DOI: 10.1007/s10535-017-0754-3 BIOLOGIA PLANTARUM 62 (1): 89-96, 2018

Assessment of reference genes for real-time quantitative PCR normalization in *Ilex paraguariensis* leaves during drought

R.M. ACEVEDO[1], E.H. AVICO[1], O.A. RUIZ[2], and P.A. SANSBERRO[1]*

Laboratorio de Biotecnología Aplicada y Genómica Funcional, Instituto de Botánica del Nordeste, Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste, W3402BKG, Corrientes, Argentina[1]
Unidad de Biotecnología 1, Instituto de Investigaciones Biotecnológicas-Instituto Tecnológico de Chascomús, B7130IWA, Chascomús, Buenos Aires, Argentina[2]

Abstract

Reverse transcription of RNA followed by real-time quantitative PCR (qPCR) is to date the most reliable method for gene expression studies. However, to control the errors introduced along the numerous experimental procedures, it requires a normalization using internal reference genes with stable expression. To address this issue, nine candidate reference genes were investigated in *Ilex paraguariensis* leaves subjected to water stress. To facilitate the selection, we analysed the real-time qPCR data with three different software programs. The obtained results support the conclusion that *RNA polymerase associated protein rtf1 homolog* (*RTF*) combined with any of the following pairs is the most suitable triad of genes to compute a normalization factor: *elongation factor 1-alpha + tubulin alpha chain* (*EF1a + α-Tub*), *actin + cyclophilin 38* (*ACT + CYP38*), or *cyclophilin 38 + vacuolar protein sorting-associated protein 18 homologs* (*CYP38 + VPS*). Our analysis constitutes the first in-depth study to identify the appropriate reference genes for the quantification of transcription in *Ilex paraguariensis* leaves during drought and provides essential information for further gene expression studies in this tree species.

Additional key words: abiotic stresses, gene expression, house-keeping genes.

Introduction

Ilex paraguariensis A. St.-Hil. (*Aquifoliaceae*), a perennial tree species native to the South American rainforest, is cultivated by agroforestry systems and monoculture in NE Argentina, SE Brazil, and Paraguay to prepare a mate-tea from its leaves and young shoots (Gottlieb *et al*. 2005). This beverage is greatly appreciated for its taste, its stimulation effect due to the caffeine content, and some medicinal properties (Heck and Mejia 2007, Lima *et al*. 2014). Commercial crops are usually subjected to water stress during the growth season, which negatively affects plant growth and yield (Sansberro *et al*. 2004). To manage drought, plants adopt a number of physiological and metabolic strategies (Acevedo *et al*. 2013, Kim *et al*. 2015), which are the result of the interaction between the genotype and the environment. The stress response and acclimatization of plants to water deficit cause a complex readjustment in the expression of a large number of genes (Shinozaki and Yamaguchi-Shinozaki 2007, Singh and Laxmi 2015). The isolation and characterization of these genes constitute a reliable way to understand the mechanisms underlying water stress (Acevedo *et al*. 2016). To quantify the expression of a particular gene, reverse transcription of RNA followed by real-time PCR quantitative (qPCR) is the most commonly used approach because of its high sensitivity, specificity, and large dynamic range

Submitted 8 February 2017, *last revision* 26 April 2017, *accepted* 11 May 2017.
Abbreviations: 4-OH - 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase; ACT - actin; Cq - quantification cycle; CYP38 - cyclophilin 38; DD - mRNA differential display; EF1α - elongation factor 1-α; FC - field capacity; MSE - mean squared error; qPCR - quantitative PCR; RTF - RNA polymerase associated protein rtf1 homolog; TDF - transcript-derived fragment; UBQ - polyubiquitin; VPS - vacuolar protein sorting-associated protein 18 homolog; α-Tub - α-tubulin; β-Tub - β-tubulin; Ψ_{soil} - soil water potential.
Acknowledgements: This work was supported by grants from the Agencia Nacional de Promoción Científica y Tecnológica (PICTO 2011-0213, PICT 2014-1246), the Secretaría General de Ciencia y Técnica- Universidad Nacional del Nordeste (PI A001/14), the Consejo Nacional de Investigaciones Científicas y Técnicas (PIP 0980), and the Establecimiento Las Marías S.A. (Gdor. Virasoro, Argentina). RMA is a Professor of the Northeast National University. OAR and PAS are members of the Research Council of Argentina (CONICET). EHA received CONICET scholarships.
* Corresponding author; fax: (+54) 379 4427131, e-mail: sansber@agr.unne.edu.ar

(Gutierrez *et al*. 2008). However, to avoid incorrect findings, an appropriate normalization strategy is required to control experimental error introduced during the numerous steps required throughout the entire workflow, *i.e*., sampling, extraction, processing of the RNA, reverse transcription, and qPCR (Die and Román 2012). Diverse methodologies have been proposed starting with the use of samples of similar size. However, this procedure is not enough to guarantee that samples from different treatments contain a similar number of cells and it should not be used as a single methodology. In a subsequent step, it is extremely important to assess the quality of the RNA and to quantify it accurately in order to introduce an equal quantity of good-quality RNA into the reverse transcription reactions. However, this procedure cannot control the variances in the enzymatic efficiency in different reactions, for example, caused by the presence of inhibitors. In addition, normalizing to total RNA has the drawback that the messenger RNA constitutes a variable fraction of about 2 - 5 %. As another alternative, the normalization to genomic DNA provides an estimation of the number of cells that were sampled. But it is infrequently used, mainly because RNA extraction procedures are not intended to co-purify DNA and its yield may vary between samples (Hugget *et al*. 2005). The most commonly utilized strategy is to normalize the expression of a target gene to an internal reference gene (usually referred to as a housekeeping gene). The ideal reference gene does not have variable expression in different tissues, developmental stages, environments, *etc.* (Andersen *et al*. 2004); therefore, its abundance is assumed to be indicative of the cDNA quantity present in each sample. Furthermore, target and reference mRNA are present in the sample even before it is collected; thus, these mRNAs should be affected similarly during all of the steps required to obtain the final qPCR measurement (Gutierrez *et al*. 2008). This might be the strongest reason for the normalization of qPCR with reference genes. Nevertheless, such a reference gene might not exist since a gene might be very stable under certain experimental conditions in which it has been tested but is often modulated under different conditions, which can introduce a significant error in the results obtained by the qPCR analyses (Zhu *et al*. 2013). This highlights the need for the validation of reference genes under the same experimental conditions in which the change in expression of target genes is investigated (Gutierrez *et al*. 2008, Kozera and Rapacz 2013). To date, validation studies of reference genes have been conducted in crops and model species, including barley (Faccioli *et al*. 2007), rice (Kim *et al*. 2003), wheat (Paolacci *et al*. 2009), *Medicago truncatula* (Kakar *et al*. 2008), poplar (Brunner *et al*. 2004), grapevine (Reid *et al*. 2006), *Arabidopsis* (Czechowski *et al*. 2005), and other plants such as *Glycyrrhiza glabra* (Maroufi 2016). Additionally, to assist in the selection of appropriate reference genes, several statistical algorithms, such as those implemented in *geNorm* (Vandesompele *et al*. 2002), *NormFinde*r (Andersen *et al*. 2004), and *BestKeeper* (Pfaffl *et al*. 2004), have been developed. These software programs calculate the stability of reference genes in a specific set of samples, allowing the establishment of a ranking of the studied candidate genes. In this context, the objective of the present study was to evaluate the suitability of nine *I. paraguariensis* genes as reference genes for quantitative qPCR normalization during drought stress and recovery to provide indispensable information for further gene expression studies in this non-model species.

Materials and methods

Two-year-old *Ilex paraguariensis* A.St.-Hil. cv. SI-49 plants grown in 4 dm^3 pots filled with lateritic soil under greenhouse conditions were transferred to a chamber set at day/night temperatures of $27 \pm 1/22 \pm 2$ ºC, a 14-h photoperiod, a photosynthetic photon flux density of 420 µmol m^{-2} s^{-1} (provided by high-pressure mercury lamps), and a relative humidity of 50 - 55 % for acclimati-zation. After 10 d, three healthy plants were watered to field capacity (FC), the pots were covered with aluminum foil to prevent evaporation, and irrigation was suspended to cause a continuous soil drying. The soil water potential was measured using a thermocouple psychrometer (*Dew Point Microvoltmeter HR-33T, Wescor*, Logan, USA) with a *PST-55* chamber according to the manufacturer´s instructions. Mature leaves were collected at midday on day 1 (control; soil water potential, Ψ_{soil}, of -0.04 MPa), during drought stress, *i.e.*, after 14 d (Ψ_{soil} = -1 MPa), 21 d (Ψ_{soil} = -2 MPa), and 26 d (Ψ_{soil} = -3 MPa), and also after 48 h of re-watering (Ψ_{soil}= -3 MPa). All harvested leaves were immediately frozen in liquid N_2 and stored at -80 °C until RNA extraction.

The total RNA was isolated using a *Spectrum™ Plant Total RNA* kit (*Sigma*, St. Louis, USA), and genomic DNA was digested with *Amplification Grade DNase I* (*Sigma*) according to the manufacturer's instructions. The RNA integrity of all samples was assessed by analysis on 2.5 % (m/v) agarose gels. The purity was calculated according to absorption ratios $(A_{260} - A_{320})/(A_{280} - A_{320})$ and $(A_{260} - A_{320})/(A_{230} - A_{320})$ measured with a *Multiskan™ GO* microplate spectrophotometer (*Thermo Scientific*, Waltham, USA). Only RNA samples with a $(A_{260} - A_{320})/(A_{280} - A_{320})$ ratio between 1.9 and 2.1 and $(A_{260} - A_{320})/(A_{230} - A_{320}) > 2.0$ were used. The RNA content was determined by measuring the A_{260}. Since we normalized the expression of the evaluated genes to the total RNA content, the samples were diluted to the same concentration, re-measured, and adjusted to ensure that the same amount of RNA was introduced in the subsequent reverse-transcription reactions. First-strand cDNA was synthesized from 4 µg of total RNA using the *ImProm-II™* reverse transcription system (*Promega*, Madison, USA) with oligo$(dT)_{20}$ primers according to the

manufacturer's instructions. For each sample, portions from the same cDNA synthesis batch were used as a template throughout the qPCR reactions, thus avoiding possible variations between different lots.

The sequences of six commonly used reference genes in plant species and three novel reference genes were obtained from the cDNA to design primer pairs suitable for the qPCR analysis. Owing to the lack of *I. paraguariensis* sequences in public databases, fragments of transcript sequences of *polyubiquitin* (*UBQ*), *elongation factor 1-alpha* (*EF1a*), *actin* (*ACT*), *tubulin alpha chain* (*α-Tub*), *tubulin beta chain* (*β-Tub*), and *cyclophilin 38* (*CYP38*) were obtained using primers designed on consensus alignments of these genes available in other plant species. In contrast, the suggested new reference genes were derived from three transcript-derived fragments (TDFs) with high stability of expression when the control plants with drought-stressed plants were earlier compared by mRNA differential display (DD) (Liang and Pardee 1992). Polyacrylamide gels from a DD assay showed a banding pattern with a uniform intensity, suggesting stable expression across the treatments and indicating that the TDFs were good candidates for qPCR normalization (Fig. 1 Suppl.). The TDFs achieved by both approaches were sequenced in *Macrogen* (Seoul, South Korea), and the homology was assessed by *BLASTN* in the GenBank database of the National Center for Biotechnology Information (*NCBI*). Afterward, the amplified sequences were extended by 5´ and 3´ rapid amplification of cDNA ends (RACE) using *GeneRacer*™ kit (*Invitrogen*™, Carlsbad, CA, USA) according to the manufacturer´s instructions and were deposited in GenBank. Based on those sequences, primers were designed using *Primer3Plus* (http:// primer3plus.com/) and selected those with an annealing temperature between 59 - 63 °C, the maximum length of the amplicon of 200 bp, and guanine-cytosine (GC) content between 40 and 60 %.

Three biological replicates were performed for each treatment with a *7500* real-time PCR system (*Applied Biosystems*, Foster City, USA). The 15 mm^3 of reaction mixture contained 10 ng of cDNA, 7.5 mm^3 of 2× *SYBR*® *Select Master Mix* (*Applied Biosystems*, Austin, USA), and 250 nM each of the specific primer pairs (Table 1 Suppl.) except for amplifying *EF1a*, *UBQ*, and *4-OH*, for which 500 nM primers were used. The qPCR programme consisted of a first step at 95 °C for 5 min and then 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Each combination of gene/sample was conducted in technical triplicate, and no-template controls were included. The specificity of the amplicons was assessed by a melting-curve analysis after the qPCR run (by heating from 60 to 95 °C). The real-time PCR amplification efficiency (E) for each primer-pair was determined by five-fold serial dilutions of the cDNAs (Rasmussen 2001) in a range of 25, 5, 1, 0.2 and 0.04 ng, and each dilution was performed in triplicate.

The real-time PCR provides the threshold cycle (Ct) of each reaction as the expression of the target. For each gene/sample combination, the average Ct value for all technical replicates was used in subsequent analyses. From this point forward, the Ct is referred to as the quantification cycle (Cq) (Lefever *et al*. 2009).

The amplification efficiencies were determined by the *7500 Software v.2.0.1* (*Applied Biosystems*, Foster City, USA) from the slope of the regression line in the standard curve according to equation $E = 10^{(-1/slope)}$ (Rasmussen 2001). Initially, to determine the influence that the biological replicates exert on the gene expression variability, an analysis of variance was conducted. The proposed statistical model considers the effect of the treatments (days 1, 14, 21, 26, and 28) and the replicates (plants A, B, and C). The mean squared error (MSE) is considered indicative of the variability introduced by the different sampled plants. Subsequently, the data were analyzed by *geNorm v3.5*, *NormFinder v0.953*, and *BestKeeper v1*, which apply different statistical algorithms to assess the stability of the candidate reference genes. Input in *geNorm* and *NormFinder* must be loaded as relative quantification data. Hence, for each gene measured, the Cq values were transformed to relative quantities Q using the formula $Q = E^{\Delta Cq}$, where E is the real-time PCR amplification efficiency of the specific primer-pair used, and ΔCq is calculated as the lowest Cq value (corresponding to the sample with the highest expression of that gene) minus the Cq value of the sample in question. The sample with the maximum expression level (the minimum Cq value) was used as a calibrator and was set to a value of 1. *BestKeeper* analyses require untransformed Cq values as input data and E.

Results

Owing to the current lack of *I. paraguariensis* sequence genes in public databases, we amplified cDNA fragments of *EF1a, UBQ, ACT, α-Tub, β-Tub,* and *CYP38* using primers designed based on related plant species. Moreover, from the DD polyacrylamide gels, we excised three transcript-derived fragments showing a good stability pattern. We sequenced and extended all of these segments by rapid amplification of the cDNA 5´ and 3´ ends when possible. By comparative analysis with *BLASTN*, all of the elongated sequences showed a high similarity (85 - 98 %) with their expected homologous sequences, and the novel reference candidates were named *RNA polymerase associated protein rtf1 homolog* (*RTF*), *4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase* (*4-OH*), *and vacuolar protein sorting-associated protein 18 homolog* (*VPS*). Ultimately, the classic reference sequences of *EF1a, UBQ, ACT, α-Tub, β-Tub,* and *CYP38* were 1 584, 1 160, 373, 1 417, 1 618, and

1 491 bp in length, respectively. The novel reference sequences, referred to as *RTF, 4-OH*, and *VPS*, were 488, 2 338, and 3 094 bp in length, respectively. These sequences were later deposited in GenBank with the accession numbers listed in Table 1 Suppl. For all primers used (Table 1 Suppl.), the melting curve analysis shows a single peak (Fig. 2 Suppl.), supporting the synthesis of a single product, which was further confirmed by electrophoresis of the qPCR products on a 2.5 % (m/v) agarose gel.

To compare the variation in the expression between treatments and different biological replicates, a heat-map made with the mean values of Cq obtained from each gene/sample combination was made (Fig. 2). The MSE

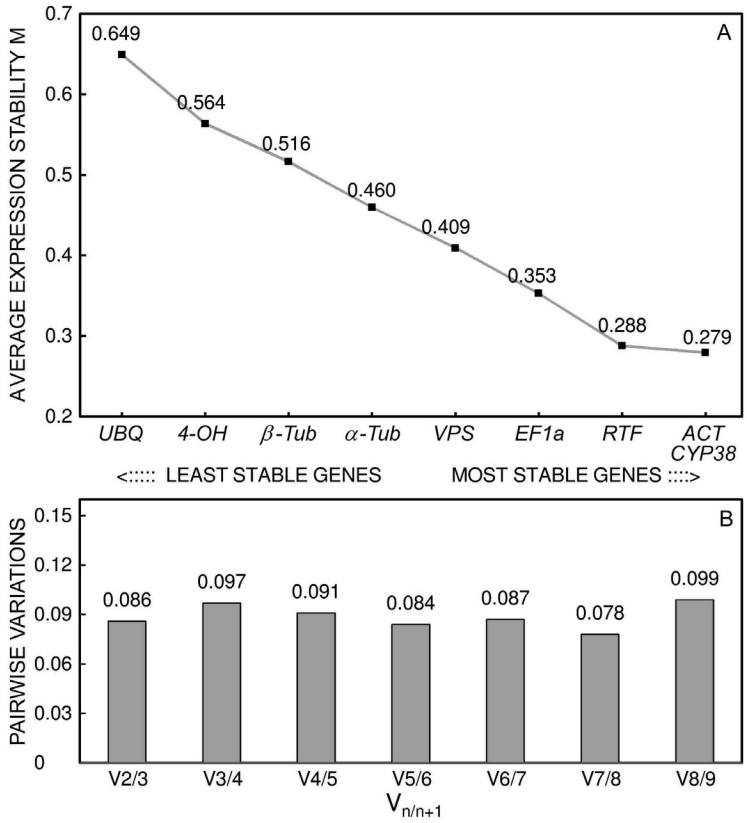

Fig. 1. Expression stability and variation analyses of the candidate reference genes by *geNorm*: *A* - average expression stability value M of reference genes at each step during stepwise exclusion of the least stable expressed reference gene; *B* - pairwise variations ($V_{n/n+1}$). Determination of the optimal number of control genes for normalization. A variation < 0.15 means no significant contribution of an additional reference gene to the normalization factor.

Table 1. Descriptive statistics of the candidate reference genes based on their quantification cycle (Cq) as calculated by *BestKeeper*. N - number of samples; GM - the geometric mean of Cq; AM - the arithmetic mean of Cq; Min and Max - the extreme values of Cq; SD - the standard deviation of the Cq; CV% - the coefficient of variation expressed as a percentage on the Cq level; Min [fold] and Max [fold] - the extreme values of expression levels expressed as an absolute x-fold over- or under-regulation coefficient; SD - standard deviation of the absolute regulation coefficients. Candidate reference genes were ranked as the most stable when have the lowest SD and CV.

Rank	Candidate gene	N	GM	AM	Min	Max	SD	CV%	Min [fold]	Max [fold]	SD
1	*RTF*	15	27.13	27.13	26.48	27.60	0.21	0.76	-1.54	1.37	± 1.15
2	*ACT*	15	27.82	27.82	27.22	28.48	0.33	1.19	-1.49	1.55	± 1.25
3	*VPS*	15	26.90	26.90	26.11	27.61	0.37	1.37	-1.70	1.62	± 1.28
4	*4-OH*	15	25.27	25.27	24.40	25.93	0.37	1.46	-1.83	1.58	± 1.28
5	*CYP38*	15	25.54	25.55	24.58	26.24	0.38	1.50	-1.91	1.60	± 1.29
6	*EF1a*	15	22.13	22.14	21.30	22.78	0.44	2.01	-1.74	1.54	± 1.34
7	*β-Tub*	15	34.08	34.09	32.91	35.21	0.57	1.68	-2.31	2.25	± 1.46
8	*α-Tub*	15	24.34	24.35	23.12	25.31	0.70	2.89	-2.24	1.91	± 1.59
9	*UBQ*	15	20.37	20.39	18.75	21.58	0.78	3.84	-2.95	2.24	± 1.68

calculated from the analysis of variance allows to conclude that the *RTF* gene was the most stable across the different biological replicates, followed by *VPS*, *CYP38*, and *EF1a*. In contrast, the expressions of *β-Tub* and *UBQ* were highly variable among different plants.

For a comprehensive analysis of gene expression stability, *BestKeeper* calculates descriptive statistics using the Cq values of each gene. Based on the standard deviation (SD) and coefficient of variation (CV) values, a first estimation of gene expression stability can be performed. Genes with an SD greater than 1 cannot be used as reference genes. The nine evaluated genes each had an SD less than 0.8; they can be ranked from the most stable, having the lowest variation, to the least stable, having the highest variation, as follows: *RTF* > *ACT* > *VPS* > *4-OH* > *CYP38* > *EF1a* > *β-Tub* > *α-Tub* > *UBQ* (Table 1). In addition, *BestKeeper* proposes using more than just one reference gene in the form of a weighted expression index, and therefore provides the *BestKeeper* index, which is the geometric mean of the Cq values from the candidate reference gene considered stably expressed, *i.e.*, SD < 1. Then, a pair-wise correlation between each candidate reference gene and the *BestKeeper* index are estimated using the Pearson correlation coefficient (r) and the probability P-value. This analysis showed a strong correlation ($r > 0.79$) with a high significance level ($P \leq 0.001$) for *α-Tub, EF1a, RTF*, and *VPS* (Table 2 Suppl.). After removing *VPS*, the correlation between the remaining genes and the index increased ($r > 0.87$). When only *α-Tub* or *EF1a* were used to obtain the index, the correlation further improved ($r = 0.987$ and 0.972, respectively).

Table 2. Expression stability of the candidate reference genes calculated by *NormFinder*.

Rank	Gene name	Stability value
1	*RTF*	0.071
2	*EF1a*	0.122
3	*ACT*	0.136
4	*CYP38*	0.143
5	*VPS*	0.145
6	*α-Tub*	0.174
7	*β-Tub*	0.213
8	*4-OH*	0.238
9	*UBQ*	0.257

From another approach, on the basis that the expression ratio of two ideal reference genes is the same in all samples independent of the cell type or the experimental condition, *geNorm* computes the reference gene stability measure M, as the average pairwise variation of a particular gene with all other reference genes in the assessment. The lower the M value, the higher the gene expression stability (Vandesompele *et al*. 2002). The gradual elimination of the gene with the greatest M value yields the most stable pair of genes. In the samples tested, they were *ACT* and *CYP38* (both M = 0.279) closely followed by *RTF* (M = 0.288) and *EF1a* (M = 0.353). The least stable one was *UBQ* with M = 0.649 (Fig. 1*A*). Moreover, Vandesompele *et al*. (2002) suggested using at least three stable genes to determine a qPCR normalization factor (NF_n, $n = 3$). Thus, a pairwise variation ($V_{n/n+1}$) analysis between the normalization factors NF_n, and NF_{n+1} was also calculated to determine the optimal number required for normalization. A $V_{n/n+1}$ value greater than 0.15 (cut-off value) gives reason to include the (n+1)th gene to calculate a reliable normalization factor. In this study, the $V_{2/3}$ was 0.086, a value below the cutoff point indicating that the two most stable reference genes (*ACT* and *CYP38*) are sufficient to obtain a reliable normalization factor, and the inclusion of the third gene (*RTF*) has no significant effect (Fig. 1*B*).

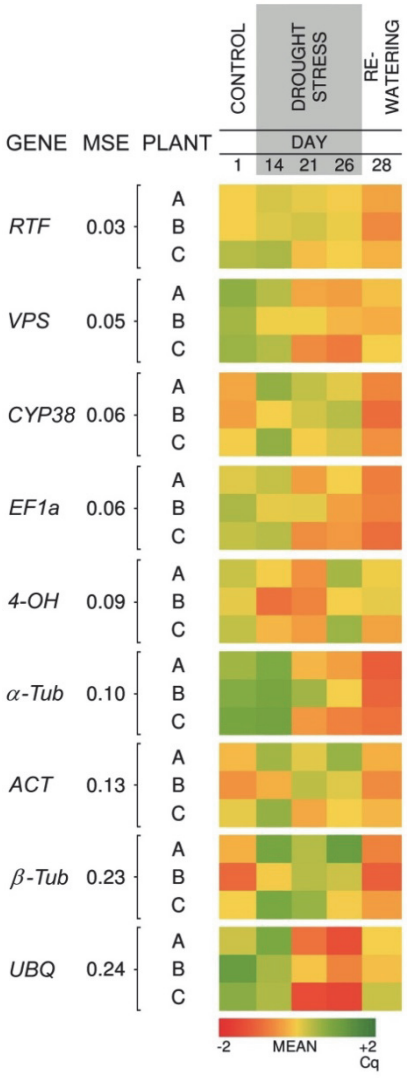

Fig. 2. Heat-map representation of the gene expression profiles in *Ilex paraguariensis* leaves subjected to drought and recovery. The mean squared error (MSE) was calculated from an analysis of variance. Colour from *red* to *green* indicates low to high expression. A, B, and C - individual plants.

In the *NormFinder* analysis, the strategy to evaluate the expression stability of candidate reference genes is rooted in the model-based approach to estimation of expression variation (Andersen *et al*. 2004). This mathematical method requires the analysis of a minimum of three genes and a minimum of two samples per group. We assessed nine genes, and three samples (plants A, B, and C) in each of the five groups (two controls and three stressed treatments). *NormFinder* computed both the intra- and inter-group expression variation, and combined the two results into a stability value; the smaller the value, the greater the gene stability. According to this, *RTF* was the most stable, followed by *EF1a, ACT, CYP38*, and *VPS*. In contrast, *UBQ* displayed the least stable expression (Table 2). In addition, *NormFinder* also indicated that *CYP38* and *VPS* were the best combination of two genes to obtain a normalization factor (NF).

Discussion

Investigating changes in gene expression is essential for identifying the function of genes and interpreting the molecular mechanisms involved in the physiological responses of plants to the environment (Gantasala *et al*. 2013). To produce accurate results, qPCR requires a robust normalization strategy by using a reference genes for relative quantification of the target gene. Many studies that have reported the results of gene expression research performed qPCR normalization with a single reference gene, but it is recommended to always use at least two or three reference genes since the use of only one can cause large errors (Nicot *et al*. 2005). The expression stability of the selected reference genes must be verified before each experiment. Preferably, the reference and target genes should have a similar expression level, but, in practice, it is not always true; therefore, it is still acceptable to use reference genes which expressions (Cq) vary between 15 and 30 (Wan *et al*. 2010). In this regard, it is necessary to consider that the Cq value is directly related to the initial amount of cDNA. In the present study, we used 10 ng of cDNA sample, and all genes assessed complied with this suggestion, except *β-Tub*, which had an arithmetic mean of Cq equal to 34.09 (Table 2).

The use of *ACT* as a reference gene in leaves of non-stressed and drought stressed *Glycyrrhiza glabra* plants is suggested, whereas *EF1a* and *β-Tub* genes are proposed only for the stressed ones (Maroufi 2016). *UBQ* and *ACT* are the most stable genes at different developmental stages and at different times of the year in poplar (Brunner *et al*. 2004). In this context, we assessed the stability of six conventional reference genes extensively utilized in plants and three novel potential reference genes to select the best candidates for qPCR normalization and to improve the accuracy of gene expression analyses in leaves of *I. paraguariensis* subjected to drought stress and recovery. To facilitate the selection of the suitable reference genes, we analysed the qPCR data with three different software programs. In this context, *BestKeeper* determined that *RTF* was the least variable gene. To obtain the *BestKeeper* index, the most correlated candidates were used, which included *RTF, EF1a*, and *α-Tub*; and the best pair was *EF1a* and *α-Tub*. In contrast, *geNorm* selected *ACT* and *CYP38* as the most stable pair of genes (both M = 0.279), closely followed by *RTF* (M = 0.288). Finally, *NormFinder* ranked *RTF* in the first place for stability, followed by *EF1a, ACT, CYP38,* and *VPS* in that order; in addition, *CYP38* and *VPS* formed the best combination of two genes for obtaining a normalization factor (NF).

After examining the algorithms underlying *geNorm*, *NormFinder*, and *BestKeeper*, it is easy to understand that the evaluated genes can be sorted in different ways depending on the software applied (Kozera and Rapacz 2013). However, we can also discern that *RTF, ACT, CYP38*, and *VPS* were ranked among the five most stable genes by the three algorithms. It is worth mentioning that two of these genes are novel references, and *RTF* emerged as the most stable gene by *BestKeeper* and *NormFinder*, and as the third most stable gene by *geNorm*; this makes it the most stable gene, and therefore suitable to be chosen when only one gene will be used for normalization of qPCR. In special cases or when the sample is scarce, use of a single reference gene is acceptable, if it was previously proven under similar experimental conditions and properly validated (Thellin *et al*. 1999). Nevertheless, Vandesompele *et al*. (2002) reported that using a single reference gene cause an error up to three-fold in 25 % and six-fold in 10 % of the cases. Using more than one gene usually increases the resolution and accuracy of the measurements.

Based on our results and considering that a general recommendation is using three genes as the most appropriate and realistic method (Vandesompele *et al*. 2002, Pfaffl *et al*. 2004, Derveaux *et al*. 2010), we suggest *RTF* combined with any of the following pairs to compute a normalization factor for qPCR: *EF1a/α-Tub, ACT/CYP38,* or *CYP38/VPS*. However, it is also expected that some differences in quantification will occur, depending on the number and selection of housekeeping genes (Nicot *et al*. 2005).

The classical reference genes utilized in plants sometimes yield good results, but often the new experiment-specific factors (even using different cultivars) cause a variation in their expression and lead to false results (Kozera and Rapacz 2013). In this sense, the expression of *UBQ* and *β-Tub* from *I. paraguariensis* samples were highly variable and their use as internal controls is discouraged. However, as in *G. glabra* and poplar, *ACT* presented a constant expression in *I. paraguariensis*. The expression of *Cyclophilin* and *β-Tub* was highly stable in the leaves of *I. paraguariensis*

in contrast with the variability observed in the pericarp of grapevine fruits by Reid *et al*. (2006). These contrasting results emphasize the need for validation of reference genes when they are used in different conditions, not previously validated.

When searching for candidate genes for novel experimental conditions, previous studies can be a good starting point for the selection of traditional reference genes, but also the evaluation of new potential reference genes is highly recommended (Kumar *et al*. 2011). The use of tools that enable global analysis of transcripts, such as microarrays (Czechowski *et al*. 2004) or serial analysis of gene expression (SAGE; Lee *et al*. 2010), is frequently recommended to find a set of genes for validation. In this context, DD is an end-point PCR technique that is intended to find genes with modulated expression (Liang and Pardee 1992) but also enables visualizing genes that are constitutively expressed across the samples (uniform band intensity, see Fig. 2 Suppl.). Unlike microarrays and SAGE, DD has the advantage that no prior sequence knowledge is required; therefore, it enables studying non-model organisms with a lack of genomic data. In addition, DD is versatile and can be easily adapted to compare samples of the new experiment in a simple, fast, affordable and reliable manner. Note that *RTF* and *VPS*, two genes obtained by this method, showed higher stability (calculated by the three algorithms) compared with the classic reference genes from the literature, possibly because the former were chosen from an experimental approach.

In conclusion, we cloned, sequenced, identified, and deposited in GenBank nine mRNAs of *I. paraguariensis* and examined their suitability as reference genes for normalization of real-time qPCR. Our data support the conclusion that *RTF* combined with any of the following pairs *EF1a/α-Tub, ACT/CYP38,* or *CYP38/VPS* are the most suitable triads of genes to compute a normalization factor. Our analysis constitutes the first in-depth study to identify the appropriate reference genes for the quantification of gene transcription in *I. paraguariensis* leaves during drought and recovery.

References

Acevedo, R.M., Maiale, S.J., Pessino, S.C., Bottini, R., Ruiz, O.A., Sansberro, P.A.: A succinate dehydrogenase flavoprotein subunit-like transcript is upregulated in *Ilex paraguariensis* leaves in response to water deficit and abscisic acid. - Plant Physiol. Biochem. **65**: 48-54, 2013.

Acevedo, R.M., Ruiz, O.A., Sansberro, P.A.: Gene expression changes in response to drought stress in *Ilex paraguariensis* leaves. - Plant Omics **9**: 334-343, 2016.

Andersen, C.L., Jensen, J.L., Ørntoft, T.F.: 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, 2004.

Brunner, A.M., Yakovlev, I.A., Strauss, S.H.: Validating internal controls for quantitative plant gene expression studies. - BMC Plant Biol. **4**: 14, 2004.

Czechowski, T., Bari, R.P., Stitt, M., Scheible, W.R., Udvardi, M.K.: Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. - Plant J. **38**: 366-379, 2004.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., Scheible, W.R.: Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. - Plant Physiol. **139**: 5-17, 2005.

Derveaux, S., Vandesompele, J., Hellemans, J.: How to do successful gene expression analysis using real-time PCR. - Methods **50**: 227-230, 2010.

Die, J.V., Román, B.: RNA quality assessment: a view from plant qPCR studies. - J. exp. Bot. **63**: 6069-6077, 2012.

Faccioli, P., Ciceri, G.P., Provero, P., Stanca, A.M., Morcia, C., Terzi, V.: A combined strategy of ‘‘in silico’’ transcriptome analysis and web search engine optimization allows an agile identification of reference genes suitable for normalization in gene expression studies. - Plant mol. Biol. **63**: 679-688, 2007.

Gantasala, N.P., Papolu, P.K., Thakur, P.K., Kamaraju, D., Sreevathsa, R., Rao, U.: Selection and validation of reference genes for quantitative gene expression studies by real-time PCR in eggplant (*Solanum melongena* L). - BMC Res. Notes **6**: 312, 2013.

Gottlieb, A., Giberti, G., Poggio, L.: Molecular analyses of the genus *Ilex* (*Aquifoliaceae*) in southern South America, evidence from AFLP and its sequence data. - Amer. J. Bot. **92**: 352-369, 2005.

Gutierrez, L., Mauriat, M., Guénin, S., Pelloux, J., Lefebvre, J.F., Louvet, R., Rusterucci, C., Moritz, T., Guerineau, F., Bellini, C., Van Wuytswinkel, O.: 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, 2008.

Heck, C.I., De Mejia, E.G.: Yerba mate tea (*Ilex paraguariensis*): a comprehensive review on chemistry, health implications, and technological considerations. - J. Food Sci. **72**: R138-R151, 2007.

Huggett, J., Dheda, K., Bustin, S., Zumla, A.: Real-time RT-PCR normalisation; strategies and considerations. - Genes Immun. **6**: 279-284, 2005.

Kakar, K., Wandrey, M., Czechowski, T., Gaertner, T., Scheible, W.R., Stitt, M., Torres-Jerez, I., Xiao, Y., Redman, J.C., Wu, H.C., Cheung, F., Town, C.D., Udvardi, M.K.: A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in *Medicago truncatula*. - Plant Methods **4**: 18, 2008.

Kim, B.R., Nam, H.Y., Kim, S.U., Kim, S.I., Chang, Y.J.: Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice. - Biotechnol. Lett. **25**: 1869-1872, 2003.

Kim, S.G., Lee, J.S., Kim, J.T., Kwon, Y.S., Bae, D.W., Bae, H.H., Son, B.Y., Baek, S.B., Kwon, Y.U., Woo, M.O., Shin, S.: Physiological and proteomic analysis of the response to drought stress in an inbred Korean maize line. - Plant Omics **8**:159-168, 2015.

Kozera, B., Rapacz, M.: Reference genes in real-time PCR. - J.

appl. Genet. **54**: 391-406, 2013.

Kumar, V., Sharma, R., Trivedi, P.C., Vyas, G.K., Khandelwal, V.: Traditional and novel references towards systematic normalization of qRT-PCR data in plants. - Aust. J. Crop Sci. **5**: 1455-1468, 2011.

Lee, J.M., Roche, J.R., Donaghy, D.J., Thrush, A., Sathish, P.: Validation of reference genes for quantitative RT-PCR studies of gene expression in perennial ryegrass (*Lolium perenne* L.). - BMC mol. Biol. **11**: 8, 2010.

Lefever, S., Hellemans, J., Pattyn, F., Przybylski, D.R., Taylor, C., Geurts, R., Untergasser, A., Vandesompele, J.: RDML: structured language and reporting guidelines for real-time quantitative PCR data. - Nucl. Acids Res. **37**: 2065-2069, 2009.

Liang, P., Pardee, A.B.: Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. - Science **257**: 967-971, 1992.

Lima, M.E., Colpo, A.C., Salgueiro, W.G., Sardinha, G.E., Ávila, D.S., Folmer, V.: *Ilex paraguariensis* extract increases lifespan and protects against the toxic effects caused by paraquat in *Caenorhabditis elegans*. - Int. J. environ. Res. Public Health **11**:10091-10104, 2014.

Maroufi, A.: Selection of reference genes for real-time quantitative PCR analysis of gene expression in *Glycyrrhiza glabra* under drought stress. - Biol. Plant. **60**: 645-654, 2016.

Nicot, N., Hausman, J.F., Hoffmann, L., Evers, D.: Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. - J. exp. Bot. **56**: 2907-2914, 2005.

Paolacci, A.R., Tanzarella, O.A., Porceddu, E., Ciaffi, M.: Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. - BMC mol. Biol. **10**: 11, 2009.

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP.: 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, 2004.

Rasmussen, R.: Quantification on the LightCycler. - In Meuer, S., Wittwer, C., Nakagawara, K.I. (ed.): Rapid Cycle Real-time PCR, Methods and Applications. Pp 21-34. Springer-Verlag, Berlin - Heidelberg 2001.

Reid, K.E., Olsson, N., Schlosser, J., Peng, F., Lund, S.T.: An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. - BMC Plant Biol. **6**: 27, 2006.

Sansberro, P.A., Mroginski, L.A., Bottini, R.: Abscisic acid promotes *Ilex paraguariensis* growth by alleviating diurnal water stress. - Plant Growth Regul. **42**: 105-111, 2004.

Shinozaki, K., Yamaguchi-Shinozaki, K.: Gene networks involved in drought stress response and tolerance. - J. exp. Bot. **58**: 221-227, 2007.

Singh, D., Laxmi, A.: Transcriptional regulation of drought response: a tortuous network of transcriptional factors. - Front. Plant Sci. **6**: 895, 2015.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., Heinen, E.: Housekeeping genes as internal standards: Use and limits. - J. Biotechnol. **75**: 291-295, 1999.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F.: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. - Genome Biol. **3**: Research0034, 2002.

Wan, H., Zhao, Z., Qian, C., Sui, Y., Malik, A.A., Chen, J.: Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber. - Anal. Biochem. **399**: 257-261, 2010.

Zhu, J., Zhang, L., Li, W., Han, S., Yang, W., Qi, L.: Reference gene selection for quantitative real-time PCR normalization in *Caragana intermedia* under different abiotic stress conditions. - PLoS ONE **8**: e53196, 2013.