Seed Science Research (2014) **24**, 341–352
© *Cambridge University Press 2014*
doi:10.1017/S0960258514000294

Identification of reference genes for gene expression studies during seed germination and seedling establishment in *Ricinus communis* L.

Paulo R. Ribeiro[1,2*], Bas J. W. Dekkers[1,3], Luzimar G. Fernandez[2], Renato D. de Castro[2], Wilco Ligterink[1*] and Henk W. M. Hilhorst[1]

[1]Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands; [2]Laboratório de Bioquímica, Biotecnologia e Bioprodutos, Departamento de Biofunção, Universidade Federal da Bahia, Reitor Miguel Calmon s/n, 40160-100 Salvador, Brazil; [3]Molecular Plant Physiology, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

(Received 18 March 2014; accepted after revision 7 August 2014; first published online 23 September 2014)

Abstract

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) is an important technology to analyse gene expression levels during plant development or in response to different treatments. An important requirement to measure gene expression levels accurately is a properly validated set of reference genes. In this context, we analysed the potential use of 17 candidate reference genes across a diverse set of samples, including several tissues, different stages and environmental conditions, encompassing seed germination and seedling growth in *Ricinus communis* L. These genes were tested by RT-qPCR and ranked according to the stability of their expression using two different approaches: GeNorm and NormFinder. GeNorm and Normfinder indicated that *ACT*, *POB* and *PP2AA1* comprise the optimal combination for normalization of gene expression data in inter-tissue (heterogeneous sample panel) studies. We also describe the optimal combination of reference genes for a subset of root, endosperm and cotyledon samples. In general, the most stable genes suggested by GeNorm are very consistent with those indicated by NormFinder, which highlights the strength of the selection of reference genes in our study. We also validated the selected reference genes by normalizing the expression levels of three target genes involved in energy metabolism with the reference genes suggested by GeNorm and NormFinder. The approach used in this study to identify stably expressed genes, and thus potential reference genes, was applied successfully for *R. communis* and it provides important guidelines for RT-qPCR studies in seeds and seedlings for other species (especially in those cases where extensive microarray data are not available).

Keywords: castor bean, gene expression, GeNorm, normalization, NormFinder, quantitative real-time polymerase chain reaction

Introduction

Ricinus communis L. (castor bean) is a drought-resistant species which is well adapted to hot climates. It is an important oilseed crop because of its unique oil composition, up to 94% of which comprises the fatty acid ricinoleic acid (12-hydroxy-*cis*-9-octadecenoic acid) (Gong *et al.*, 2005). This fatty acid confers unique properties to the oil and the biodiesel produced from it (Conceicao *et al.*, 2007; Salimon *et al.*, 2010; Anjani, 2012). *R. communis* oil is used in the chemical industry worldwide, but its production is limited by insufficient supply of castor bean seeds (Severino *et al.*, 2012). India, Brazil and China are the major oil crop producers in the world and, recently, biodiesel production has received wide attention in Brazil, where the Program for Production and Use of Biodiesel (PNPB) has identified castor bean as the ideal species for oil production which, in addition, could help to promote social development in the semi-arid regions of Brazil (Cesar and Batalha, 2010). Its agronomic importance stimulated molecular research in *R. communis* related to seed maturation (Cagliari *et al.*, 2010; Sánchez-García *et al.*, 2010; Loss-Morais *et al.*, 2013), plant development (O'Leary *et al.*, 2011), and responses to biotic and abiotic stresses

*Correspondence
Fax: +31317 418094
Emails: paulodc3@gmail.com; wilco.ligterink@wur.nl

(Wei *et al.*, 2010; Maciel *et al.*, 2011). Characterization and expression profiles of several genes from *R. communis* have been reported (Cagliari *et al.*, 2010; Chileh *et al.*, 2010; Wei *et al.*, 2010; Maciel *et al.*, 2011). In this study we focus on seed germination and seedling growth, since a better understanding of the biochemical and molecular aspects of these processes in *R. communis* is crucial for the breeding of high-yielding varieties adapted to various growing environments (Severino and Auld, 2013).

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) is the method of choice for gene expression profiling studies, due to its high sensitivity, accuracy and speed. Accurate gene expression measurements involve normalization to correct for differences in sample input due to errors in measurements, pipetting and differences in enzymatic efficiencies (Vandesompele *et al.*, 2002). Internal control genes or 'reference genes' are used to normalize for such differences. Reference genes are presumed to be expressed stably across different tissues and at various developmental stages and growth conditions. However, the stability of used reference genes needs to be properly validated for accurate and reliable normalization of gene expression data, rather than to trust their presumed expression stability. This is an important requirement for a reliable RT-qPCR experiment (Vandesompele *et al.*, 2002). In the past decade several tools have become available to determine which candidate reference genes are stably expressed across a set of samples, e.g. GeNorm (Vandesompele *et al.*, 2002), BestKeeper (Pfaffl *et al.*, 2004) and NormFinder (Andersen *et al.*, 2004). These tools have helped to identify stable reference genes in a range of plant species (Artico *et al.*, 2010; Jarosová and Kundu, 2010; Dekkers *et al.*, 2012; Pettengill *et al.*, 2012). However, so far, no validated reference genes are available for RT-qPCR studies during germination and early seedling establishment in *R. communis*.

In this study, the expression of 17 candidate reference genes was analysed across a set of 14 samples – using three biological replicates per sample – that include several tissues at different stages under different environmental conditions, including light/dark, water/nutrient solution and temperature treatments. Furthermore, to validate the candidate reference genes for analysis of gene expression purposes, and to illustrate their usefulness, we analysed the expression levels of three *R. communis* genes, oil-body lipase 2 (*OBL2*), chlorophyll *a*/*b* binding protein (light-harvesting complex – *LHCB3*) and α-amylase (*AMY3*), which are involved in primary energy metabolism. This is the first systematic report on the selection of reference genes for *R. communis*, providing useful guidelines for future accurate gene expression profiling experiments by RT-qPCR.

Materials and methods

Plant material

Ricinus communis L. seeds (genotype MPA11) used in this work were kindly supplied by Empresa Baiana de Desenvolvimento Agrícola S.A (EBDA), Salvador-Bahia, Brazil. After seed-coat removal, dry seed samples were collected and additional seeds were allowed to imbibe and germinate, using paper rolls as substrate at 25°C in the dark. In the first part of the experimental design, germinated seeds were allowed to grow on paper rolls at 25°C, in the dark, for 9 d. After collection, dry seeds and the roots, endosperm and cotyledons of 3-, 6- and 9-day-old germinated seeds were immediately frozen in liquid nitrogen, freeze-dried, ground and stored at −80°C prior to analysis. Three biological replicates of 30 seeds each were used. In the second part of the experimental design, germinated seeds were transferred from paper rolls to moist vermiculite and were allowed to grow at 20°C in continuous light for 10 d. Half of the 10-day-old seedlings were then transferred to an incubator at 35°C with continuous light, while the remaining seedlings were kept under the original conditions. After 4 d, the roots and green cotyledons (three biological replicates of 15–18 seedlings each) were collected, immediately frozen in liquid nitrogen, freeze-dried, ground and stored at −80°C prior to analysis. Details of the different samples are shown in Table 1.

RNA extraction and cDNA synthesis

Total RNA was extracted from 20 mg of dry material using the RNeasy Plant Mini Kit (Qiagen, Hilden, North Rhine-Westphalia, Germany), including the DNase digestion of genomic DNA performed using the RNase-Free DNase Set (Qiagen), according to the manufacturer's instructions. RNA quantification and quality control were performed spectrophotometrically using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, USA). Samples showing A_{260}/A_{280} and A_{260}/A_{230} ratios of 1.8–2.2 were used for subsequent analysis. RNA quality was further inspected by the integrity of ribosomal RNA bands on a 1% agarose gel. Sharp and intense 18S and 28S ribosomal RNA bands, without visible degradation, confirmed the suitability of the isolation method. The first-strand cDNA was synthesized from 1 μg of total RNA using the iScript™ cDNA synthesis kit (Bio-Rad, Hercules, California, USA) according to the manufacturer's instructions. Reverse transcription was performed at 37°C for 40 min followed by 85°C for 5 min. The cDNA was diluted 20 times for the use in RT-qPCR reactions. All cDNAs were stored at −20°C prior to the analysis.

https://doi.org/10.1017/S0960258514000294 Published online by Cambridge University Press

Table 1. Description of tissue samples used for reference gene validation

Sample code	Sample type	Age (d)	Growth condition and treatment
C3	Cotyledons[a]	3	Paper roll, imbibed in water, dark, 25°C
C6	Cotyledons[a]	6	Paper roll, imbibed in water, dark, 25°C
C9	Cotyledons[a]	9	Paper roll, imbibed in water, dark, 25°C
C14a	Green cotyledons + first real leaves[b]	14	Vermiculite, full light, nutrient solution, 20°C
C14b	Green cotyledons + first real leaves[b]	14	Vermiculite, full light, nutrient solution, 35°C
E3	Endosperm[a]	3	Paper roll, imbibed in water, dark, 25°C
E6	Endosperm[a]	6	Paper roll, imbibed in water, dark, 25°C
E9	Endosperm[a]	9	Paper roll, imbibed in water, dark, 25°C
R3	Root[a]	3	Paper roll, imbibed in water, dark, 25°C
R6	Root[a]	6	Paper roll, imbibed in water, dark, 25°C
R9	Root[a]	9	Paper roll, imbibed in water, dark, 25°C
R14a	Root[b]	14	Vermiculite, full light, nutrient solution, 20°C
R14b	Root[b]	14	Vermiculite, full light, nutrient solution, 35°C
Dry	Dry seeds[a]	–	–

[a]Three biological replicates of 30 seeds. [b]Three biological replicates of 15–18 seeds.

Reference gene selection, primer design and measurement of amplification efficiency

Potential reference genes were selected based on available data in the literature on commonly used reference genes in RT-qPCR studies in Arabidopsis, tomato, *Jatropha curcas* L. and *R. communis*. We started our selection for candidate reference genes by a literature search for reference genes commonly used in *R. communis*, which includes: *18S*, protein Pob (*POB*), NADH-ubiquinone oxidoreductase (*NADH_OXI*) and elongation factor 1-beta (*EF1B*) (Eastmond, 2004; Chen *et al.*, 2007; Cagliari *et al.*, 2010; H.L. Li *et al.*, 2012; Arroyo-Caro *et al.*, 2013; Loss-Morais *et al.*, 2013). In addition to these candidates, 40S ribosomal protein (*40S*), 60S acidic ribosomal protein (*60S*), β-tubulin (*β-TUB*), translation elongation factor G (*EFG*) and structural constituent of ribosome (*SCR*) genes were also selected because of their use as reference genes for other species (Fig. 1 and see supplementary Table S1, available online) (Le *et al.*, 2012; X.S. Li *et al.*, 2012; Rapacz *et al.*, 2012). However, only few of these studies have validated the expression stability of these putative reference genes (Cagliari *et al.*, 2010). Since the number of candidate reference genes was very low we extended our search and looked for putative reference genes that were described as stably expressed in seeds of the model plant *Arabidopsis thaliana*, tomato and also for putative reference genes in the closely related species *J. curcas*. For Arabidopsis, tomato and *J. curcas* candidate genes, the sequences of possible orthologues were identified through a TBLASTX against the *R. communis* translated nucleotide database (NCBI, http://www.ncbi.nlm.nih.gov/). Only sequences that showed high similarity (E-value $<$ 1e−70) were considered putative orthologues in *R. communis* and were selected for primer design. This selection consisted of genes for ubiquitin-conjugating enzyme E2 (*UBI_E2*), serine/threonine protein phosphatase 2a regulatory subunit A (*PP2AA1*), Type 2A phosphatase activator TIP41 (*TIP41*) and GTP-binding protein Sar1 (*SAR1*) used for Arabidopsis and tomato seeds (Dekkers *et al.*, 2012) and actin (*ACT*), ubiquitin (*UBI*), α-tubulin (*α-TUB*) and glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) used for developing *J. curcas* seeds (Fig. 1, supplementary Table S1) (Gu *et al.*, 2012; Fan *et al.*, 2013; Rocha *et al.*, 2013). After an initial quality assessment, primer pairs for 17 putative reference genes remained (supplementary Table S1). The gene identifier (GI) number, gene name and description, amplicon length, primer efficiency and primer sequences (reverse and forward) are listed in supplementary Table S1. Primers for the 17 potential reference genes were designed using CLCbio software (CLCbio, Aarhus, Denmark) with melting temperatures (T_m) of 58–62°C, primer lengths of 18–22 bp and amplicon lengths of 80–200 bp. Information about all primers used in this study is summarized in supplementary Table S1. The PCR amplification efficiency was evaluated based on a standard curve generated by twofold serial dilutions of a pooled cDNA sample. The specificity of the primers was verified by loading the amplicons on a 2.5% agarose gel and melting curve analysis.

RT–qPCR conditions

For RT-qPCR experiments, reactions were prepared in a total volume of 10 µl, containing 2.5 µl of cDNA, 0.5 µl of primer-mix (from a 10 µM working solution), 5 µl of iQ SYBR Green Supermix (Bio-Rad) and 2 µl of water. The RT-qPCRs were run on a CFX (Bio-Rad). The following qPCR programme was used for all PCR reactions: 95°C for 3 min, followed by 40 cycles of 95°C for 15 s and 60°C for 30 s. Melting curves were recorded

https://doi.org/10.1017/S0960258514000294 Published online by Cambridge University Press

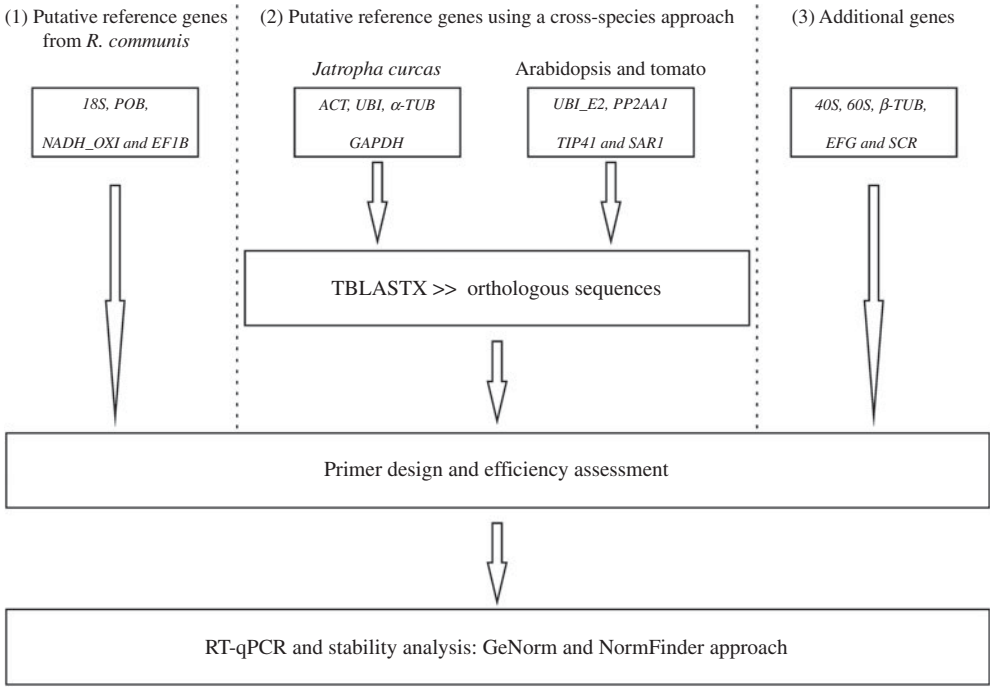

Figure 1. Flowchart showing the steps followed in this study from the selection of putative reference genes to gene expression stability analysis.

after cycle 40 by heating from 65 to 95°C, increasing the temperature stepwise every 5 s by 0.5°C.

Evaluation of reference gene expression stability: GeNorm and NormFinder

The expression levels of 17 candidate reference genes were determined by the number of cycles (Cq) required to reach fluorescence above a specific threshold level. GeNorm (Vandesompele *et al.*, 2002; Hellemans *et al.*, 2007) and NormFinder (Andersen *et al.*, 2004) were used for expression stability evaluation of the candidate reference genes across all tested samples.

GeNorm ranks the candidate reference genes according to the expression stability index, *M*. The most stable genes show the lowest *M* values. The pairwise variation (V_n/V_{n+1}) was calculated to determine the optimal number of reference genes (Vandesompele *et al.*, 2002). Alternatively, Normfinder (Andersen *et al.*, 2004) was used to identify the optimal reference genes. Normfinder ranks the putative reference genes according to their stability value (Andersen *et al.*, 2004).

Reference gene validation: normalization of three target genes

Three genes involved in primary energy metabolism were used to demonstrate the usefulness of the candidate reference genes and to validate them. Oil-body lipase 2 (*OBL2*), chlorophyll *a*/*b* binding protein (light-harvesting complex–*LHCB3*) and α-amylase (*AMY3*) expression levels were normalized using the most stable candidate reference genes across all tested samples. qBase (Biogazelle, Ghent, Belgium) was used to normalize the expression levels using these reference genes and the most stable candidate reference genes across three sample subsets: root, cotyledons and endosperm samples. qBase allows gene expression levels to be normalized by using up to five reference genes (Hellemans *et al.*, 2007). Statistical analyses were conducted using one-way ANOVA and the Tukey post-hoc test with IBM SPSS Statistics® (Windows version 19, SPSS Inc., Chicago, Illinois, USA) and Microsoft Excel 2010®. Values of $P < 0.05$ were considered to be statistically significant.

Results and discussion

Primer specificity, efficiency and expression profile of the 17 candidate reference genes

Gene expression measurements by RT-qPCR rely on normalization by stably expressed reference genes. A set of validated reference genes is of utmost importance to ensure reliable and accurate data normalization. However, such a set is not available for *R. communis*. Therefore we selected a range of putative reference genes and tested their expression

https://doi.org/10.1017/S0960258514000294 Published online by Cambridge University Press

stability over a diverse set of samples encompassing seed germination and seedling establishment in *R. communis*. A successful approach to identify stably expressed genes, and thus potential reference genes, in a given sample panel is to mine microarray data (Czechowski *et al.*, 2005; Popovici *et al.*, 2009; Dekkers *et al.*, 2012; De Oliveira *et al.*, 2012). To the best of our knowledge, microarray data for seed germination and seedling establishment in *R. communis* are not yet available. Therefore, we selected 17 putative reference genes from a literature search. These include four candidate reference genes commonly used in *R. communis* and 13 candidate reference genes used in other species (Fig. 1).

To evaluate the primer amplification specificity of these putative reference genes, electrophoresis was performed with the product of an RT-qPCR on a pooled cDNA template with the primers for each candidate reference gene. Additionally, analysis of the melting curves confirmed that all primers generated only a single amplicon (see supplementary Fig. S1, available online) and that neither primer–dimers nor unexpected products were found. The absence of genomic DNA was confirmed by comparing RT-qPCR with the cDNA template with that of RNA samples which were not reverse transcribed (minus RT control). Most of the minus RT controls did not show any amplification. In some samples, low signals were detected in the minus RT control, but these were at least 7 Cq (quantification cycles) higher than with the cDNA samples. This is clearly above the limit of 5 Cq as proposed by Nolan *et al.* (2006). Amplification efficiency ranged from 88.77% for *GAPDH* to 111.6% for *β-TUB*, and the coefficient of determination (R^2) varied from 0.9709 to 0.9997 (see supplementary Table S1). This indicates that the reaction conditions are optimal and that the results obtained are highly repeatable.

The expression levels of the 17 candidate reference genes across all samples are presented as Cq values (Fig. 2 and supplementary Table S1) which represent the relative abundance of a particular transcript. The mean values of most candidate reference genes were between a Cq of 21 (*EF1B*) and 29 (*SCR*). This difference represents a 150-fold higher abundance of *EF1B* over *SCR*. The mean Cq value of *18S* was 8.41. *SAR1* showed the lowest variation in expression across all tested samples ($Cq_{max} - Cq_{min} = 6.22$), whereas *β-TUB* showed the highest variation ($Cq_{max} - Cq_{min} = 13.34$) (see supplementary Table S1).

Gene expression stability analysis

First, gene expression stability was analysed across all samples, aiming at selecting reference genes that could be used to normalize between tissues. Then the samples were subdivided into three tissue-specific subsets, i.e. root, cotyledon and endosperm samples (Table 1). These four sample panels were analysed using two methods to assess the gene expression stability: GeNorm and NormFinder. GeNorm calculates the average expression stability (*M*) based on the average pairwise variation between all tested genes. A lower value of *M* indicates more stable gene expression and vice versa (Vandesompele *et al.*, 2002). The ranking order according to the *M* value is shown in Fig. 3 and Table 2. For more heterogeneous sets of samples (all tested samples), *M* values lower than 1 were used (Hellemans *et al.*, 2007). *M* values lower than 0.5 were used to identify stably expressed reference genes in relatively homogeneous sets of samples (tissue-specific).

POB, *PP2AA1*, *ACT* and *SAR1* were defined as the four most stable reference genes ($0.575 < M < 0.625$) for the entire set of samples (inter-tissue) (Fig. 3a). GeNorm also determines the optimal number of reference genes (GeNorm $V < 0.15$) required for calculating an accurate normalization factor, which is

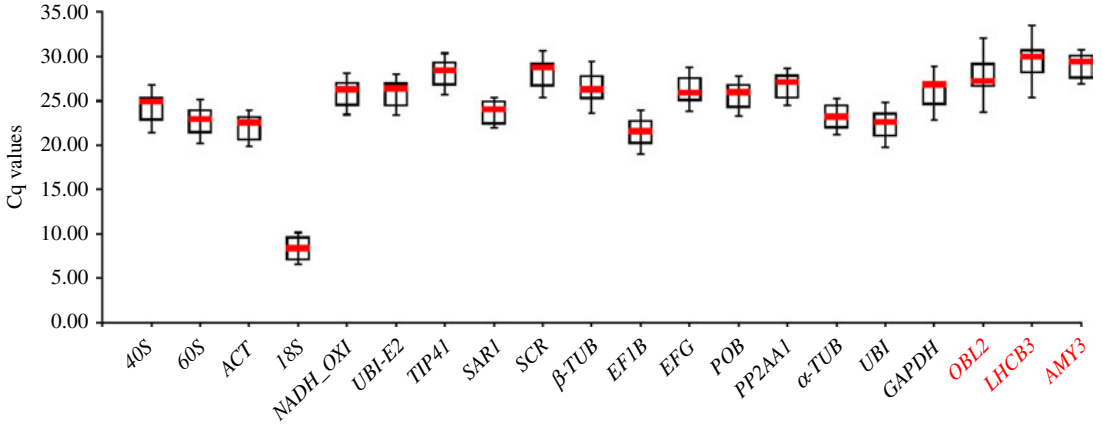

Figure 2. (colour online) Quantification cycle values (Cq) of all 17 candidate reference and 3 target genes across all samples. The horizontal line across the boxes depicts the median Cq value.

https://doi.org/10.1017/S0960258514000294 Published online by Cambridge University Press

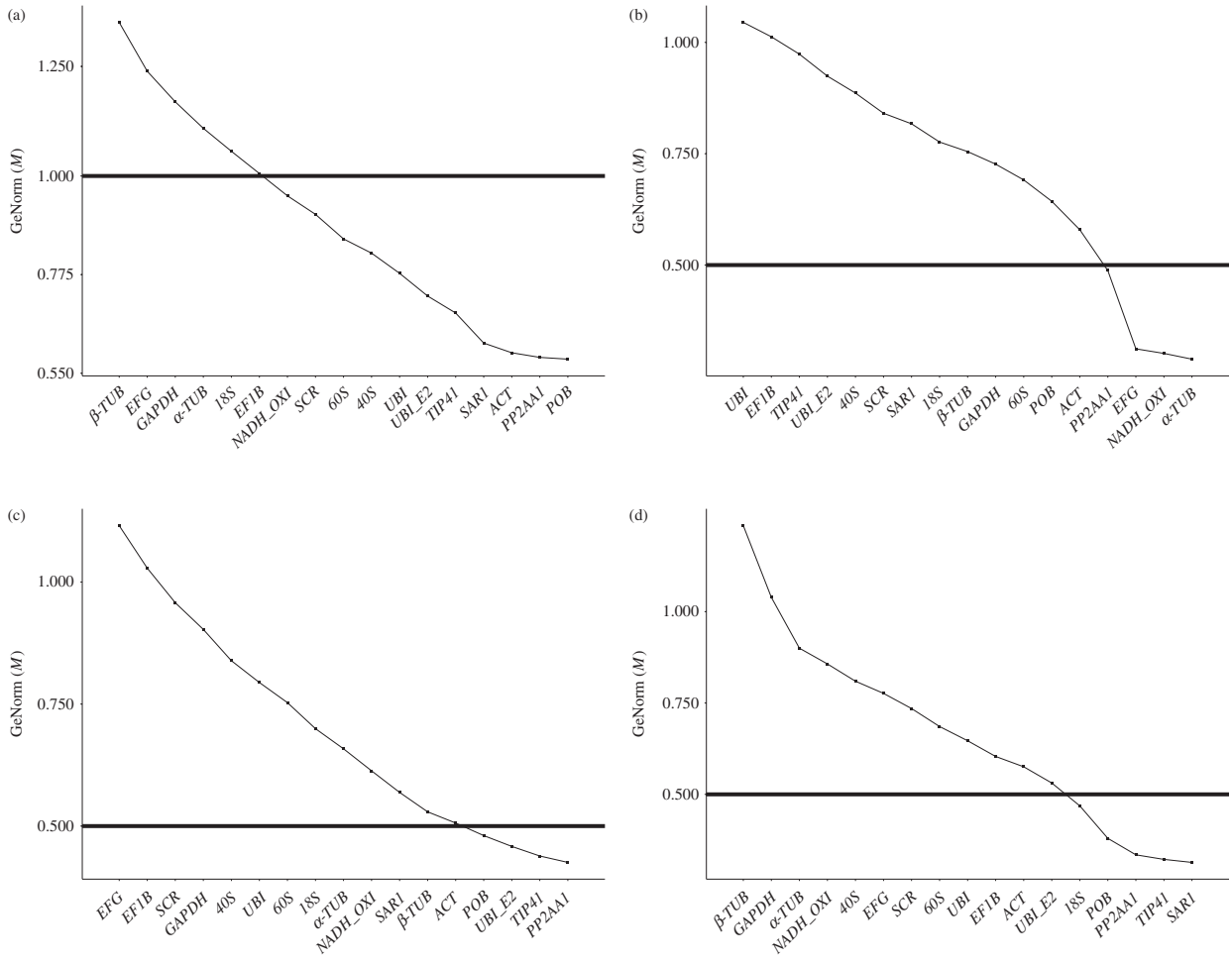

Figure 3. Expression stability values (*M*) of 17 candidate reference genes. (a) All *R. communis* samples; (b) root samples in different developing stages and temperatures; (c) cotyledon samples in different developing stages and temperatures; (d) endosperm samples in different developing stages. A lower *M* value indicates more stable expression.

based on pairwise variation (V_n/V_{n+1}). In order to accurately normalize gene expression data for the entire set of samples, the use of the three most stable genes is suggested ($V_3/V_4 < 0.15$) (Fig. 4a, Table 2). For accurate normalization of root and endosperm subset samples two reference genes should be used (Fig. 4b and 4d; Table 2), while for cotyledons the optimal number is three (Fig. 4c; Table 2). *NADH_OXI* and *α-TUB* were the most stable genes across all root samples, *UBI_E2*, *TIP41* and *PP2AA1* for cotyledons and *TIP41* and *SAR1* for endosperm samples (Fig. 3; Table 2). As expected, the most stable genes showed the lowest variation in expression across the entire sample panel ($Cq_{max} - Cq_{min} \approx 6–7$) while the least stable genes showed the

Table 2. Expression stability analysis of 17 candidate reference genes by GeNorm

Experimental sets	Most stable genes	Optimal combination ($V < 0.15$)	Least stable genes
Total	*ACT*, *TIP41*, *SAR1*, *POB* and *PP2AA1* (M < 1.0)	*ACT*, *POB* and *PP2AA1*	*β-TUB*, *EFG*, *α-TUB* and *GAPDH*
Root	*NADH_OXI*, *EFG*, *PP2AA1* and *α-TUB* (M < 0.5)	*NADH_OXI* and *α-TUB*	*UBI_E2*, *TIP41*, *EF1B* and *UBI*
Cotyledons	*UBI_E2*, *TIP41*, *POB* and *PP2AA1* (M < 0.5)	*UBI_E2*, *TIP41* and *PP2AA1*	*SCR*, *EF1B*, *EFG* and *GAPDH*
Endosperm	*18S*, *TIP41*, *SAR1*, *POB* and *PP2AA1* (M < 0.5)	*TIP41* and *SAR1*	*NADH_OXI*, *β-TUB*, *α-TUB* and *GAPDH*

https://doi.org/10.1017/S0960258514000294 Published online by Cambridge University Press

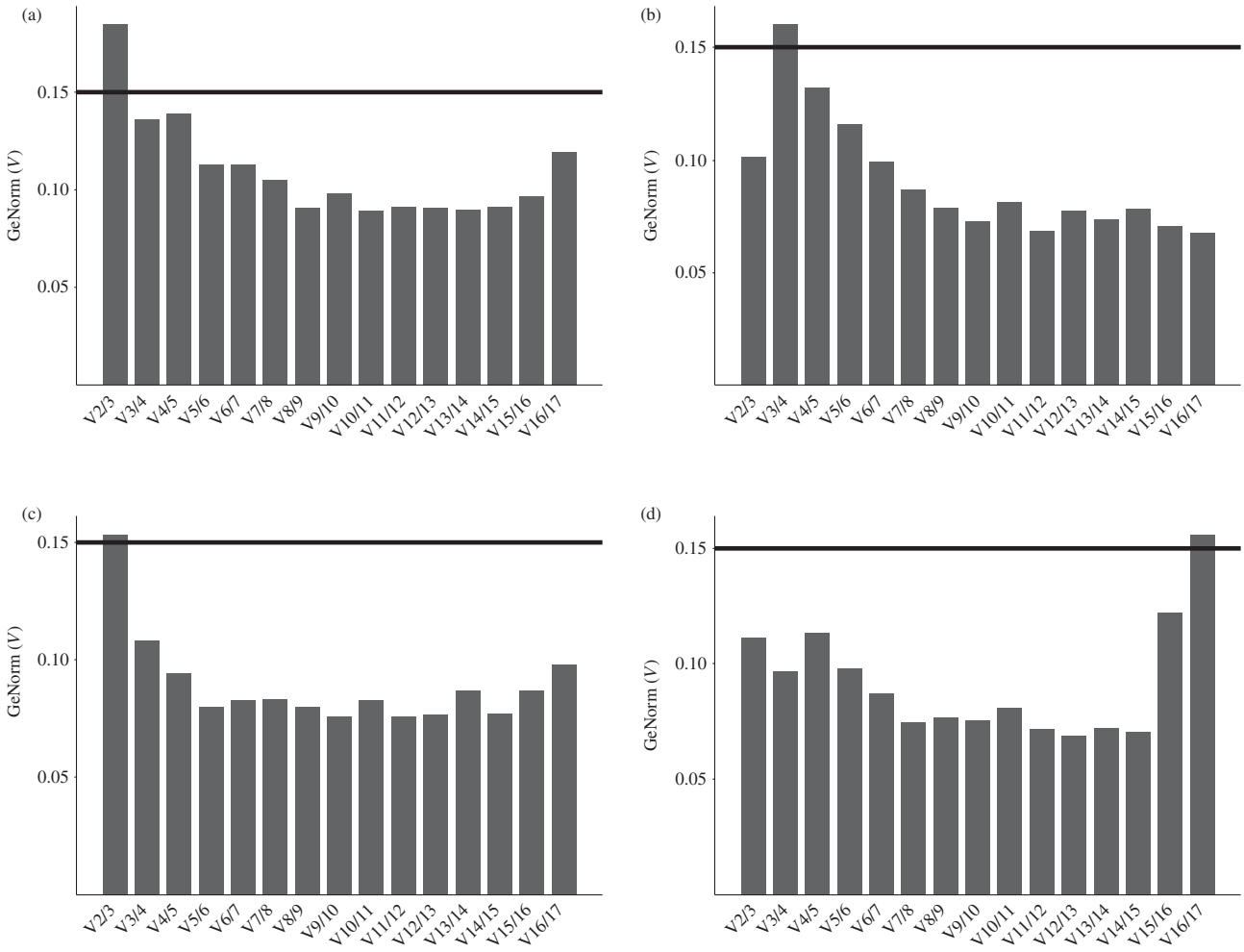

Figure 4. Determination of optimal number of reference genes based on pairwise variation (*V*) analysis of normalization factors of the candidate reference genes. (a) All *R. communis* samples; (b) root samples in different developing stages and temperatures; (c) cotyledon samples in different developing stages and temperatures; (d) endosperm samples in different developing stages. The V_n/V_{n+1} values were calculated to determine the optimal number of reference genes for use in RT-qPCR data normalization.

highest variation ($Cq_{max} - Cq_{min} \approx 9–13$) (Fig. 2, Table 2, supplementary Table S1).

NormFinder is a freely available Visual Basic application for Microsoft Excel, which automatically calculates the stability value for all candidate reference genes tested on a sample set containing any number of samples organized in any given number of groups (Andersen *et al.*, 2004). NormFinder ranks all the candidate reference genes according to their stability value in a robust manner which shows low sensitivity toward co-regulation of the candidate reference genes (Andersen *et al.*, 2004). The ranking of reference genes based on the NormFinder algorithm for the entire set of samples indicated *ACT*, *PP2AA1* and *POB* as the three most stable genes (Table 3), which is very consistent with the output obtained from GeNorm (Fig. 3a, Table 2). Furthermore, NormFinder indicated *ACT*, *PP2AA1* and *POB* as the three most stable genes across all root samples, *UBI_E2*, *PP2AA1* and *POB* for cotyledons and *ACT*, *EFG* and *18S* for endosperm samples (Table 3).

Additionally, genes encoding the ribosomal units 18S, 40S and 60S were suggested by NormFinder as suitable reference genes for normalization of RT-qPCR data in *R. communis* (Table 3). In order to rule out a potential sensitivity of this algorithm to co-regulation of these ribosomal genes, we re-calculated the stability value of all candidate reference genes without the gene expression data of *40S*. However, the removal of *40S* hardly affected the candidate reference gene output of NormFinder (Table 4).

Expression profiles of three functional genes

To evaluate our sets of reference genes we measured and normalized the expression of three target genes involved in energy metabolism. OBL2 is an

https://doi.org/10.1017/S0960258514000294 Published online by Cambridge University Press

Table 3. Expression stability analysis of 17 candidate reference genes by NormFinder

Total		Roots		Cotyledons		Endosperm	
Best gene	Stability value	Best gene	Stability value	Best gene	Stability value	Best gene	Stability value
ACT	0.277	*ACT*	0.014	*UBI_E2*	0.012	*ACT*	0.064
PP2AA1	0.278	*POB*	0.023	*PP2AA1*	0.015	*EFG*	0.074
POB	0.289	*PP2AA1*	0.074	*POB*	0.028	*18S*	0.091
NADH_OXI	0.327	*60S*	0.221	*TIP41*	0.049	*POB*	0.198
60S	0.345	*40S*	0.230	*β-TUB*	0.066	*NADH_OXI*	0.200

oil-body-associated lipase from *R. communis* endosperm that is induced upon seed germination and is responsible for the hydrolysis of a range of triacylglycerols (Eastmond, 2004). OBL-like proteins are present in many species and they probably play an important role in regulating lipolysis during germination (Eastmond, 2004). LHCB3 (light-harvesting chlorophyll *a*/*b*) is responsible for the conversion of free energy from absorbed light into ATP, reduced ferredoxin (Fd) and NADPH that can be used to drive metabolism in the chloroplast and cytosol (Foyer *et al.*, 2012). Down-regulation or disruption of any member of the LHC family reduces responsiveness of stomatal movement to abscisic acid (ABA) and therefore results in a decrease in tolerance to drought stress in *Arabidopsis thaliana* (Xu *et al.*, 2012). α-Amylase is involved in the degradation of starch and it has been demonstrated that in Arabidopsis leaves it is induced by biotic and abiotic stress and secreted (Doyle *et al.*, 2007). The expression levels of the three target genes were measured and normalized by: (1) using *ACT*, *POB* and *PP2AA1* as reference genes; and (b) using tissue-specific reference genes for which the expression stability value, *M*, was below the recommended value of 0.5 for each subset. The results show that *OBL2*, *LHCB3* and *AMY3* genes were expressed in all samples (Fig. 5) and the mean Cq values varied between 27.23 (*OBL2*) and 29.96 (*LHCB3*) (Fig. 2, supplementary Table S1). *LHCB3* showed the highest variation in expression across all tested samples ($Cq_{max} - Cq_{min} = 13.3$), whereas *AMY3* showed the lowest variation ($Cq_{max} - Cq_{min} = 6.81$). The observed wide range of Cq values for *LHCB3* is due to the fact that photosynthetic tissues, such as 14-day-old greening cotyledons, showed relatively higher expression as compared to the non-photosynthetic endosperm tissues (data not shown). Expression levels of *OBL2* were more than a 1000-fold higher in 9-day-old endosperm as compared to the expression in dry seeds (Fig. 5g). This is consistent with the function of the gene, which regulates lipolysis during germination (Eastmond, 2004). To illustrate the effect of reference gene set selection on the normalization and the accuracy of the RT-qPCR data, we compared the normalized expression data of the target genes obtained by using *ACT*, *POB* and *PP2AA1* as reference genes with the values obtained by using tissue-specific reference genes. For most of the genes, a high correlation was found between the normalized data obtained by using *ACT*, *POB* and *PP2AA1* as reference genes and the data normalized by the respective tissue-specific reference genes (Fig. 5). The correlation coefficient (R^2) varied from 0.4348 (*AMY3*, roots) to 0.9986 (*OBL2*, endosperm) (Fig. 5a–i). We only found a weak correlation for *OBL2* (Fig. 5a) and *AMY3* (Fig. 5c) in the root sample panel. However, it is important to highlight that the combination of genes suggested for the entire dataset by GeNorm (as well as Normfinder) is the same as that suggested by NormFinder for the root subset and thus similar! This difference is likely caused by the different algorithms used in both methods.

The fact that most stable genes suggested by GeNorm are very consistent with those indicated by NormFinder for the entire sample panel and for the cotyledon subset, highlights the strength of the

Table 4. Expression stability analysis of 17 candidate reference genes by NormFinder (excluding *40S*)

Total		Roots		Cotyledons		Endosperm	
Best gene	Stability value	Best gene	Stability value	Best gene	Stability value	Best gene	Stability value
ACT	0.270	*ACT*	0.014	*PP2AA1*	0.004	*ACT*	0.054
PP2AA1	0.278	*POB*	0.033	*UBI_E2*	0.008	*EFG*	0.084
POB	0.308	*PP2AA1*	0.060	*POB*	0.023	*18S*	0.084
NADH_OXI	0.312	*EFG*	0.224	*β-TUB*	0.048	*NADH_OXI*	0.157
60S	0.352	*60S*	0.228	*TIP41*	0.064	*UBI_E2*	0.202

https://doi.org/10.1017/S0960258514000294 Published online by Cambridge University Press

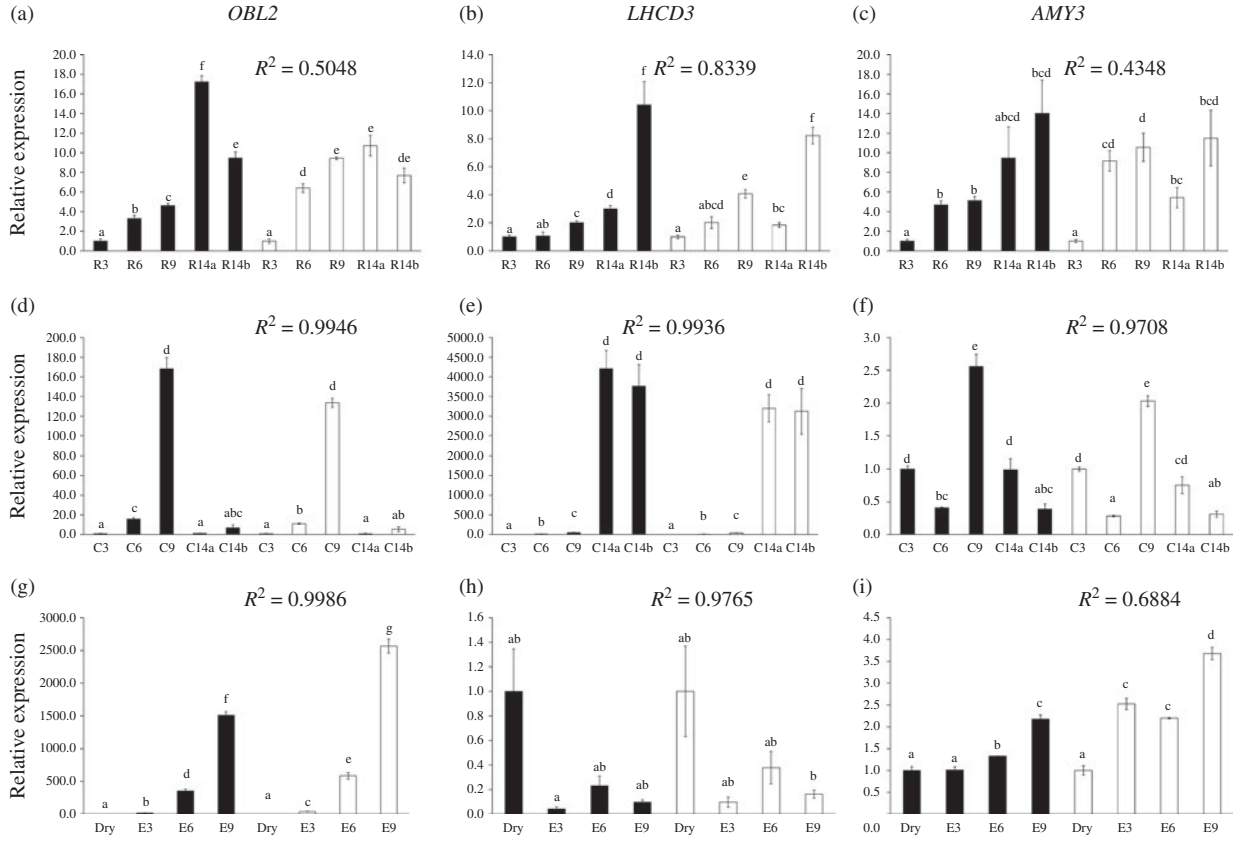

Figure 5. Relative expression of *OBL2* (a, d and g), *LHCB3* (b, e and h) and *AMY3* (c, f and i) in root (a–c), cotyledon (d–f) and endosperm (g–i) samples. Black bars represent the expression levels normalized by using *ACT*, *POB* and *PP2AA1* reference genes as the optimal combination selected by GeNorm to normalize an inter-tissue (heterogeneous) sample panel. White bars represent the expression levels normalized by using reference genes found as the optimal combination selected by GeNorm for the sample-specific sets: *NADH_OXI* and *α-TUB* in root samples (a–c), *UBI_E2*, *TIP41* and *PP2AA1* in cotyledon samples (d–f) and *TIP41* and *SAR1* in endosperm samples (g–i). Averages and standard errors of three biological and two technical replicates are shown. Expression levels were normalized towards the first column (R3, C3 and Dry) for roots, cotyledons and endosperm sample panels, respectively. Results are expressed as mean ± SD. Letters above the bars indicate significant differences between different samples by Tukey's HSD ($P < 0.05$).

selection of the reference genes in our study. However, little overlap was observed for endosperm samples (Tables 2 and 3). Therefore, we also normalized the expression of the three studied target genes by using the most stable genes suggested by NormFinder (*ACT* and *EFG*) for endosperm samples (Fig. 6). The expression pattern of *OBL2* and *LHCB3* were extremely consistent between all three sets of reference genes used to normalize the expression levels (Fig. 6a, b), however the relative expression levels of *OBL2* were slightly higher for the samples normalized by using *TIP41* and *SAR1* reference genes (white bars) as compared to the samples normalized by using *ACT*, *POB* and *PP2AA1* (black bars) and *ACT* and *EFG* [grey bars (red online)] as reference genes for the subset of endosperm samples (Fig. 6a). The expression pattern and expression levels of *AMY3* for the samples normalized by using the optimal combination selected by GeNorm to normalize the complete inter-tissue sample panel (*ACT*, *POB* and *PP2AA1*) and the optimal combination selected by NormFinder for the subset of endosperm samples (*ACT* and *EFG*) were more similar to each other than to the values obtained by using *TIP41* and *SAR1* as reference genes (Fig. 6c). Taken together, these results suggest that the approach used to select reference genes for *R. communis* is robust and provides useful guidelines for future accurate gene expression profiling experiments by RT-qPCR.

Conclusions

This study reports a systematic analysis aimed at determining the optimal combination of reference genes for the normalization of gene expression data for *R. communis* seed germination and seedling establishment, providing useful guidelines for future accurate gene expression profiling experiments by RT-qPCR.

https://doi.org/10.1017/S0960258514000294 Published online by Cambridge University Press

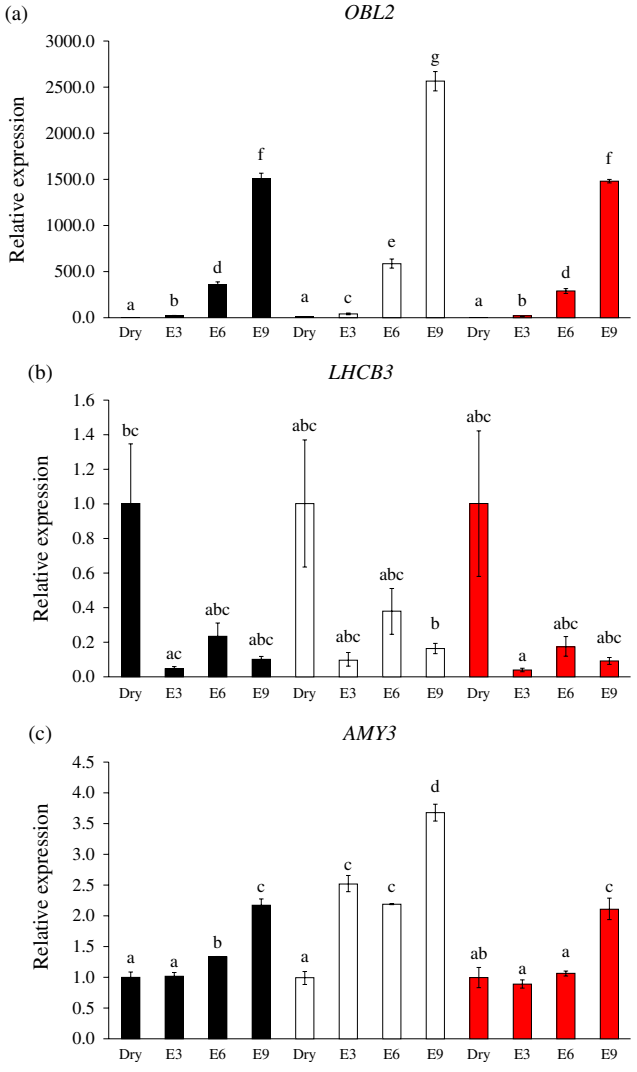

Figure 6. (colour online) Relative expression of *OBL2* (a), *LHCB3* (b) and *AMY3* (c) in endosperm samples. The black bars represent the expression levels normalized by using *ACT, POB* and *PP2AA1* reference genes as the optimal combination selected by GeNorm to normalize an inter-tissue (heterogeneous) sample panel. The white bars represent the expression levels normalized by using *TIP41* and *SAR1* reference genes found as the optimal combination selected by GeNorm for the subset of endosperm samples. The grey bars (red online) represent the expression levels normalized by using *ACT* and *EFG* reference genes found as the optimal combination selected by NormFinder for the subset of endosperm samples. Averages and standard errors of three biological and two technical replicates are shown. Expression levels were normalized towards dry seeds. Results are expressed as mean ± SD. Letters above the bars indicate significant differences between different samples by Tukey's HSD ($P < 0.05$).

Genes described in the literature as stably expressed in *J. curcas* (*ACT*), Arabidopsis and tomato (*TIP41, SAR1* and *PP2AA1*) were proven to be stable for *R. communis* seed and seedling samples (Fig. 1 and Tables 2 and 3). Therefore these genes are likely candidates for successful gene expression studies in seeds and seedlings of a broader range of species. More importantly, our data showed that some of the widely used reference genes were not suitable reference genes for our samples (Table 2), highlighting the importance of a proper validation of candidate reference genes for each study. Normalization of expression data of three target genes involved in energy metabolism highlights the reliability of the selected reference genes in our study. The approach used in this study to identify stably expressed genes, and thus potential reference genes, was successfully applied for *R. communis* and it provides important guidelines for finding suitable reference genes for RT-qPCR studies in seeds and seedlings of other species (especially in those cases where extensive microarray data are not available).

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0960258514000294

Authors' contributions

P.R.R. carried out the physiological and molecular experiments, data processing, statistical analysis and draft of the manuscript. L.G.F. and R.D.C. participated in the design of the study. B.J.W.D., W.L. and H.W.M.H. participated in the design of the study, coordination and critical reading of the manuscript. All authors read and approved the manuscript.

Acknowledgements

The authors would like to thank Empresa Baiana de Desenvolvimento Agrícola S.A (EBDA) for kindly supplying the seeds used in this work.

Financial support

Financial support was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (grant number 200745/2011-5), Petróleo Brasileiro S.A. (grant number 0050.0062805.10.9), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – Brazil, and the Netherlands Organization for Scientific Research – NWO.

Conflicts of interest

None.

References

Andersen, C.L., Jensen, J.L. and Ørntoft, T.F. (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**, 5245–5250.

Anjani, K. (2012) Castor genetic resources: a primary gene pool for exploitation. *Industrial Crops and Products* **35**, 1–14.

Arroyo-Caro, J.M., Chileh, T., Kazachkov, M., Zou, J., Alonso, D.L. and García-Maroto, F. (2013) The multigene family of lysophosphatidate acyltransferase (LPAT)-related enzymes in *Ricinus communis*. Cloning and molecular characterization of two *LPAT* genes that are expressed in castor seeds. *Plant Science* **199–200**, 29–40.

Artico, S., Nardeli, S.M., Brilhante, O., Grossi-de-Sa, M.F. and Alves-Ferreira, M. (2010) Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biology* **10**, 49.

Cagliari, A., Margis-Pinheiro, M., Loss, G., Mastroberti, A.A., de Araujo Mariath, J.E. and Margis, R. (2010) Identification and expression analysis of castor bean (*Ricinus communis*) genes encoding enzymes from the triacylglycerol biosynthesis pathway. *Plant Science* **179**, 499–509.

Cesar, A.d.S. and Batalha, M.O. (2010) Biodiesel production from castor oil in Brazil: a difficult reality. *Energy Policy* **38**, 4031–4039.

Chen, G.Q., Turner, C., He, X., Nguyen, T., McKeon, T.A. and Laudencia-Chingcuanco, D. (2007) Expression profiles of genes involved in fatty acid and triacylglycerol synthesis in castor bean (*Ricinus communis* L.). *Lipids* **42**, 263–274.

Chileh, T., Esteban-García, B., Alonso, D.L. and García-Maroto, F. (2010) Characterization of the 11S globulin gene family in the castor plant *Ricinus communis* L. *Journal of Agricultural and Food Chemistry* **58**, 272–281.

Conceicao, M.M., Candeia, R.A., Silva, F.C., Bezerra, A.F., Fernandes, V.J. and Souza, A.G. (2007) Thermo analytical characterization of castor oil biodiesel. *Renewable & Sustainable Energy Reviews* **11**, 964–975.

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

Dekkers, B.J.W., Willems, L., Bassel, G.W., Van Bolderen-Veldkamp, R.P.M., Ligterink, W., Hilhorst, H.W.M. and Bentsink, L. (2012) Identification of reference genes for RT-qPCR expression analysis in Arabidopsis and tomato seeds. *Plant and Cell Physiology* **53**, 28–37.

De Oliveira, L.A., Breton, M.C., Bastolla, F.M., Camargo, S.D.S., Margis, R., Frazzon, J. and Pasquali, G. (2012) Reference genes for the normalization of gene expression in eucalyptus species. *Plant and Cell Physiology* **53**, 405–422.

Doyle, E.A., Lane, A.M., Sides, J.M., Mudgett, M.B. and Monroe, J.D. (2007) An α-amylase (At4g25000) in Arabidopsis leaves is secreted and induced by biotic and abiotic stress. *Plant, Cell and Environment* **30**, 388–398.

Eastmond, P.J. (2004) Cloning and characterization of the acid lipase from Castor beans. *Journal of Biological Chemistry* **279**, 45540–45545.

Fan, Z., Li, J., Lu, M., Li, X. and Yin, H. (2013) Overexpression of phosphoenolpyruvate carboxylase from *Jatropha curcas* increases fatty acid accumulation in *Nicotiana tabacum*. *Acta Physiologiae Plantarum* **35**, 2269–2279.

Foyer, C.H., Neukermans, J., Queval, G., Noctor, G. and Harbinson, J. (2012) Photosynthetic control of electron transport and the regulation of gene expression. *Journal of Experimental Botany* **63**, 1637–1661.

Gong, Q., Li, P., Ma, S., Indu Rupassara, S. and Bohnert, H.J. (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. *Plant Journal* **44**, 826–839.

Gu, K., Yi, C., Tian, D., Sangha, J.S., Hong, Y. and Yin, Z. (2012) Expression of fatty acid and lipid biosynthetic genes in developing endosperm of *Jatropha curcas*. *Biotechnology for Biofuels* **5**, 47.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. and Vandesompele, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**, R19.

Jarosová, J. and Kundu, J.K. (2010) Validation of reference genes as internal control for studying viral infections in cereals by quantitative real-time RT-PCR. *BMC Plant Biology* **10**, 146.

Le, D.T., Aldrich, D.L., Valliyodan, B., Watanabe, Y., van Ha, C., Nishiyama, R., Guttikonda, S.K., Quach, T.N., Gutierrez-Gonzalez, J.J., Tran, L.S.P. and Nguyen, H.T. (2012) Evaluation of candidate reference genes for normalization of quantitative RT-PCR in soybean tissues under various abiotic stress conditions. *PLoS ONE* **7**, 9.

Li, H.L., Zhang, L.B., Guo, D., Li, C.Z. and Peng, S.Q. (2012) Identification and expression profiles of the WRKY transcription factor family in *Ricinus communis*. *Gene* **503**, 248–253.

Li, X.S., Yang, H.L., Zhang, D.Y., Zhang, Y.M. and Wood, A.J. (2012) Reference gene selection in the desert plant *Eremosparton songoricum*. *International Journal of Molecular Science* **13**, 6944–6963.

Loss-Morais, G., Turchetto-Zolet, A.C., Etges, M., Cagliari, A., Körbes, A.P., Maraschin, F.S., Margis-Pinheiro, M. and Margis, R. (2013) Analysis of castor bean ribosome-inactivating proteins and their gene expression during seed development. *Genetics and Molecular Biology* **36**, 74–86.

Maciel, F.M., Salles, C.M.C., Retamal, C.A., Gomes, V.M. and Machado, O.L.T. (2011) Identification and partial characterization of two cysteine proteases from castor bean leaves (*Ricinus communis* L.) activated by wounding and methyl jasmonate stress. *Acta Physiologiae Plantarum* **33**, 1867–1875.

Nolan, T., Hands, R.E. and Bustin, S.A. (2006) Quantification of mRNA using real-time RT-PCR. *Nature Protocols* **1**, 1559–1582.

O'Leary, B., Fedosejevs, E.T., Hill, A.T., Bettridge, J., Park, J., Rao, S.K., Leach, C.A. and Plaxton, W.C. (2011) Tissue-specific expression and post-translational modifications of plant- and bacterial-type phosphoenolpyruvate

carboxylase isozymes of the castor oil plant, *Ricinus communis* L. *Journal of Experimental Botany* **62**, 5485–5495.

Pettengill, E.A., Parmentier-Line, C. and Coleman, G.D. (2012) Evaluation of qPCR reference genes in two genotypes of *Populus* for use in photoperiod and low-temperature studies. *BMC Research Notes* **5**, 366.

Pfaffl, M.W., 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**, 509–515.

Popovici, V., Goldstein, D.R., Antonov, J., Jaggi, R., Delorenzi, M. and Wirapati, P. (2009) Selecting control genes for RT-QPCR using public microarray data. *BMC Bioinformatics* **10**, 42.

Rapacz, M., Stepień, A. and Skorupa, K. (2012) Internal standards for quantitative RT-PCR studies of gene expression under drought treatment in barley (*Hordeum vulgare* L.): the effects of developmental stage and leaf age. *Acta Physiologiae Plantarum* **34**, 1723–1733.

Rocha, A.J., Soares, E.L., Costa, J.H., Costa, W.L.G., Soares, A.A., Nogueira, F.C.S., Domont, G.B. and Campos, F.A.P. (2013) Differential expression of cysteine peptidase genes in the inner integument and endosperm of developing seeds of *Jatropha curcas* L. (Euphorbiaceae). *Plant Science* **213**, 30–37.

Salimon, J., Noor, D.A.M., Nazrizawati, A.T., Firdaus, M.Y.M. and Noraishah, A. (2010) Fatty acid composition and physicochemical properties of Malaysian castor bean *Ricinus communis* L. seed oil. *Sains Malaysiana* **39**, 761–764.

Sánchez-García, A., Moreno-Pérez, A.J., Muro-Pastor, A.M., Salas, J.J., Garcés, R. and Martínez-Force, E. (2010) Acyl-ACP thioesterases from castor (*Ricinus communis* L.): an enzymatic system appropriate for high rates of oil synthesis and accumulation. *Phytochemistry* **71**, 860–869.

Severino, L.S. and Auld, D.L. (2013) A framework for the study of the growth and development of castor plant. *Industrial Crops and Products* **46**, 25–38.

Severino, L.S., Auld, D.L., Baldanzi, M., Candido, M.J.D., Chen, G., Crosby, W., Tan, D., He, X.H., Lakshmamma, P., Lavanya, C., Machado, O.L.T., Mielke, T., Milani, M., Miller, T.D., Morris, J.B., Morse, S.A., Navas, A.A., Soares, D.J., Sofiatti, V., Wang, M.L., Zanotto, M.D. and Zieler, H. (2012) A review on the challenges for increased production of castor. *Agronomy Journal* **104**, 853–880.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**, 7.

Wei, W., Dai, X., Wang, Y., Chuan, Y., Gou, C.B. and Chen, F. (2010) Cloning and expression analysis of 1-ı-myo-inositol-1-phosphate synthase gene from *Ricinus communis* L. *Zeitschrift für Naturforschung–Section C Journal of Biosciences* **65 C**, 501–507.

Xu, Y.H., Liu, R., Yan, L., Liu, Z.Q., Jiang, S.C., Shen, Y.Y., Wang, X.F. and Zhang, D.P. (2012) Light-harvesting chlorophyll *a*/*b*-binding proteins are required for stomatal response to abscisic acid in Arabidopsis. *Journal of Experimental Botany* **63**, 1095–1106.