60(2): 487–493.

Gutierrez L, Mauriat M, Guénin S, Pelloux J, Lefebvre JF, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C and Van Wuytswinkel O. (2008a). 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 Biotechnology Journal 6(6): 609–618.

Gutierrez L, Mauriat M, Pelloux J, Bellini C and Van Wuytswinkel O. (2008b). Towards a systematic validation of references in real-time rt-PCR. The Plant Cell 20(7) and 1734–1735.

Henry VJ, Bandrowski AE, Pepin AS, Gonzalez BJ, Desfeux A. (2014). OMICtools: an informative directory for multi-omic data analysis. Database: The Journal of Biological Databases and Curation 2014 (13): 1–5. doi:10.1093/database/bau069.

Huggett JF, Dheda K, Bustin SA and Zumla A. (2005). Real-time RT-PCR normalisation; strategies and considerations. Genes and Immunity 6(4): 279–84.doi:10.1038/sj.gene.6364190.

Iskandar H.M, Simpson RS, Casu RE, Bonnett GD, Maclean DJ and Manners JM. (2004). Comparison of Reference Genes for Quantitative Real-Time Polymerase Chain Reaction Analysis of Gene Expression in Sugarcane. Plant Molecular Biology Reporter 22: 325–337.

Kozera B, Rapacz M. (2013) Reference genes in real-time PCR. Journal of Applied Genetics 54:391-406. Doi:10.1007/s13353-013-0173-x.

Lanes CFC, Fernandes JMO, Kiron V, Babiak I. (2012). Profiling of key apoptotic, stress, and immune-related transcripts during embryonic and postembryonic development of Atlantic cod (*Gadus morhua L.*). Theriogenology 78(7): 1583–1596. doi:10.1016/j.theriogenology.2012.07.003.

Libault M, Thibivilliers S, Bilgin DD, Radwan O, Benitez M, Clough SJ and Stacey G. (2008). Identification of four soybean reference genes for gene expression normalization. The Plant Genome 1(1): 44-54.

Lilly ST, Drummond RSM, Pearson MN and Macdiarmid RM. (2011). Identification and Validation of Reference Genes for Normalization of Transcripts from Virus-Infected *Arabidopsis thaliana*, Molecular Plant-Microbe Interaction 24:294–304.

Liu D, Shi L, Han C, Yu J, Li D and Zhang Y. (2012). Validation of reference genes for gene expression studies in virus-infected *Nicotiana benthamiana* using quantitative real-time PCR. PloS One, 7(9), e46451. doi:10.1371/journal.pone.0046451

Livak KJ and Schmittgen TD. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCT Method. Methods 25(4): 402-408.

Lopez-Pardo R, Ruiz de Galarreta JI and Ritter E. (2013). Selection of housekeeping genes for qRT-PCR analysis in potato tubers under cold stress. Molecular Breeding 31(1): 39–45..

Løvdal T, Lillo C. (2009). Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. Analytical Biochemistry 387(2): 238–242.

Ma S, Niu H, Liu C, Zhang J, Hou C and Wang D. (2013). Expression Stabilities of Candidate Reference Genes for RT-qPCR under Different Stress Conditions in Soybean. PLoS one 8(10). doi: 10.1371/journal.pone.0075271

Mafra V, Kubo KS., Alves-Ferreira M, Ribeiro-Alves M, Stuart RM, Boava LP, Machado MA. (2012). Reference genes for accurate transcript normalization in citrus genotypes under different experimental conditions. PloS One, 7(2), e31263. doi: 10.1371/journal.pone.0031263

Maksup S, Supaibulwatana K and Selvaraj G. (2013). High-quality reference genes for quantifying the transcriptional responses of *Oryza sativa* L. (ssp. indica and japonica) to abiotic stress conditions. Chinese Science Bulletin 58(16): 1919–1930. doi:10.1007/s11434-013-5726-1.

Manoli A, Sturaro A,Trevisan S, Quaggiotti S and Nonis A. (2012) Evaluation of candidate reference genes for qPCR in maize. Journal of Plant Physiology 169(8): 807-815.

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. Molecular Plant Pathology 11(6): 805–816.

McMaugh SJ, Lyon BR. (2003). Real time quantitative RT-PCR assay of gene expression in plant roots during Fungal pathogenesis. BioTechniques 34(5): 3–6.

Nicot N, Hausman JF, Hoffmann L and Evers D. (2005). Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. Journal of Experimental Botany 56(421): 2907–2914.

O'Shaughnessy PJ, Monteiro A and Fowler PA. (2011). Identification of stable endogenous reference genes for real-time PCR in the human fetal gonad using an external standard technique. Molecular Human Reproduction 17(10): 620–5. doi:10.1093/molehr/gar032.

Park SC, Kim YH, Ji CY, Park S, Jeong J, Lee HS, Kwak SS. (2012). Stable Internal Reference Genes for the Normalization of Real-Time PCR in Different Sweetpotato Cultivars Subjected to Abiotic Stress Conditions. PLoS ONE 7(12). doi :10.1371/journal.pone.0051502

Pfaffl MW. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research 29(9): 45e–45. doi:10.1093/nar/29.9.e45

Pfaffl MW., Tichopad A, Prgomet C and Neuvians T.P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. Biotechnology Letters 26(6): 509–515.

Sestili S, Sebastiani MS., Belisario A and Ficcadenti N. (2014). Reference gene selection for gene expression analysis in melon infected by *Fusarium oxysporum* f.sp. *melonis*. Journal of Plant Biochemistry and Biotechnology 23(3): 238–248. doi:10.1007/s13562-013-0207-9.

Schmidt GW and Delaney SK. (2010). Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. Molecular Genetics and Genomics 283(3): 233–241. doi:10.1007/s00438-010-0511-1

Schmittgen TD and Zakrajsek BA. (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. Journal of Biochemical and Biophysical Methods 46(1-2): 69–81. doi:10.1016/S0165-022X(00)00129-9

Smith RD, Brown B and Ikonomi PSA. (2003). Exogenous Reference RNA for normalization of Real Time Quantitative PCR. BioTechniques 34(1): 88–91.

412 Toppino L, Valè G and Rotino GL. (2008). Inheritance of Fusarium wilt resistance introgressed from *Solanum*
413 *aethiopicum* Gilo and Aculeatum groups into cultivated eggplant (*S. melongena*) and development of associated PCR-
414 based markers. Molecular Breeding 22(2): 237-250.

415 Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A and Heinen E. (1999).
416 Housekeeping genes as internal standards: use and limits. Journal of Biotechnology 75:291–295.

417 Turabelidze A, Guo S and DiPietro LA. (2010). Importance of housekeeping gene selection for accurate reverse
418 transcription-quantitative polymerase chain reaction in a wound healing model. Wound Repair and Regeneration: Official
419 Publication of the Wound Healing Society & the European Tissue Repair Society 18(5): 460–466.

420 Vandesompele J, Preter KDe, Poppe B, Roy N Van and Paepe ADe. (2002). Accurate normalization of real-time
421 quantitative RT -PCR data by geometric averaging of multiple internal control genes. Genome Biology 3(7): 1–12.

422 Vandesompele J, Kubista M and Pfaffl MW. (2009) Reference gene validation software for improved normalization. In
423 Real-time PCR: current technology and applications. Caister Academic Press Vol. 4, pp. 47–64. , 47-64.

424 Wan H, Yuan W, Ruan M, Ye Q, Wang R, Li Z, Zhou G, Yao Z, Zhao J, Liu S and Yang Y. (2011). Identification of
425 reference genes for reverse transcription quantitative real-time PCR normalization in pepper (*Capsicum annuum* L.).
426 Biochemical and Biophysical Research Communications, 416(1-2): 24–30.

427 Whitham SA, Quan S, Chang H, Cooper B, Estes B, Zhu T, Wang X and Hou Y. (2003). Diverse RNA viruses elicit the
428 expression of common sets of genes in susceptible *Arabidopsis thaliana* plants, Plant Journal. 33(2):271–283.

429 Wieczorek P, Wrzesinska B and Obrepalska-Steplowska A. (2013). Assessment of reference gene stability influenced
430 by extremely divergent disease symptoms in *Solanum lycopersicum* L. Journal of Virological Methods 194(1): 161–168.

431 Zhou X, Liu J and Zhuang Y. (2014). Selection of appropriate reference genes in eggplant for quantitative gene expression
432 studies under different experimental conditions. Scientia Horticulturae 176: 200–207.