Application of qRT-PCR and RNA-Seq analysis for the identification of housekeeping genes useful for normalization of gene expression values during *Striga hermonthica* development

M. Fernández-Aparicio · K. Huang · E. K. Wafula · L. A. Honaas ·
N. J. Wickett · M. P. Timko · C. W. dePamphilis · J. I. Yoder · J H. Westwood

Abstract *Striga* is a root parasitic weed that attacks many of the staple crops in Africa, India and Southeast Asia, inflicting tremendous losses in yield and for which there are few effective control measures. Studies of parasitic plant virulence and host resistance will be greatly facilitated by the recent emergence of genomic resources that include extensive transcriptome sequence datasets spanning all life stages of *S. hermonthica*. Functional characterization of *Striga* genes will require detailed analyses of gene expression patterns. Quantitative real-time PCR is a powerful tool for quantifying gene expression, but correct normalization of expression levels requires identification of control genes that have stable expression across tissues and life stages. Since no *S. hermonthica* housekeeping genes have been established for this purpose, we evaluated the suitability of six candidate housekeeping genes across key life stages of *S. hermonthica* from seed conditioning to flower initiation using qRT-PCR and high-throughput cDNA sequencing. Based on gene expression analysis by qRT-PCR and RNA-Seq across heterogeneous *Striga* life stages, we determined that using the combination of three genes, *UBQ1*, *PP2A* and *TUB1* provides the best normalization for gene expression throughout the parasitic life cycle. The housekeeping genes characterized here provide robust standards that will facilitate powerful descriptions of parasite gene expression patterns.

M. Fernández-Aparicio (✉)
Department of Plant Breeding, IAS-CSIC, Institute for
Sustainable Agriculture, 14080 Córdoba, Spain
e-mail: monica.fernandez@ias.csic.es

M. Fernández-Aparicio · JH. Westwood
Department of Plant Pathology, Physiology and Weed Science,
Virginia Tech, Blacksburg, VA 24061, USA

K. Huang · M. P. Timko
Department of Biology, University of Virginia, Charlottesville,
VA 22904, USA

E. K. Wafula · L. A. Honaas · N. J. Wickett ·
C. W. dePamphilis
Department of Biology, Institute for Molecular Evolutionary
Genetics, Penn State University, State College, PA 16801, USA

N. J. Wickett
Chicago Botanic Garden, Glencoe, IL 60022, USA

J. I. Yoder
Department of Plant Sciences, University of California-Davis,
Davis, CA 95616, USA

Abbreviations

CV	Coefficient of variation
DMBQ	2,6-dimethoxy-*p*-benzoquinone
ESTs	Expressed sequence tags
GAPC-2	Glyceraldehyde-3-phosphate dehydrogenase-2
GR24	Growth regulator 24: strigol analogue
MFC	Maximum fold change of expression
PP2A	Phosphoprotein phosphatase 2A subunit A3
PPGP	Parasitic plant genome project
qRT-PCR	Quantitative real-time PCR
RLI	RNase L inhibitor protein
RNA-Seq	High-throughput sequencing of RNA
RPKM	Reads per kilobase per million mapped reads
TUB1	Beta tubulin 1
TUB5	Beta tubulin 5
UBQ1	Ubiquitin 1

Introduction

Striga is a root-parasitic weed that seriously affects the production of staple crops in Africa, India and Southeast Asia [1]. In sub-Saharan Africa it has been estimated that more than 50 million cultivated hectares of legumes and cereal crops are infected by *Striga* species, affecting more than 300 million farmers and causing annual yield losses of US$7 billion [2]. *Striga hermonthica* is the most damaging of the *Striga* species, attacking mainly sorghum, maize and pearl millet. Damage due to this parasite can range from minor to near complete crop loss annually [1].

The life cycle of parasitic plants is morphologically and physiologically specialized for heterotrophy, having evolved under strong selection for host recognition, host invasion, and nutrient acquisition. The plant family Orobanchaceae comprises the full trophic spectrum of plant nutrition, ranging from obligate holoparasites and hemiparasites, through autotrophic plants that upon specific conditions develop facultative hemiparasitism, to autotrophic non parasitic plants [3]. *Striga* species are hemiparasites that rely on host derived cues to initiate their life-cycle through germination and following attachment to the host root via the haustorium rely upon host acquired nutrients and water during their early subterranean stage of growth. Once the parasite shoot emerges from the soil, it becomes photosynthetically competent but remains linked to the host, [4]. Other parasitic life strategies developed by members of Orobanchaceae are strict holoparasitism, as represented by *Orobanche* and *Phelipanche* spp., which lack photosynthetic ability and completely depend on their hosts from germination to anthesis and seed production. In contrast, *Triphysaria* spp. are facultative parasites that can live independent of a host, but use parasitism as a energy-efficient life strategy in the presence of a suitable host.

Striga has evolved a number of attributes that enable its success as a plant parasite. *Striga* seeds require a period of after-ripening or dormancy that last several months due to incomplete maturity of the embryos at the time of seed dispersal. After complete maturation, the seeds require a period of warm and humid stratification (called seed conditioning) in order to germinate. During seed conditioning, physical, chemical and metabolic changes lead the parasite seed to break dormancy and become sensitive to host germination stimulants. The final requirement is exposure to a germination stimulant exuded by host roots [5–11].

Upon germination *S. hermonthica* seeds produce a radicle that has determinate growth based upon available seed storage reserves. In general radicle growth only extends a few millimeters in length. In the presence of a second chemical signal released by host roots termed the haustorial initiation factor the radicle stops elongating and differentiates into a terminal infective organ called the haustorium

[12]. The terminal *Striga* haustorium invades the host root and establishes a host vascular connection through which nutrients flow toward the parasite. Using the host-acquired resources the parasite develops a shoot, which emerges from the soil surface and initiates photosynthesis. In addition, new lateral roots develop that serve as anchorage roots with the capacity of form new, lateral haustoria and attachments to host roots [3]. Soon after shoot emergence, protandrous allogamous flowers develop [13].

One approach for parasitic weed control is to identify genes underlying unique parasitic processes that can be targeted through gene silencing or breeding for new host resistances, but this approach has been hampered by the lack of genomic information on parasites [3]. Recently, this has changed dramatically as the Parasitic Plant Genome Project (PPGP; [14]; http://ppgp.huck.psu.edu/) and other researchers (https://database.riken.jp/sw/en/Striga_hermonthica_EST_Database__/crib151s2rib151s128i/) have produced extensive transcriptome sequence datasets spanning the parasitic life cycle of *S. hermonthica*. Functional characterization of these genes will require the application of genomic and transcriptomic analyses focused on the examination of critical parasite life stages and tissues during the interaction of the parasite with its host. Quantitative real-time PCR (qRT-PCR) is a powerful tool for measuring gene expression and will be essential for deciphering parasitic plant developmental programs. qRT-PCR allows sensitive and specific assay of gene expression, but is underused in plants in part due to the lack of characterized genes that enable robust normalization. High-throughput sequencing of RNA (RNA-Seq) is a novel method for transcriptional profiling [15, 16]. RNA-Seq comprises significant technical advances in transcriptional profiling when compared to micro-arrays in non-model species in terms of detection range and transcriptome profiling [17]. Here we tested the suitability of candidate housekeeping genes across key life stages of *S. hermonthica* from seed conditioning to flower initiation using two different strategies: qRT-PCR and RNA-Seq.

Materials and methods

Plant material and sample collection

A total of eight life stages of the parasitic plant *Striga hermonthica* were selected for the current study. The stage identification system is the one used in the PPGP [14], with relevant stages summarized here. (A) Pre-attachment stages: StHe0: conditioned seeds; StHe1: germinated seeds; StHe2: DMBQ (2,6-dimethoxy-*p*-benzoquinone) induced seedlings. (B) Young parasitic stages during early post-attachment: StHe3: early established parasite with haustoria attached to host root establishing pre-vascular connections and StHe4:

early established parasite after vascular connection. (C) Late post-attachment: StHe5.1: underground shoots, StHe5.2: underground roots; StHe6.1: vegetative aboveground tissue and StHe6.2: reproductive aboveground tissue.

For sample collection of pre-attachment stages, *Striga hermonthica* seeds used were collected originally from plants parasitizing *Sorghum bicolor* (local landrace) in Kano, Nigeria. *Striga* seeds were surface sterilized by incubating with 10 % sodium hypochlorite for 5 min. They were then washed with 150 ml distilled water and placed onto a moistened glass fiber filter paper in a 90 mm Petri dish. The petri dish was sealed with parafilm and incubated in the dark at 30 °C for 4 days (StHe0.4), 7 days (StHe0.7) and 14 days (StHe0.14) to allow the seed to undergo conditioning. Subsequently one ml of a 0.1 ppm solution of the synthetic germination stimulant GR24 (purchased from B. Zwanenburg, Radboud University Nijmegen, the Netherlands), was applied to each Petridish containing 14 days conditioned *Striga* seeds [18]. After 16 h of incubation with GR24 at 30 °C, the tip of the parasitic seed radicle started to emerge from the seed coat and seedlings were harvested (StHe1). For collection of haustorial induced seedlings, 1 ml of 10 μmol haustorial stimulant DMBQ (Pfaltz and Bauer, Inc., Waterbury, CT 06708) was applied to each petri dish of pre-germinated *Striga* seeds [12]. The dishes were sealed, incubated at 30 °C for 16 h and then harvested (StHe2).

For sample collection of young parasitic stages during early post-attachment stages, sorghum seeds (local landrace collected in Mokwa, Nigeria) were placed on a moistened glass fiber filters in a 90 mm petri dish. The seeds were incubated at 30 °C overnight to allow them to germinate, and then placed between two blocks of moistened rockwool separated by a glass fiber filter paper. The rockwool blocks were placed in a controlled environment growth room for 7 days to allow the root and shoot systems to develop. The controlled environment growth room provided an irradiance of 500 μmol m^{-2} s^{-1} at plant height, a 12 h photoperiod and a relative humidity of 60 %. Sorghum seedlings were then transferred to rhizotrons. Each rhizotron consisted of a 150 mm^2 × 20 mm petri dish which was filled with rockwool to hold the nutrient solution. Two holes were cut at the top and bottom of the rhizotron to allow the plant to grow and to let the nutrient solution drain through. A piece of nylon mesh was placed on top of the rockwool to prevent the roots from penetrating into the rockwool. Each sorghum seedling was placed into a rhizotron, the lid replaced and the rhizotron wrapped in aluminium foil to exclude light from the roots. Plants were watered from above twice a day with a total of 50 ml of 40 % Long Ashton nutrient solution [19]. Plants were allowed to grow for 14 days before inoculating with pre-germinated *Striga* seeds, which were aligned along the host roots using a fine paint brush. Rhizotrons were then placed back in the growth room after inoculation and *Striga*

tissue was collected under a microscope using forceps. This was done at two different time points: early attachments without vascular connections were collected at 24–48 h after inoculation (StHe3; *Striga hermonthica*, haustoria attached to host root, pre-vascular connection) and early attachments with vascular connection established were collected at 72 h after inoculation (StHe4; *Striga hermonthica,* early established parasite after vascular connection).

For sample collection of late post-attachment stages. *Striga* plants were cut from the host root growing in the rhizotrons and separated into shoots (StHe5.1, *Striga hermonthica,* underground shoots) and roots (StHe5.2; *Striga hermonthica,* roots) at 7–10 days after inoculation. For collection of stages 6.1 and 6.2, *Striga* and sorghum were cultivated in pots with a diameter of 20 cm and depth of 30 cm filled with sand. Forty mg of un-preconditioned *S. hermonthica* seeds were mixed into the sand 6 cm below the surface. Five pre-germinated sorghum seeds were then planted in the pot. The pot was placed in the growth chamber (described above) and was watered with 150 ml of nutrient solution every day. *Striga* shoots emerged from the surface in 4–6 weeks time. Stage StHe6.1 tissues (*Striga hermonthica*; vegetative structures, leaves and stems) and stage StHe6.2 tissues (*Striga hermonthica*; reproductive structures; floral buds up to anthesis) were then harvested.

Two step-qRT-PCR analyses

Selection of candidate reference genes and primer design

Based on their identification as reference genes in other plant species [20–24], six candidate genes were selected for analysis from *S. hermonthica* and their coding sequences obtained from the PPGP database (http://ppgp.huck.psu.edu/). The selected genes are as follows: β tubulin1 [*TUB1*, Unigene accession StHe1GB1_52449 (705 bp)], β tubulin5 [*TUB5*, Unigene accession StHe1GB1_53023 (1343 bp)], glyceraldehyde-3-phosphate dehydrogenase-2 [*GAPC-2*, Unigene accession StHe1GB1_72584 (1185 bp)], phosphoprotein phosphatase 2A subunit A3 [*PP2A*, Unigene accession StHe1GB1_55080 (894 bp)], RNase L inhibitor protein [*RLI*, Unigene accession StHe1G2B1_39249 (1447 bp)] and ubiquitin 1 [*UBQ1*, Unigene accession StHe61FB1_199 (674 bp)].

Specific primer pairs (Table 1) were designed to amplify products of 100–150 bp with an optimal melting temperature of ∼60 °C and a GC content between 40 and 60 %. In order to avoid genomic DNA amplification, the reverse primer for each primer pair was designed spanning a predicted exon–exon junction [25, 26]. Due to the intimate contact of host and parasitic tissue in some stages of the parasitic process, we designed the primers to be parasite-specific in order to avoid amplification of host transcripts potentially present as contaminants. This allows discrimination between host and

Table 1 Oligonucleotide primer sequences used for qRT-PCR and qRT-PCR amplification efficiencies

Gene symbol	PPGP unigene number	Unigene length (bp)	Orientation	5′–3′ sequence	qRT-PCR amplicon size (bp)	Efficiency of qRT-PCR reaction
TUB1	StHe1GB1_52449	705	Forward	ATACAGTGGTCGAGCCTTAC	132	2.00
			Reverse	TAGGTGTCGTGAGCTTAAGC		
TUB5	StHe1GB1_53023	1343	Forward	AACTTCGTCTTTGGGCAGTC	140	1.94
			Reverse	ACTTGAAATCCCTGGAGAC		
GAPC-2	StHe1GB1_72584	1185	Forward	GCTGCCATCAAGGAAGAATC	105	2.00
			Reverse	TATGCTAGACCTGCTGTCAC		
PP2A	StHe1GB1_55080	894	Forward	CAAGATTCTGTGCGGTTGC	123	1.91
			Reverse	CCAAGACTTATCCTGCGAGA		
RLI	StHe1G2B1_39249	1447	Forward	AATGCCGTCAGGAGTGCA	141	1.91
			Reverse	ACGGGCATTTCTTCACACA		
UBQ1	StHe61FB1_199	674	Forward	CATCCAGAAAGAGTCGACTTTG	122	1.80
			Reverse	CATAACATTTGCGGCAAATCA		

parasite transcripts and provide accurate normalization of gene expression regardless of whether parasite tissue is isolated or embedded in host tissue. Homologous housekeeping genes were identified in the typical *Striga* host species rice, sorghum and maize by reciprocal Blast searches in NCBI and PPGP. For each gene, parasite and host sequences were aligned using Muscle in Geneious software (Biomatters Ltd), and the more divergent regions were selected as amplification targets for qRT-PCR analysis.

Total RNA isolation and cDNA synthesis

Triplicate samples of frozen tissue from stages StHe0.4, StHe0.7, StHe0.14, StHe1, StHe4, StHe5.1, StHe5.2, StHe6.1 were individually ground using a Qiagen Tissue Lyzer II and RNA was isolated using Trizol (Invitrogen) following the manufacture instructions. RNA was then DNase treated (Roche Applied Science) and purified using Qiagen RNeasy Minikit.The quantity and quality of RNA samples were assessed by Agilent Bioanalyzer 2100 analyses (Agilent Technologies, USA).

cDNA was synthesized in 20 μl reaction volumes with 1 μg of total RNA, 1× RT buffer, 4 mM dNTP mix, 1× RT random primers and 50 U multi-scribe reverse transcriptase. PCR cycling conditions followed manufacturer's protocol (High Capacity cDNA Reverse Transcription kit, Applied Biosystems).

qRT-PCR conditions

qRT-PCR was performed using SYBR Green detection. For each gene, the samples were run in a 96 well plate on a qRT-PCRsystem (Applied Biosystems7300). PCRs were performed in 20 μl total volume per well containing 10 μl of 2× SYBR® Green, 1 μl of cDNA (0.05 μg of RNA equivalents) and 1 μl (10 μM) of each gene-specific primer. PCR cycling conditions followed manufacturer's protocol (Power SYBR® Green PCR Master Mix).

An external standard curve was generated for each target gene using cDNA from *S. hermonthica* shoots. The cDNA region that covers the amplification target for each reference gene was amplified from total RNA using SuperScript one step RT-PCR with platinum TAQ kit (Invitrogen). PCR products were separated by electrophoresis through a 1 % agarose gel, yielding a single correct sized band that was excised from the gel, purified using the QIAquick Gel extraction kit (Qiagen) and its concentration quantified using nanodrop. The absolute number of target molecules was calculated using the following formula: Copy number $= C \times N_A/M$ where Copy number, number of molecules/μl contained in the purified cDNA region; C, concentration of the purified cDNA region (g/μl); M, the molecular weight of the purified region; N_A, Avogadro's number $= 6.023 \times 10^{23}$ molecules/mole.

A dilution series was created for each gene, starting from 2.0×10^8 molecules/μl of the purified cDNA target region of each candidate housekeeping gene, and diluted down to 2.0×10^1 using 8-fold serial dilutions with nuclease-free water (Sigma). The standard curve was run in the same 96-well plate with the experimental samples. The standard curve information (slope, intercept and R2) was also calculated using 7300 System SDS Software.

Data analysis

The amplification efficiency (E) of each qRT-PCR primer set (Table 1) was calculated from the standard

amplification curve [24, 27] using the equation of Ramakers et al. [28]:

$$E = 10^{(-1/k)}$$

where k is the standard curve slope calculated by the 7300 System SDS Software.

The expression levels obtained by qRT-PCR were converted to transcript copy numbers using the standard curve by the 7300 System SDS Software. Log2 transformation was performed on transcript copy number in order to allow visual comparison in the expression profile across stages graphed (Fig. 3a).

Stability analysis

For each gene, maximum fold change (MFC) and coefficient of variation (CV) of expression were calculated across stages on log2 transformed transcript copy numbers (Table 2). Genes with CV values below 0.04 and MCF values below 1.99 can be considered as stably expressed, potential housekeeping genes [29]. In addition the stability of gene expression levels across samples was calculated using the statistical algorithm geNorm [30] at default settings. The software provides a measure of gene expression stability (M). The geNorm software establishes a maximum value of M < 1.5 for assuming that a gene is stably expressed [30]. The geNorm analysis was performed on the total sample group and on each of 4 subgroups (germination, early stages of host infection, shoot vegetative development and post-haustorial development). The program also computes a normalization factor (NF) based on the geometric mean of the expression levels of the best performing reference genes. For total samples and the four subgroups, the optimal number of genes necessary for accurate normalization was calculated by geNorm using the pair-wise variation V_n/V_{n+1}.

Stability confirmation by RNA-Seq using the Illumina GA-II platform

In order to examine the reliability of gene expression data produced by qRT-PCR raw Illumina GAIIx sequence reads for stages StHe0, StHe1, StHe2, StHe3, StHe4, StHe5.1, StHe5.2, StHe6.1 and StHe6.2 (http://ppgp.huck.psu.edu/data_summary_results.php?species = StHe) were mapped onto the amplicon sequences used in qRT-PCR for each *S. hermonthica* candidate housekeeping gene.

To avoid artificial expression values, RNA-Seq data for each stage were first pre-processed in order to remove PCR duplicates that are introduced during library preparation [31]. All stage-specific filtered reads were then stringently mapped onto 100-150 bp amplicons targeted by qRT-PCR for each *S. hermonthica* candidate housekeeping gene.

Normalized measures of expression intensity, Reads Per Kilobase per Million mapped reads (RPKM) were computed from the read counts, the length of the targeted region, and the total number of mapped reads in each library or developmental stage. The high-throughput Sequencing RNA-Seq Analysis program of CLC Genomics Workbench ver. 5.0 (http://www.clcbio.com/index.php) was used for mapping and RPKM computation (parameters: length fraction = 1.0, similarity = 1.0, min insert size = 100, and max insert size = 250). By requiring mapped read fragments to be perfect matches, we eliminated reads that would have mapped elsewhere if the whole transcriptome build was used.

Log2 transformation was performed on the RPKM data in order to facilitate graphical comparisons of the expression profiles across stages (Fig. 3b). For each gene, MFC and CV of expression were calculated across stages on log2 transformed RPKM data (Table 2). Correlation between the expression values detected by RNA-Seq and qRT-PCR for the 6 genes tested and 6 *Striga* stages that were analyzed in common by both methods was estimated by calculating the Pearson correlation in the JMP501 statistical software package (SAS Institute INC., Cary, NC). This was not done for GAPC-2 expressed in stages StHe3 and StHe5.2 since no GAPC-2 transcripts were detected by RNA-seq analysis in these developmental stages.

Results

qRT-PCR amplification specificity and efficiency

Total RNA was extracted from tissues representing key stages of *S. hermonthica* life cycle and treated with DNase. All samples were analysed with an Agilent Bioanalyzer 2100 and found to contain clear 28S and 18S peaks with low noise between the peaks and a low abundance of low molecular weight species. The RNA integrity number (RIN) was above 9.0. Electropheretograms indicated minimal RNA degradation and no genomic DNA contamination in the samples. qRT-PCR was performed on eight developmental stages of *S. hermonthica* spanning the parasitic life cycle. This was done by using three independent biological replicates per stage and using the same pool of cDNA for testing specific primer pairs for 6 candidate housekeeping genes. To confirm primer quality, the primers were subjected to a dissociation analysis following the qRT-PCR cycle. The specificity of the amplifications was indicated by the single-peak melting curves of the PCR products (Fig. 1). A single significant peak was observed in all genes except for *RLI*, in which a minor secondary peak was observed in the dissociation curve, but no additional band was observed in agarose gel of qRT-PCR products (Fig. 2).

Table 2 Comparison of RNA expression stability data for 6 candidate housekeeping genes calculated by qRT-PCR and RNA-Seq

| Gene symbol | Amplicon size (bp) | qRT-PCR | | | RNA-Seq | | |
| | | Mean (transcript copy number) | Log2 transformed transcript copy number | | Mean (RPKM) | Log2 transformed RPKM | |
			CV*	MCF		CV	MCF
TUB1	132	98,716	0.027	1.30	1,010,535	0.022	1.08
TUB5	140	74,635	0.046	1.36	2,41,580	0.028	1.07
GAPC-2	105	1,33,380	0.045	1.25	60,956	0.072	1.22
PP2A	123	73,952	0.033	1.27	5,44,797	0.018	0.98
RLI	141	78,246	0.033	1.21	1,50,445	0.035	0.98
UBQ1	122	4,634,756	0.028	1.18	5,904,393	0.011	1.04

RNA-Seq values were obtained from sequences identical to those amplified for qRT-PCR

CV coefficient of variation across developmental stages, *MFC* ratio of the maximum and minimum values observed

* Seed conditioning excluded

The efficiency for each PCR primer pair was deduced from the slope of the standard curve. The maximum efficiency possible in PCR is 2 (every PCR product is replicated every cycle) and the minimum value is 1, corresponding to no amplification [25]. PCR efficiencies based on standard curve slopes indicated that tested genes had high efficiencies ranged from 1.91 to 2.00 except for *UBQ1*, which had an efficiency of 1.80 (Table 1).

Expression profile and stability across parasite life cycle

Transcription profiles obtained by qRT-PCR of the candidate housekeeping genes across each developmental stage of *S. hermonthica* were calculated as Ct and converted to transcript copy numbers (Fig. 3a). Expression levels for most genes were relatively constant across growth stages, with the potential exception of seed conditioning stages. Because seeds can be recalcitrant to RNA extraction, the possible interaction between RIN and Ct was tested but showed no effect as observed by a non-significant Pearson correlation analysis (data not shown). When considering all stages used by qRT-PCR, the MCF obtained for each gene was below the stability limit of 1.99 (Table 2). Inclusion of seed conditioning data in the qRT-PCR analysis inflated the CV for all genes above the stability limit of 0.04 (data not shown). When we excluded seed conditioning data from the analysis, all the genes except for *GAPC-2* and *TUB5* were stable, showing a CV value below the stability limit of 0.04 (Table 2).

The statistical algorithm geNorm was used to analyze the stability of transcript copy numbers across the various tissues. Tested genes were ranked according to a value stability measure (M) and analyzed across all stages as well as four substages (germination, early parasite development, shoot vegetative stage and post-haustorial development). When all *Striga* stages were analysed together, the six genes showed a

M < 0.9, which is below the default geNorm software limit of 1.5 [30] and below the M value of 1 established for heterogeneous panels [22, 32]. *PP2A*, *TUB1* and *UBQ1* genes exhibited the most stable expression levels, and *RLI* and *GAPC-2* the most variable (Fig. 4a).

Greater control-gene reliability was observed when the *Striga* life cycle was examined in distinct sections. When only parasite seed conditioning and germination were considered, all six genes showed a M < 0.8, with *PP2A*, *TUB1* and *TUB5* being the more stable genes during these stages (M ≤ 0.5), and *RLI* the most variable (Fig. 4c). Dekkers et al. [24] used a cut-off for M value of ≤0.5 for stably expressed genes during seed development in tomato and Arabidopsis.

Both *UBQ1* and *TUB5* were the more stable genes (M < 0.4) for studies of gene expression during post-attachment stages, regardless of whether developmental stages were analyzed all together (Fig. 4i), or divided into initial heterotrophic development (Fig. 4e) and vegetative shoot system (Fig. 4g). *RLI* gene expression was the most variable in vegetative shoots, whereas *PP2A* was the most variable in roots or when the all post-attachment stages were analysed together.

Optimal number of reference genes required for normalization

Using the geNorm program it is possible to compute a NF and calculate the optimal number of reference genes required for normalization. The pairwise variation between two sequential NFs V_n/V_{n+1}, starting with the most stable genes, are used to determine if adding the next most stable gene is required for proper normalization. Vandesompele et al. [30] established an estimated limit of 0.15. In the current study, pairwise variation was calculated independently using all developmental stages in the *Striga* life cycle and using the subsets of developmental stages
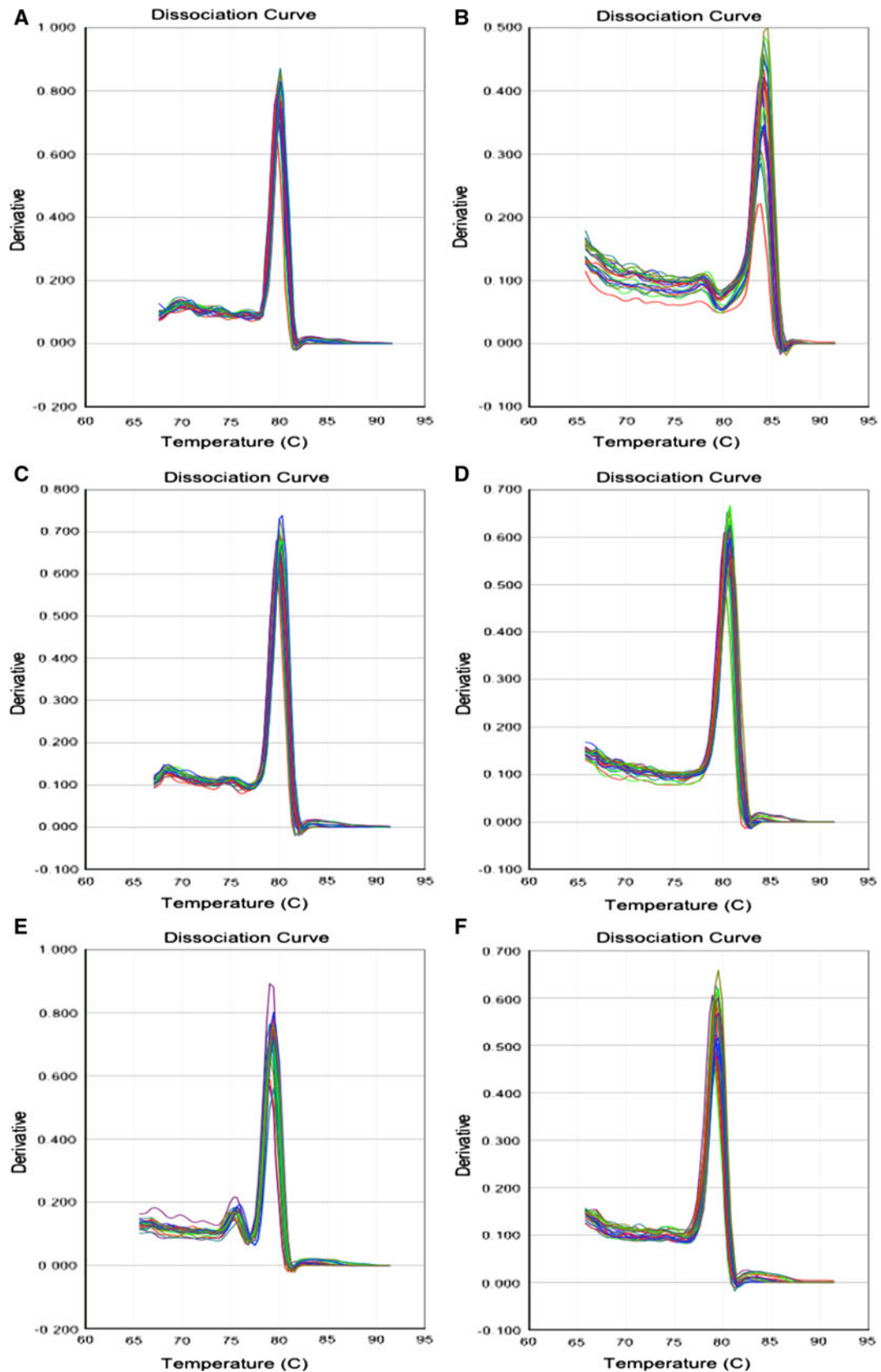
Fig. 1 Verification of amplification specificity as indicated by the single-peak melting curves of the qRT-PCR products. **a** *TUB1*, **b** *TUB5*, **c** *GAPC-2*, **d** *PP2A*, **e** *RLI*, **f** *UBQ1*
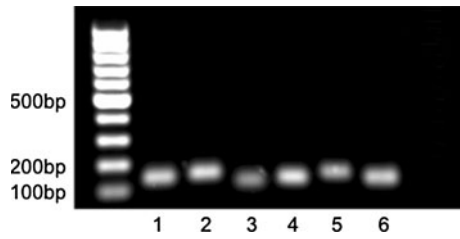
Fig. 2 Verification of amplification specificity by electrophoresis of qRT-PCR amplification products on 1 % agarose gels. Note the single sized amplification product in the various sample lanes. 1 TUB1, 2 TUB5, 3 GAPC-2, 4 PP2A, 5 RLI, 6 UBQ1

Fig. 4 GeNorm stability analysis of the gene expression profile ▶ detected by qRT-PCR. **a**, **b** all samples (from conditioning to aboveground shoot photosynthetic stage), **c**, **d** seed development (conditioning StHe 0–4, 0–7, 0–14 and germination StHe1), **e**, **f** young parasitic development (StHe4 and StHe5.2), **g**, **h** shoot vegetative stage (StHe5.1 and StHe6.1), **i**, **j** post haustorial stage (StHe4, StHe5.1, StHe5.2 and StHe6.1)

described for the M value calculations above. For all stages together, pairwise variation showed that it was necessary to include the 4 most stable reference genes for proper normalization, as $V_{2/3}$, $V_{3/4}$, $V_{4/5} > 0.15$ and $V_{5/6} < 0.15$ (Fig. 4b). However, since the $V_{4/5} = 0.154$ was close to the recommended threshold cut-off of 0.15 reported in the geNorm manual, three genes could be sufficient for normalization across the *Striga* life cycle. Taking into account only seed conditioning and germination, the pairwise

variation $V_{2/3}$, $V_{3/4} > 0.15$ and $V_{4/5} < 0.15$ meaning that the combination of the three more stable genes (*TUB1*, *TUB5*, and *PP2A*) are required for normalization during seed conditioning, germination and seedling growth (Fig. 4d). For the subset of samples comprising initial heterotrophic development and parasitic root development, $V_{2/3} < 0.15$, indicating that only one reference gene is required for normalization of this subset of stages (Fig. 4f). Normalization during shoot development would also only require the use of one reference gene as indicated by a $V_{2/3} = 0.122$ (Fig. 4h). For normalization of the complete post-haustorial subset of samples (Fig. 4j), use of the two most stable references genes will be sufficient since $V_{2/3} = 0.170$ and $V_{3/4} = 0.140$.

Fig. 3 Transcript of 6 different candidate housekeeping genes at different developmental stages in the *S. hermonthica* life cycle. **a** Transcript levels determined by qRT-PCR and presented as log2 (copy number), **b** Transcript levels determined as RPKM from RNA-Seq analysis and presented as log2

Expression stability (M)

Pairwise Variation (V)
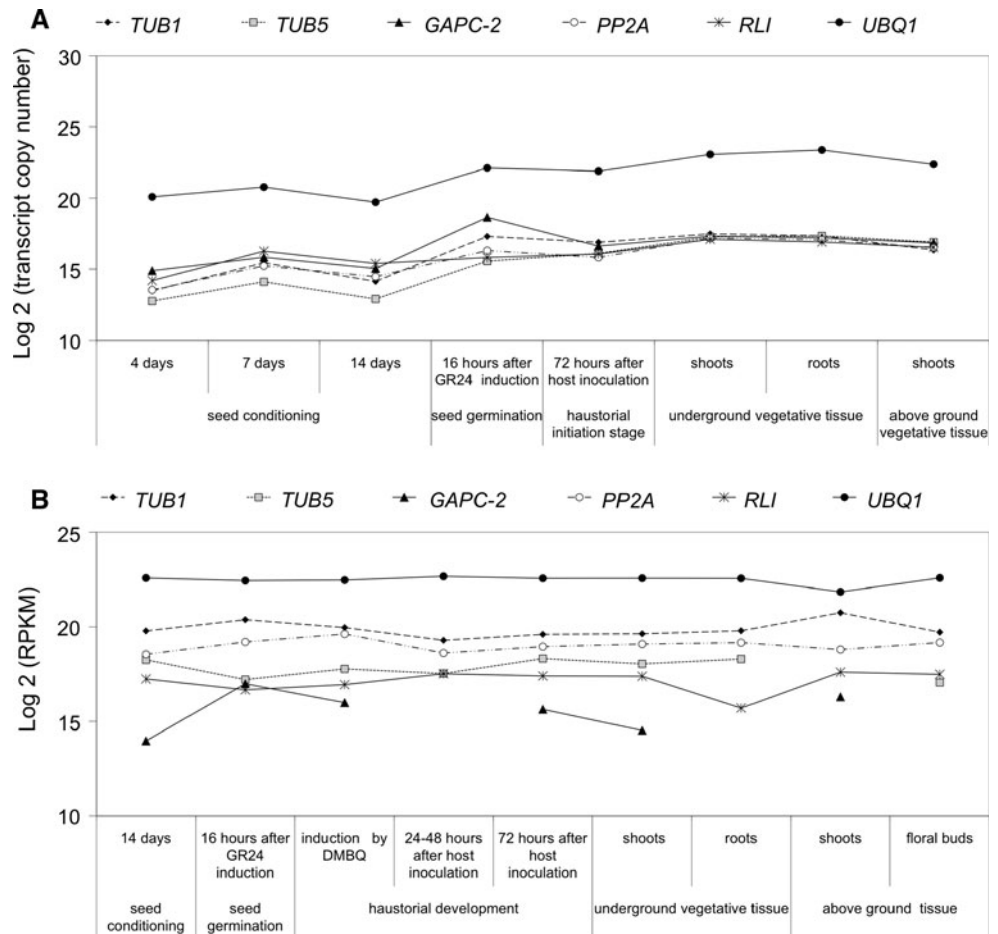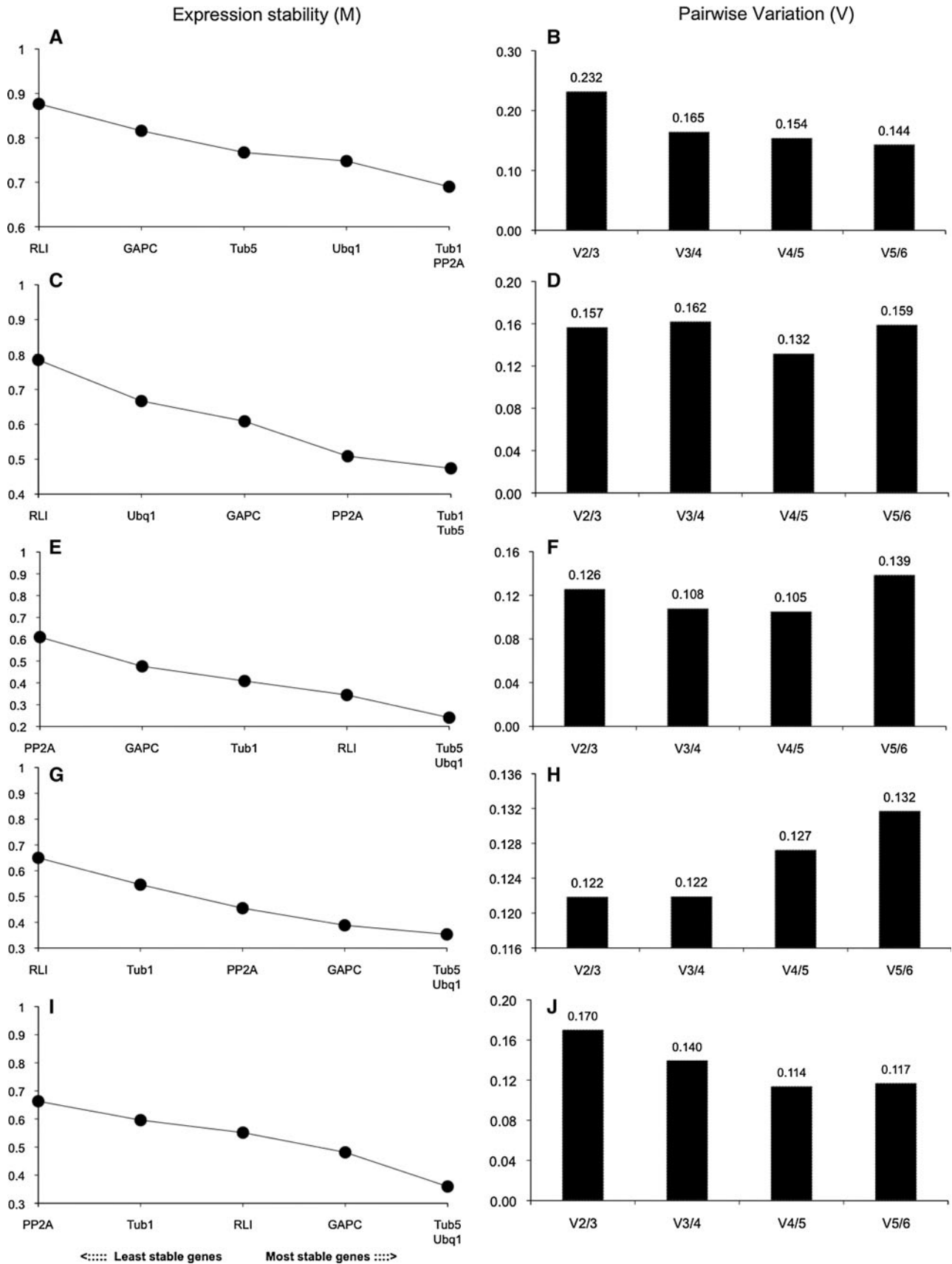
Stability confirmation by RNA-Seq analysis

The RPKM values for the 6 candidate housekeeping genes derived from the eight *S. hermonthica* libraries is presented in Fig. 3b. We found a significant Pearson correlation between the qRT-PCR and RNA-Seq data (r = 0.843, $p < 0.001$, $N = 34$) for when comparable stages are compared. Similar to what was observed by qRT-PCR analysis, *UBQ1* displayed the higher expression level across libraries spanning the *Striga* life cycle with an approximated average of 6 million RPKM. Genes with CV values < 0.04 and MCF < 1.99 can be considered as stably expressed and thus potential housekeeping genes, and RNA-Seq analysis confirmed that all genes except for *GAPC-2* met this criterion for stable expression [29] (Fig. 3b; Table 2). Remarkably, the highest stability was displayed by *TUB1, PP2A* and *UBQ1* in qRT-PCR analysis across the highly heterogeneous set of samples spanning the life cycle of *S. hermonthica* (Fig. 4a) and these were confirmed by RNA-Seq (CV ≤ 0.022 and MCF ≤ 1.08, Table 2). RNA-Seq also agreed with qRT-PCR on the 3 less stable genes found by geNorm qRT-PCRanalysis when considering all *Striga* stages. *GAPC-2*, *RLI* and *TUB5* displayed the higher CV for RNA-Seq data (Table 2) and the lower stability measured by the higher value M for qRT-PCR data (Fig. 4a).

For studies on *Striga* gene expression during seed and seedling pre-attachment development including seed conditioning, germination and haustorium induction, RNA-Seq analysis indicated that *UBQ1* was the most stable gene (CV = 0.003) followed by *TUB1*, *TUB5*, *PP2A*, and *RLI* (CV ≤ 0.03 and MCF ≤ 1.99). Again, *GAPC-2* (CV = 0.10) was the only gene studied that did not pass the cut-off value of stability CV < 0.04.

For studies of gene expression spanning the underground and aboveground post-attachment stages, all genes studied except for *GAPC-2*, were stable according to RNA-Seq data (CV ≤ 0.03 and MCF ≤ 1.99). RNA-Seq showed disagreement with qRT-PCR during *Striga* post attachment due to the low and unstable expression of *GAPC-2* (CV = 0.06).Discrepancies between qRT-PCR and RNA-Seq are expected to be higher among genes expressed at low levels [33]. In addition, low quality reads were observed during stages StHe3 (haustorial development at 24–48 h after host inoculation), stage StHe5.1 (roots) and stage StHe6.2 (floral buds) that did not pass the stringency threshold set. Consequently, no data is presented for *GAPC-2* for these three *Striga* life stages. Stages StHe3 and StHe6.2 were not analyzed by qRT-PCR but *GAPC-2* showed stable expression level beyond stage StHe4 including StHe5.1 with a M ≤ 0.5 (Fig. 3a and 4i).

An increase of more than 6-fold in the RNA expression was observed for *TUB5* during stage StHe6.1 (vegetative shoots); however, no similar increase was observed by qRT-PCR analysis. This discrepancy between the two approaches cannot be attributed to a low amplification efficiency (Table 1), nor to issues associated with low quality/integrity of RNA (since all three biological replicates of StHe6.1 RNA had integrity values >9.6). Based on this lack of correspondence, we did not consider the increase in RNA expression during stage StHe6.1 to be reliable, and therefore, we do not present the corresponding data.

Discussion

Accurate calculations of quantitative gene expression require comparison to one or more internal control genes whose expression if constant throughout development. These genes are often referred to as housekeeping genes because they perform an essential function in the cell and their expression levels are relatively constant throughout growth and development and under various metabolic programs. These make them useful for detailed dissection of changes in expression of single target genes by providing a baseline for expression normalization [26]. It is often difficult to find a universal housekeeping gene, that is, a gene that keeps stable expression across all cell types and conditions, and thus each biological system must be tested for genes that maintain relatively stable expression patterns across different developmental and physiological stages [27, 34, 35]. In this work we evaluated six *Striga* genes belonging to five different functional classes and involved in different cellular functions, such as structural constituents of cytoskeleton (*TUB1* and *TUB5*), glucose metabolism (*GAPC-2*), regulation of phosphorylation (*PP2A*), structural constituent of ribosome (*UBQ1*), and transporter activity (*RLI*) [20–22, 24], in order to minimize the chance of co-regulation [27]. The selected genes are among those previously described in other plant species as high quality reference genes.

Several technical issues were addressed at the outset to ensure proper evaluation of candidate genes. These included determining the quality, integrity, and purity of the RNA used for cDNA synthesis. This is a critical factor since compromised RNA quality can easily lead to unreliable results during gene expression analysis [27, 36, 37]. In addition, an exact estimate of concentration is required when relying on total RNA standardization. Accurate determination of RNA quality and quantity was facilitated through the use of Bioanalyzer [36–39]. For amplification we used two-step qRT-PCRas it reduces unwanted primer dimer formation when SYBR® Green is used for detection [40]. In the first step of the process, reverse transcription was carried out to produce cDNA and was initiated using random primers, which anneal preferentially to abundant species [21]. Pilot experiments had suggested that all six candidate housekeeping genes were expressed at high levels, so the challenge

of amplifying rare transcripts was not a concern. As additional quality control measures, reverse primers for each primer pair were designed to target the exon–exon junction in order to avoid amplification of genomic DNA [26] and both primers targeted the more divergent regions of host and parasitic sequences in each gene in order to avoid amplification of potential host RNA contamination.

In order to allow comparisons among each biological replicate and developmental stage sampled, the same pools of cDNAs were used for qRT-PCR amplification of all the candidate genes per developmental stage and biological replicate. The primer pairs generated single amplicons of the expected size as indicated by single bands in agarose gels and single-peak melting curves, with the only exception being that *RLI* yielded a second minor melting curve peak that could not be visualized following agarose gel electrophoresis and ethidium bromide staining. Quantification of DNA formation in qRT-PCR using SYBR$^{\circledR}$ Green I is based on monitoring the increasing fluorescence after each amplification cycle of the PCR reaction. This is based on the determination of the threshold cycle (Ct), which represents the fractional cycle number at which a fixed amount of DNA is formed [28]. This calculation assumes a constant PCR efficiency value of 2.00. However, PCR efficiency can vary over time and across samples [28]. Here, we observed a high efficiency value of 2 for *TUB1* and *GAPC-2*; the rest of the genes behave within an efficiency range between 1.91 and 1.94. Only *UBQ1* is characterized by a low efficiency of 1.80. This variation should be taken into account when studying target gene expression in order to avoid an error in fold-difference calculations.

All genes evaluated in this study were stable when analyzed by qRT-PCR, showing M values below the geNorm stability threshold. Several previous studies have pointed out that normalization based on a single gene can lead to significant errors in expression quantification, and have suggested the necessity of using several genes for confident normalization [30, 35]. Taking into account the entire life cycle from seed conditioning to above ground vegetative tissues analyzed by qRT-PCR, the genes *TUB1*, *PP2A*, *UBQ1* and *TUB5* exhibit the steadiest levels of expression. Pairwise variation showed that these four genes are necessary and sufficient for normalization of target genes across the *Striga* cycle. The use of four reference genes was also determined as necessary for normalization of expression from different physiological stages of *Cucurbita pepo* [41]. However, the pairwise variation value that described four genes as necessary for normalization of the *Striga* cycle was very near to the cut-off value of 0.15 [30] and this value should not be viewed as a strict cut-off.

In our studies, RNA-Seq analysis also identified *TUB1*, *PP2A*, *UBQ1* as stably expressed across development stages. Therefore, the combination of *TUB1*, *PP2A*, and *UBQ1* should be sufficient to normalize the whole parasitic life cycle of *S. hermonthica*. Two of these genes (*TUB1* and *UBQ1*) were also described as stable throughout the life cycle of *Phelipanche ramosa* (syn.*Orobanche ramosa*) [20]. Although *P. ramosa* is a root parasitic plant related to *S. hermonthica*, there are fundamental differences between the species, such as the nature of vascular connection (e.g., *P. ramosa* exhibits phloem continuity between host and parasite whereas this continuity is absent in *Striga*-host associations), *P. ramosa* has the ability to develop a tubercle (storage organ) whereas S. *hermonthica* does not, and photosynthesis is present in *Striga* but absent in the *Phelipanche*.

Stable expression of genes can vary with biological context [42]. That is, comparing gene expression across different developmental stages may require normalization using distinct reference genes because transcriptomes vary across different plant tissues [24]. Accordingly, we observed that the ranking of the most stable genes varied slightly across developmental subsets in *Striga*. Three genes (*TUB1*, *TUB5*, and *PP2A*) were identified stable, specifically for analyses of gene expression in seeds. The observed stability of tubulin expression during germination in *Striga* is in contrast with the results of tubulin stability during Arabidopsis and tomato seed gene expression studies, in which tubulin-4 was the least stable gene and was not recommended for normalization [24].

Based on our observations we recommend that when analyzing transcriptomic changes at the early stages of host infection by *S. hermonthica* or during Striga shoot development, the use of *UBQ1* is best, since this gene displayed the more stable expression both in qRT-PCR and RNA-Seq. For studies of gene expression post-attachment the use of 2 genes (*UBQ1* and *TUB5*) will be required for normalization.

The growing use of next-generation sequencing technology and the ability of RNA-Seq to provide gene expression information suggest that more investigators will turn to these data for assistance in identifying housekeeping genes for their research organisms. Our experience provides evidence that good correlations exist between RNA-Seq and qRT-PCR and confirms that RNA-Seq databases are a good starting point for finding control genes. An important cautionary note is that technical problems with read quality and read mapping can easily lead to the disqualification of otherwise promising candidate genes, as in the case of GAPC-2 described here.

We expect the recent increase in available *S. hermonthica* gene sequences to contribute to a proliferation of studies seeking to understand the role of specific genes in parasite development and interactions with hosts. Parasitic plants have many remarkable features, and understanding the genetic basis of this unusual lifestyle holds potential benefits for basic plant science and agriculture. This understanding will arise from the focused examination of gene expression in critical parasite life stages and tissues during the interaction of the parasite with its host. The housekeeping genes characterized

here provide robust standards that will facilitate powerful descriptions of parasite gene expression patterns.

References

1. Parker C (2009) Observations on the current status of *Orobanche* and *Striga* problems worldwide. Pest Manag Sci 65:453–459

2. Ejeta G (2007) The *Striga* scourge in Africa: a growing problem. In: Ejeta G, Gressel J (eds) Integrating new technologies for *Striga* control: toward ending the witch-hunt. World Scientific Publishing Co., Hackensack, pp 3–16

3. Westwood JH, Yoder JI, Timko MP, Depamphilis CW (2010) The evolution of parasitism in plants. Trends Plant Sci 15:227–235

4. Press MC, Smith S, Stewart GR (1991) Carbon acquisition and assimilation in parasitic plants. Funct Ecol 5:278–283

5. Brown R, Edwards M (1946) The germination of the seeds of *Striga lutea*. II. The effect of time of treatment and concentration of the host stimulant. Ann Bot 10:133–142

6. Vallance KB (1951) Studies on the germination of the seeds of *Striga hermonthica*. III. On the nature of pretreatment and after ripening. Ann Bot 15:109–128

7. Kust CA (1966) A germination inhibitor in *Striga* seeds. Weeds 14:327–329

8. Babiker AGT, Cai T, Ejeta G, Butler LG, Woodson WR (1994) Enhancement of ethylene biosynthesis and germination with thidiazuron and some selected auxins in *Striga asiatica* seeds. Physiol Plant 91(529):536

9. Babiker AGT, Ejeta G, Butler LG, Woodson WR (1993) Ethylene biosynthesis and strigol-induced germination of *Striga asiatica*. Physiol Plant 88(359):365

10. Babiker AGT, Ma Y, Sugimoto Y, Inanaga S (2000) Conditioning period, CO_2 and GR24 influence ethylene biosynthesis and germination of *Striga hermonthica*. Physiol Plant 109:75–80

11. Poneleit LS, Dilley DR (1993) Carbon dioxide activation of 1-aminocyclopropane-1-carboxylate (ACC) oxidase in ethylene biosynthesis. Postharv Biol Technol 3:191–199

12. Lynn DG, Chang M (1990) Phenolic signals in cohabitation: implications for plant development. Annu Rev Plant Physiol Plant Mol Biol 41:497–526

13. Musselman LJ, Matteson PC, Fortune S (1983) Potential pollen vectors of *Striga hermonthica* (Scrophulariaceae) in West Africa. Ann Bot 51:859–862

14. Westwood JH, dePamphilis CW, Das M, Fernandez-Aparicio M, Honaas LA, Timko MP, Wickett NJ, Yoder JI (2012) The Parasitic Plant Genome Project: new tools for understanding the biology of *Orobanche* and *Striga*. Weed Sci 60:295–306

15. Szövényi P, Rensing SA, Lang D, Wray GA, Shaw AJ (2011) Generation-biased gene expression in a bryophyte model system. Mol Biol Evol 28:803–812

16. Srivastava A, Ohm RO, Oxiles L, Brooks F, Lawrence CB, Grigoriev IV, Cho Y (2012) A zinc-finger-family transcription factor, AbVf19, is required for the induction of a gene subset important for virulence in *Alternaria brassicicola*. Mol Plant Microbe Interact 25(4):443–452

17. Schmidt A, Schmid MA, Grossniklaus U (2012) Analysis of plant germline development by high-throughput RNA profiling: technical advances and new insights. Plant J 70:18–29

18. Johnson AW, Rosebery G, Parker C (1976) A novel approach to *Striga* and *Orobanche* control using synthetic germination stimulants. Weed Res 16:223–227

19. Hewitt EJ (1966) Sand and water culture methods used in the study of plant nutrition. Commonwealth Agricultural Bureaux, Wallingford

20. Gonzalez-Verdejo CI, Die JV, Nadal S, Jimenez-Marín A, Moreno MT, Roman B (2008) Selection of housekeeping genes for normalization by real-time RT-PCR: analysis of Or-MYB1 gene expression in *Orobanche ramosa* development. Anal Biochem 379:38–43

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

22. Die JV, Roman B, Nadal S, Gonzalez-Verdejo C (2010) Evaluation of candidate reference genes for expression studies in *Pisum sativum* under different experimental conditions. Planta 232:145–153

23. Giménez MJ, Pistón F, Atienza SG (2010) Identification of suitable reference genes for normalization of qPCR data in comparative transcriptomics analyses in the Triticeae. Planta 233:163–173

24. Dekkers BJW, Willems L, Bassel GW, van Bolderen-Veldkamp RP, Ligterink W, Hilhorst HWM et al (2012) Identification of reference genes for RT-qPCR expression analysis in Arabidopsis and tomato seeds. Plant Cell Physiol 53:28–37

25. Wilkening S, Bader A (2004) Quantitative real-time polymerase chain reaction: methodical analysis and mathematical model. J Biomol Tech 15:107–111

26. Lu Y, Xie L, Chen J (2012) A novel procedure for absolute real-time quantification of gene expression patterns. Plant Methods 8:9. doi:10.1186/1746-4811-8-9

27. Silver N, Best S, Jiang J, Thein SL (2006) Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol Biol 7:33

28. Ramakers C, Ruijter JM, Deprez RH, Moorman AF (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 13:62–66

29. de Jonge HJM, Fehrmann RSN, de Bont ESJM, Hofstra RMW, Gerbens F et al (2007) Evidence based selection of housekeeping genes. PLoSONE 2(9):e898. doi:10.1371/journal.pone.0000898

30. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real time quantitative RT–PCR data by geometric averaging of multiple internal control genes.Genome Biol 3(7):RESEARCH0034

31. Kozarewa I, Ning ZM, Quail MA, Sanders MJ, Berriman M, Turner DJ (2009) Amplification-free illumina sequencing-library preparation facilitates improved mapping and assembly of (G plus C)-biased genomes. Nat Methods 6(4):291–295

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

33. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349

34. Haberhausen G, Pinsl J, Kuhn CC, Markert-Hahn C (1998) Comparative study of different standardization concepts in quantitative competitive reverse transcription-PCR assays. J Clin Microbiol 36:628–633

35. Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A, Heinen E (1999) Housekeeping genes internal standards: use and limits. J Biotechnol 75:291–295

36. Fleige S, Walf V, Huch S, Prgomet C, Sehm J, Pfaffl MW (2006) Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. Biotechnol Lett 28:1601–1613

37. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 7:3

38. Riedmaier I, Bergmaier M, Pfaffl MW (2010) Comparison of two available platforms for determination of RNA quality. Biotechnol Biotechnol Equip 24(4):2154–2159

39. Vermeulen J, De Preter K, Lefever S, Nuytens J, Derveaux S, Hellemans J et al (2011) Measurable impact of RNA quality on gene expression results from quantitative PCR. Nucleic Acids Res 39(9):e63. doi:10.1093/nar/gkr065

40. Wong ML, Medrano JF (2005) Real-time PCR for mRNA quantification. Biotechniques 39:75–85

41. Obrero A, Die JV, Roman B, Gomez P, Nadal S, Gonzalez-Verdejo CI (2011) Selection of reference genes for gene expression studies in Zucchini (*Cucurbita pepo*) Using qPCR. J Agric Food Chem 59(10):5402–5411

42. Hruz T, Wyss M, Docquier M, Pfaffl MW, Masanetz S, Borghi L et al (2011) RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC Genomics 12:156