Planta (2019) 250:2101–2110
https://doi.org/10.1007/s00425-019-03289-x

ORIGINAL ARTICLE

Selection of miRNA reference genes for plant defence studies in rice (*Oryza sativa*)

Bruno Verstraeten[1] · Lien De Smet[1] · Tina Kyndt[1] · Tim De Meyer[2]

Received: 6 August 2019 / Accepted: 26 September 2019 / Published online: 3 October 2019

Abstract

Main conclusion **MicroRNAs miR390-5p, miR7694-3p miR1868 and miR1849 were found to be suitable miRNA reference genes for rice, under either infection with the root-knot nematode *Meloidogyne graminicola* or treatment with BABA.**

Abstract RT-qPCR is a widely used method to investigate the expression levels of genes under certain conditions. A key step, however, to have reliable results is the normalization of expression. For every experimental condition, suitable reference genes must be chosen. These reference genes must not be affected by differences in experimental conditions. MicroRNAs are regulatory RNA molecules, able to direct the expression levels of protein coding genes. In plants, their attributed functions range from roles in development to immunity. In this work, microRNAs (miRNAs) are evaluated for their suitability as reference genes in rice after infection with root-knot nematode *Meloidogyne graminicola* or after priming with beta-amino butyric acid. The evaluation was based on their amplification efficiency and their stability estimates according to geNorm, NormFinder and BestKeeper. All tested miRNAs, excluding one, were considered acceptable for normalization. Furthermore, miRNAs were validated using miRNA sequencing data. The set of microRNAs miR390-5p and miR7694-3p was found to be the most stable combination under the tested conditions. Another miRNA set consisting of miR7694-3p, miR1868 and miR1849 also shows potential to be used for miRNA expression normalization under experimental conditions beyond the scope of this study. This work is the first report on reference miRNAs in rice for the purpose of plant defence studies.

Keywords RT-qPCR · BABA · Rice · *Meloidogyne graminicola* · Plant defence

Abbreviations

BABA	Beta-amino butyric acid
CPM	Counts per million mapped reads
CV	Coefficient of variation

Introduction

Rice (*Oryza sativa*) is a staple food worldwide with an annual yield of over 770 million tons (FAO 2017). It is mainly grown under anaerobic conditions, such as in rice paddies, and faces constant challenges from biotic stresses, such as fungi and nematodes. Nematodes alone can cause yield losses up to 25% (Bridge et al. 2005). Because of its detailed gene annotation and relatively small genome of ca. 420 Mb, it is often used as a monocot model organism. At the same time, there is an interest in exploring new avenues for biotechnological improvement of rice yield and immunity. These studies are increasingly focusing on microRNAs (miRNAs), 20–24-nt-long RNAs that regulate gene expression by targeting complementary mRNAs. In plants, this is generally followed by RNA-Induced Silencing Complex-directed cleavage of the mRNA, while animal miRNAs mainly function by inhibiting translation (Millar and Waterhouse 2005). Plant microRNAs play important

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00425-019-03289-x) contains supplementary material, which is available to authorized users.

✉ Tina Kyndt
Tina.Kyndt@UGent.be

[1] Department of Biotechnology, Ghent University, Ghent, Belgium

[2] Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium

roles in numerous processes ranging from development to stress responses (Djami-Tchatchou et al. 2017). In rice, for example, overexpression of miR319 led to increased cold tolerance, while expression of miR397 has a positive effect on yield parameters such as grain size and panicle branching (Yang et al. 2013; Zhang et al. 2013; Wang et al. 2014). To understand their functionality, targeting effectiveness and role in biological pathways, it is critical that the expression of miRNAs is accurately assessed.

The most commonly used method for the evaluation of gene expression is reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR). To account for technical variation due to factors such as differences in sample size, pipetting errors and sample quality, normalization is a crucial step in expression analysis (Pritchard et al. 2012). Mean expression value normalization can be used but is only valid if the number of miRNAs tested is sufficiently large and if there is no general trend towards miRNA over- or under-expression. This strategy is, therefore, generally not applicable to experiments that focus on a specific selection of miRNAs (Mestdagh et al. 2009). A very common alternative normalization technique is to use endogenous reference genes, which are known to be stably expressed under the conditions analyzed. Ribosomal RNA genes such as 18S ribosomal RNA are often chosen for this purpose (Chugh and Dittmer 2012). Ideally, genes used for normalization should belong to the same RNA class as the genes of interest, to avoid bias created by differences in extraction efficiency, reverse transcription and/or PCR amplification (Chugh and Dittmer 2012). Furthermore, reference genes should be validated for new experimental conditions since their expression may be treatment dependent (Kozera and Rapacz 2013), which has been demonstrated even for commonly used reference genes such as actin and GAPDH (Schmittgen and Zakrajsek 2000; Selvey et al. 2001). There is evidence that no gene can be universally used as a reference gene, which emphasizes the need for systematic validation of reference genes (Gutierrez et al. 2008).

In this work, several miRNAs are tested for their suitability as reference genes in rice plant defence studies. A first treatment assessed constitutes the infection of rice plants with the root-knot nematode *Meloidogyne graminicola*. This endoparasitic pathogen is able to penetrate roots after which it creates feeding structures known as giant cells through the reorganization of vascular cells. These cells go through multiple mitosis cycles, but without cytokinesis, resulting in large multinucleic cells. The cells also undergo metabolic and developmental reprogramming, while the defense response is suppressed (Gheysen and Mitchum 2011). As second treatment, beta-amino butyric acid (BABA) was applied on rice plants. BABA is a well-known plant defence activating molecule against both biotic as well as abiotic stresses (Slaughter et al. 2012; Ji et al. 2015; Jisha and Puthur 2016; Buswell et al. 2018). Candidate reference miRNAs were evaluated in both conditions. The best scoring pair of candidates was validated using a publicly available miRNA-sequencing dataset of rice under infection by rice stripe virus. This allows to compare the performance of the candidate reference miRNAs with state-of-the-art RNA-sequencing normalization strategies by means of standard quality control plots and statistics.

Materials and methods

Plant materials

Seeds of *Oryza sativa* cv. 'Nipponbare' (GSOR 100, USDA) were germinated for 3 days at 30 °C on paper cloths that were drenched with tap water. Seedlings were transferred into SAP substrate (sand-absorbent polymer) (Reversat et al. 1999) and grown at 26 °C under a 16-h/8-h light/dark regime. We aimed at comparing three treatment groups. Firstly, at 14 days old, each plant was inoculated with ca. 200 stage 2 juveniles of *Meloidogyne graminicola*. Galls were collected 3 days after inoculation, forming the first group (day 17). Secondly, at 16 days, plants were root-drenched with 3.5-mM BABA (Sigma Aldrich) dissolved in distilled water. Root tips were collected 24 h after application, composing the second treatment group (day 17). Thirdly, at 14 days old, control plants were mock-treated with distilled water. Root tips were collected after 3 days, i.e., the control group (day 17). Plant materials were immediately frozen in liquid nitrogen and stored at − 80 °C. Each treatment consisted of four biological replicates. Each biological replicate consisted of ten plants.

Selection of candidate reference genes

Candidate reference genes were selected from an in-house small RNA sequencing dataset [to be published elsewhere, Verstraeten et al. (in preparation)]. The selection was made based on q-values relating to the significance of expression changes of miRNAs between roots of uninfected rice plants and galls of rice plants 3 days post inoculation, i.e., non-differentially expressed miRNAs were selected. A second filtering step was done by selecting mature miRNAs whose sequences are unique in the rice genome, which was verified by the BLAST service of miRbase (http://www.mirbase.org/) (Kozomara et al. 2019). The names of the candidate reference genes are miR166m, miR5149, miR164e, miR156l-5p, miR7694-3p, miR535-3p, miR1876, miR390-5p, miR168a-3p, miR1882e-3p, miR1849 and miR1868. The miRBase accession numbers of these miRNAs are presented in Supplementary Table S1.

Primer design

RT-qPCR was performed using the stem-loop PCR method of Varkonyi-Gasic et al. (2007). Therefore, two primers were designed for each candidate reference gene: a stem-loop primer for the reverse transcriptase step and a forward primer for amplification. A universal reverse primer was used. All primers were manufactured by Sigma Aldrich. The list of primers can be found in Supplementary Table S1.

RT-qPCR

Frozen samples were ground and total RNA was extracted using the ZR Plant RNA Miniprep kit (Zymo Research) according to the manufacturer's instructions. The bead beater step was performed using a FastPrep-24 homogenizer (MP Biomedicals) at a speed of 4 m/s for 45 s. All centrifugal steps were performed at 16,000*g*.

For DNase treatment, 1 µg of total RNA per sample was combined with 3.6 µL DNase I buffer + $MgCl_2$ (10×, B43, Thermo Scientific), 1 µL RiboLock RNase Inhibitor (40 U/µL, EO0381, Thermo Scientific) and 1 µL DNAse I (1 U/µL, EN0521, Thermo Scientific). Afterwards, RNase-free water was added until a total volume of 36 µL followed by incubation for 30 min at 37 °C. Finally, 4 µL EDTA (25 mM, Thermo Scientific) was added before incubating for 10 min at 65 °C. RNA quality was checked using a NanoDrop 2000 spectrophotometer (Thermo Scientific).

Reverse transcription (RT) was performed using the Tetro cDNA synthesis kit (Bioline, Memphis, TN, USA). For every RT reaction, two mixes were made. The first mix consisted of 0.5 µL 10 mM dNTP mix (RNAse free), 8.5 µL nuclease free H_2O and 1 µL 1 mM of appropriate stem-loop primer (RNAse free). This mix was incubated at 65 °C for 5 min after which it was cooled on ice for 2 min. The second mix consisted of 4 µL RT buffer, 0.5 µL Ribosafe RNase inhibitor (10 U/µL) and 0.25 µL Tetro reverse transcriptase (200 U/µL). The two mixes were subsequently combined with 1 µL RNA and 4.25 µL H_2O for a total reaction volume of 20 µL. Afterwards, samples were loaded into a T100 thermal cycler (Bio-Rad) and incubated for 30 min at 16 °C, followed by pulsed RT of 60 cycles at 30 °C for 30 s, 42 °C for 30 s and 50 °C for 1 s.

qPCR was performed using the SensiMix SYBR Hi-ROX kit (Bioline). Three technical replicates were used per sample. Each sample mix consists of 10 µL 2 × SensiMix, 1 µL 10 µM forward primer, 1 µL 10 µM reverse primer, 1 µL cDNA sample and 7 µL H_2O. Non template controls (NTCs) were made consisting of 10 µL 2 × SensiMix, 1 µL 10 µM forward primer, 1 µL 10 µM reverse primer and 8 µL H_2O. Sample mixes and NTCs were appropriately combined in a 96-well plate using a CAS-1200 Liquid Handling Robot (Corbett Robotics). Finally, the plate was inserted into a CFX Connect Real-Time system (Bio-Rad). The thermocycling steps were as follows:

1. 95 °C for 10 min
2. 95 °C for 25 s
3. 58 °C for 25 s
4. 72 °C for 20 s

Repeat 49 times

5. Increment temperature from 65 to 95 °C to obtain melting curves.

Data analysis

CFX Manager (version 3.1.1217.0823) was used for qPCR analysis. C_q values were determined using the regression modus. Uniformity of the melting curves was assessed as a quality control step. Amplification curves were fitted using the CFX Manager. The amplification efficiency was calculated using the value of the slope of the fitted amplification efficiency curves with the following formula:

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

The standard error (SE) of the efficiency values was calculated using a Taylor's series approximation (Gene Quantification Platform 2019):

$$\text{SE}(E) = \frac{E \times \log_e(10) \times \text{SE(slope)}}{\text{slope}^2}$$

$$\text{with SE(slope)} = \sqrt{\frac{\sum_{i=1}^{N} \sum \left(Y_i - Y_i'\right)^2}{N}}$$

Slope is the slope of the fitted amplification curve, Y and Y', respectively, represent actual and predicted Cq values; while N is the number of data points per amplification curve.

geNorm, NormFinder and BestKeeper were used for expression stability assessment (Vandesompele et al. 2002; Andersen et al. 2004; Pfaffl et al. 2004). geNorm was used as part of qbase+ that was installed on a local server. The expression stability of candidate reference genes is estimated by NormFinder through the calculation of a stability value based on intra- and intergroup variation. Lower intra- and intergroup variation will result in a lower stability value, which reflects more stable expression, essential for candidate reference genes. For NormFinder, C_q values of technical replicates were averaged and the averages were transformed into Q values. This transformation is necessary since NormFinder expects the input values to be on a linear scale. The following formula was used:

$$Q = E^{(C_{q,\min} - C_q)}$$

$C_{q,min}$ is the lowest C_q value for a candidate reference gene across all samples.

BestKeeper makes pair-wise correlations between all samples as well as between samples and an index that is calculated as the geometric mean of the *C*q values. A higher coefficient of correlation indicates a more stable reference gene.

Validation

A relevant miRNA-sequencing dataset of rice infected with rice stripe virus was obtained from the GEO repository (Accession No. GSE74498) (Yang et al. 2016). All treatments and controls had three biological replicates. For all samples, trimming was done with Trimmomatic (v0.38) using the following parameters: ILLUMINACLIP:3:30:10, MAXINFO:23:1, SLIDINGWINDOW:5:30, MINLEN:17 (Bolger et al. 2014). STAR (v2.6.1d) was used for mapping using the following parameters: —outFilterMismatchNoverLmax 0.05, —outFilterMatchNmin 16, —outFilterScoreMinOverLread 0, —outFilterMatchNminOverLread 0, —alignIntronMax 1 (Zaleski et al. 2012). These settings were also used within the ENCODE project specifically for small RNA mapping (Dobin 2013). Afterwards, samtools (v1.3) was used to merge multiplexed samples and extract unique mappings (Li et al. 2009). A GTF file was downloaded from miRbase (v22) with all known mature rice miRNas. Count tables were generated by the summarizeOverlaps function in the GenomicAlignments R package (v1.16.0) (Lawrence et al. 2013). The following options were used for counting: mode='Union' and singleEnd = TRUE.

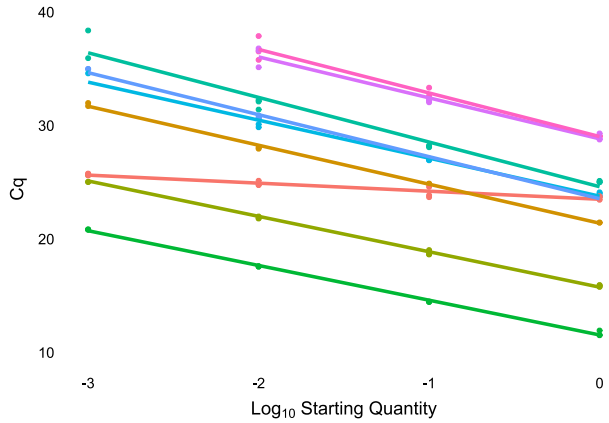
Multiple normalization methods were used: Firstly, naive library size scaling, performed by calculating count per million mapped reads (CPM) per sample. Secondly, the default DESeq2 (v1.22.2) normalization was applied, which involves calculation of the size factors determined by the median ratio of gene counts relative to geometric mean per gene (Love et al. 2014). Thirdly, edgeR (v 3.24.3) normalization, the effective library size was estimated using the trimmed mean of *M* values method and taken into account for CPM value calculation (Robinson and Oshlack 2010). Finally, reference gene miRNA normalization, performed by calculating a scaling factor which is the geometric mean of the counts of the selected reference genes per sample. A pseudocount of 1 was added to the count table to enable use of logarithmic transformation. Differences between normalization strategies were evaluated by considering three types of plots: MA plots, density plots and scree plots.

Results

Twelve miRNA genes were selected to test their suitability as reference genes in *Oryza sativa*. The setup included three experimental conditions, galls sampled at three days after inoculation with *Meloidogyne graminicola*, soil drenching with 3.5 mM BABA and mock-treated control plants. As a quality control step, the melt curves of the candidate reference genes were analyzed. For miR5149 and miR164e, no unimodal melt peaks could be generated; whereas for miR156l-5p, melt peaks were not uniform over all samples. Therefore, these genes were discarded from further analysis.

Amplification efficiency

The amplification efficiency of the candidate reference genes was determined using a tenfold dilution series of a pooled cDNA sample containing all samples over all experimental



miRNA	E	SE(E)	R²
miR166m	25.734	29.787	0.907
miR168a-3p	1.955	0.064	0.998
miR1849	2.092	0.102	0.997
miR1868	2.121	0.078	0.998
miR1876	1.796	0.204	0.968
miR1882e-3p	1.989	0.179	0.984
miR390-5p	1.863	0.076	0.996
miR535-3p	1.893	0.165	0.973
miR7694-3p	1.822	0.156	0.971

Fig. 1 Amplification efficiency curves. The expression levels of the candidate reference genes were assessed in a tenfold dilution series of a pooled sample of all conditions. Three technical replicates were used. Amplification efficiency curves were fitted to the Cq values using linear regression and are shown on a log scale. The corresponding efficiency value (*E*), standard error of the efficiency value [SE(*E*)] and coefficient of determination of the fitted curve (R^2) are shown per candidate reference gene. C_q values > 40 were omitted

conditions (Fig. 1). *C*q values higher than 40 were deemed unreliable and were omitted. A clear outlier was miR166m which had an amplification efficiency of 2473.4% and was, thus, discarded. The amplification efficiencies of the other candidate reference genes ranged from 79.2 to 112.1%, while the correlation coefficients of their fitted dilution curves varied from 0.968 to 0.998.

geNorm analysis

geNorm utilizes two metrics to evaluate the reference potential of candidate genes. The first metric is the *M* value which signifies the expression stability of a gene relative to the other genes that are tested. A higher *M* value indicates lower stability. The second metric is the coefficient of variation (CV), which indicates how stably a gene is expressed across conditions. A higher CV value indicates lower stability.

The *M* and CV values are shown in Table 1. *M* values vary between 0.342 and 0.991 across conditions, while CV values vary between 0.115 and 0.578. Candidate reference genes miR390-5p and miR7694-3p score the best: in terms of *M* value, they are among the best three candidates in every sample grouping with miR7694-3p topping the list in two sample groupings (Infection and BABA), while miR390-5p has the lowest *M* value in the All grouping. In terms of CV value, miR7694-3p and miR390-5p are the best candidates in every sample grouping.

The *M* value of a candidate reference gene is dependent on the set of samples that are considered in the analysis as well as on the stability of other candidate reference genes. If samples of all conditions are taken into account, i.e., BABA treatment, infection and control (All grouping), then miR7694-3p and miR390-5p are the most stable reference genes after stepwise exclusion of the least stable candidate reference gene (Fig. 2a). Similarly, if only control and infection samples are taken into account (Infection grouping), then miR1882e-3p and miR1876 are the most stable reference genes (Fig. 2b). Lastly, if only control and BABA treatment samples are considered (BABA grouping), then miR7694-3p and miR1882e-3p are identified as the most stable candidates (Fig. 2c). Candidate miR535-3p is the least stable gene in the All and Infection grouping, while miR1849 was the least stable candidate in the BABA grouping.

It is preferred to use multiple reference genes for the normalization of qPCR data (Vandesompele et al. 2002). To determine the optimal number of reference genes, *V* values are calculated per added gene to the reference gene set. *V* values represent the pairwise variation between a normalization factor based on a set of reference genes and a normalization factor that is based on that same set of reference genes and one additional reference gene. Thus, the *V* value reflects to which extent expression normalization is affected by adding an extra gene to the set of reference genes. If the *V* value is lower than 0.15, it is no longer considered beneficial to add an extra gene to the reference gene set. In every sample combination, the $V_{2/3}$ value is already below the 0.15 threshold indicating that two reference genes are sufficient (Fig. 2d) (Vandesompele et al. 2002).

NormFinder

NormFinder ranks candidate reference genes by calculating the inter- and intragroup variation for every gene and combining these metrics into a stability value (Andersen et al. 2004). Lower stability values indicate stabler gene expression. In all groupings, miR390-5p and miR7694-3p are the top two most stable reference genes, while miR168a-3p ranks third (Table 2). In the BABA grouping, miR390-5p and miR7694-3p share the first place, since they have an identical stability score. Overall, stability values varied between 0.104 and 0.549.

Table 1 *M* and CV values of candidate reference miRNAs for all sample groupings

	M			CV		
	All	Infection	BABA	All	Infection	BABA
miR168a-3p	0.529	0.540	0.377	0.208	0.253	0.123
miR1849	0.680	0.523	0.731	0.349	0.232	0.433
miR1868	0.593	0.551	0.382	0.328	0.292	0.141
miR1876	0.474	0.454	0.361	0.164	0.194	0.132
miR1882e-3p	0.527	0.526	0.376	0.257	0.286	0.143
miR390-5p	0.462	0.441	0.367	0.123	0.146	0.116
miR535-3p	0.991	0.818	0.689	0.578	0.541	0.401
miR7694-3p	0.475	0.440	0.342	0.150	0.148	0.118

Three sample groupings were used: *All* all samples, *Infection* infected + control samples, *BABA* BABA treated + control samples. Four biological replicates were used per treatment. Three technical replicates were used per biological replicate

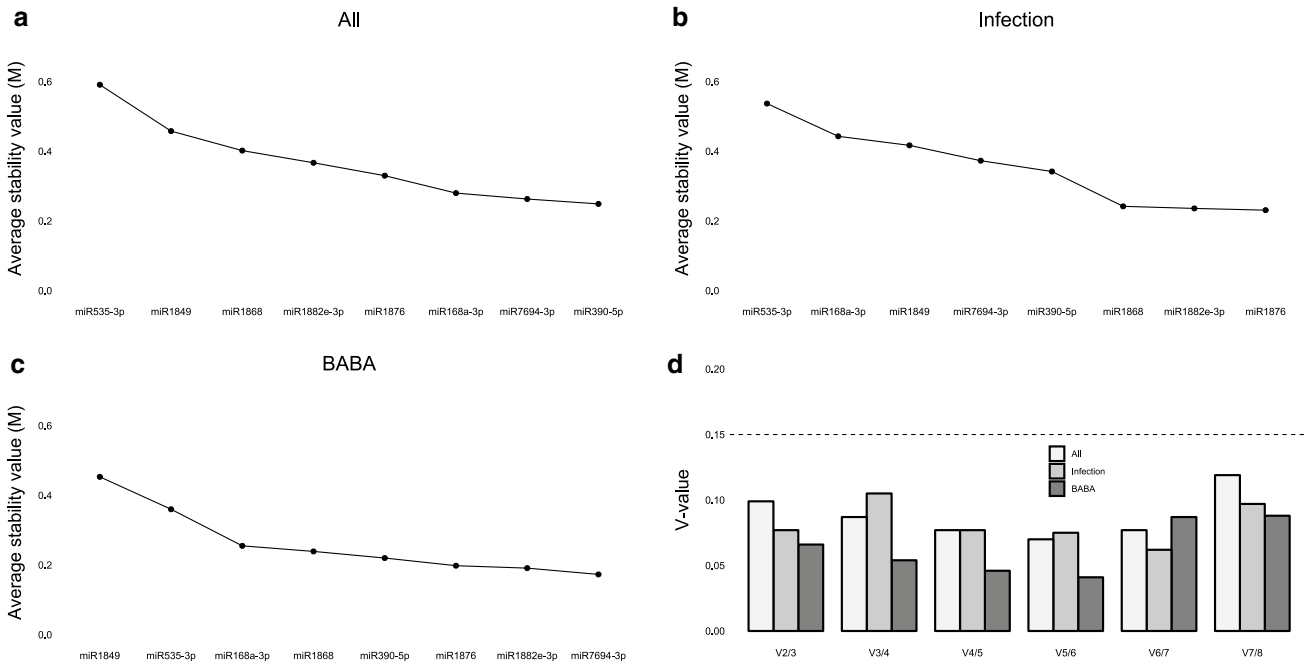


Fig. 2 geNorm assessment of candidate reference gene stability. Change of average *M* value of the candidate reference gene set after stepwise exclusion of the least stable reference gene involving **a** all samples, **b** infected + control samples, **c** BABA treated + control samples. **d** Change in *V* value after stepwise inclusion of candidate reference genes in the reference gene set. Four biological replicates were used per treatment. Three technical replicates were used per biological replicate

BestKeeper

BestKeeper uses a multi-step approach towards the determination of the most stable reference genes (Pfaffl et al. 2004). Firstly, it labels candidate reference genes as stable or unstable based on its SD (standard deviation). An SD higher than 1 marks a gene as unstable. The remaining genes are used to create an index by taking a geometric mean per sample of the C_q values. Then, a Pearson correlation coefficient is calculated between each candidate reference gene and the index which is used to rank the candidate reference genes (Table 2). A higher correlation coefficient indicates a higher stability. The BABA grouping shows lower correlation coefficients (0.016–0.725) than the All grouping (0.185–0.906) or Infection grouping (0.293–0.958). Notably, miR390-5p is considered to be the most stable reference gene in two (All, BABA) of the three sample groupings.

The rankings of the candidate reference genes obtained by analyses in geNorm, NormFinder and BestKeeper are combined to determine an overall stability ranking: per candidate reference gene, rankings are used to calculate a geometric mean (Table 3). For geNorm, rankings from the stepwise exclusion analysis were used (Fig. 2a–c). Based on this comparison, miR390-5p and miR7694-3p are the highest ranked and the final choice as reference genes.

Validation

To validate the suitability of miR390-5p and miR7694-3p as miRNA reference genes, a public miRNA-sequencing dataset of rice under infection by rice stripe virus was used (Yang et al. 2016). This strategy allows to use common quality control methods to assess the performance of the reference genes in comparison with 4 normalization strategies: No normalization, naive library size normalization, DESeq2 normalization and edgeR normalization (trimmed mean of *M* values).

First, MA plots were created to compare the effects of the normalization strategies. As can be seen in Fig. 3a, the miRNA normalization performs equally well as the other normalization methods. Log_2 fold change values show a symmetrical spread and are centered around 0 for increasing counts. A successful normalization of count data results in highly overlapping count distributions of the samples. Additionally, density plots were generated for all samples (Fig. 3b). Unnormalized counts and counts normalized by edgeR exhibit highest degree of overlap of density profiles between samples. Sample density profiles of miRNA normalized counts show a slight spread but are still highly overlapping. Finally, normalization effects were verified using principal component analysis. Successful normalization reduces technical dependency between variables, resulting in a reduced amount of variance that can be explained by a

Table 2 Ranking of candidate reference genes according to NormFinder and BestKeeper using stability values and Pearson correlation coefficients (*r*), respectively

NormFinder						BestKeeper					
All		Infection		BABA		All			Infection			BABA		
miRNA	Stability value	miRNA	Stability value	miRNA	Stability value	miRNA	SD	r	miRNA	SD	r	miRNA	SD	r
miR390-5p	0.118	miR7694-3p	0.104	miR7694-3p	0.110	miR390-5p	0.44	0.906	miR1849	0.54	0.958	miR390-5p	0.24	0.725
miR7694-3p	0.141	miR390-5p	0.115	miR390-5p	0.110	miR535-3p	1.05	0.9	miR535-3p	1.01	0.923	miR168a-3p	0.24	0.719
miR168a-3p	0.163	miR168a-3p	0.197	miR168a-3p	0.114	miR7694-3p	0.44	0.879	miR7694-3p	0.47	0.894	miR1868	0.16	0.621
miR1876	0.171	miR1876	0.199	miR1882e-3p	0.129	miR168a-3p	0.39	0.846	miR390-5p	0.43	0.880	miR535-3p	0.63	0.542
miR1882e-3p	0.246	miR1849	0.242	miR1876	0.143	miR1876	0.31	0.801	miR168a-3p	0.47	0.795	miR7694-3p	0.15	0.469
miR1868	0.325	miR1882e-3p	0.270	miR1868	0.151	miR1849	0.48	0.648	miR1876	0.26	0.794	miR1849	0.49	0.385
miR1849	0.352	miR1868	0.326	miR535-3p	0.365	miR1882e-3p	0.19	0.559	miR1882e-3p	0.19	0.494	miR1876	0.19	0.353
miR535-3p	0.549	miR535-3p	0.458	miR1849	0.432	miR1868	0.16	0.185	miR1868	0.15	0.293	miR1882e-3p	0.11	0.016

Three sample grouping were used: *All* all samples, *Infection* infected + control samples, *BABA* BABA treated + control samples. Four biological replicates were used per treatment. Three technical replicates were used per biological replicate

SD standard deviation

single principal component. Scree plots shown in Fig. 3c demonstrate that the distribution of the variance percentages explained by the principal components is the most evenly spread in the case of miRNA normalization. Another selection of miRNAs was chosen to be tested as reference genes for their potential to be used in experimental conditions beyond those described in this work: miR7694-3p, miR1868 and miR1849 have, to the best of our knowledge, not yet been described as differentially expressed under abiotic and biotic stress conditions. Normalization was done by scaling counts sample-wise with the geometric mean of counts of miR7694-3p, miR1849 and miR1868 and performed on par with the other normalization techniques (Supplementary Fig. S1).

Discussion

For reliable RT-qPCR results, normalization is a critical step in the analysis. The Minimum Information for publications of Quantitative real-time PCR Experiments (MIQE) guidelines recommend the use of multiple reference genes of which the optimal number and choice must be experimentally determined (Bustin et al. 2009). In this study, 12 miRNAs were evaluated for their potential as reference miRNAs under plant defence affecting conditions, namely nematode infection and treatment with BABA. Overall, miR390-5p and miR7694-3p were determined to be the most stable reference genes. While NormFinder does not suggest a cut-off in stability value, geNorm and BestKeeper do have guidelines: in a geNorm analysis, *M* and CV values would ideally be lower than 0.5 and 0.25, respectively, and be acceptable up to values of 1 and 0.5, respectively (Hellemans et al. 2007). All candidate reference genes can be considered acceptable in all sample groupings (Table 1). Candidate reference genes miR390-5p, miR7694-3p and miR1876 have M and CV values below the "ideal" threshold for all sample groupings, while miR168a-3p, miR1868 and miR1882e-3p are considered ideal in the BABA sample grouping. BestKeeper, on the other hand, considers genes with a standard deviation > 1 to be unstable. Only miR535-3p has a standard deviation greater than 1 (in the All and Infection sample grouping, Table 2). The standard deviations of the other candidate reference genes were well below the threshold. This indicates that, with the exception of miR535-3p, all candidate reference genes show a stable expression and have potential to be used for normalization.

MicroRNA miR390-5p has also been shown to be a suitable reference miRNA in cotton under biotic stress (Fausto et al. 2017). However, since miR390-5p has been shown to be differentially expressed in rice under a number of conditions such as heavy metal stress, drought stress, salt stress, UV stress and infection by *Magnaporthe grisea,* it

Table 3 Rankings by geNorm, NormFinder and BestKeeper per sample grouping for each candidate reference gene

	geNorm			NormFinder			BestKeeper			Geometric mean
	All	Infection	BABA	All	Infection	BABA	All	Infection	BABA	
miR390-5p	1	4	4	1	2	1	1	4	1	1.71
miR7694-3p	2	5	1	2	1	1	3	3	5	2.13
miR168a-3p	3	7	6	3	3	3	4	5	2	3.72
miR1876	4	1	3	4	4	5	5	6	7	3.88
miR1882e-3p	5	2	2	5	6	4	7	7	8	4.61
miR535-3p	8	8	7	8	8	7	2	2	4	5.28
miR1849	7	6	8	7	5	8	6	1	6	5.32
miR1868	6	3	5	6	7	6	8	8	3	5.47

Three sample grouping were used: *All* all samples; *Infection* infected + control samples; *BABA* BABA treated + control samples

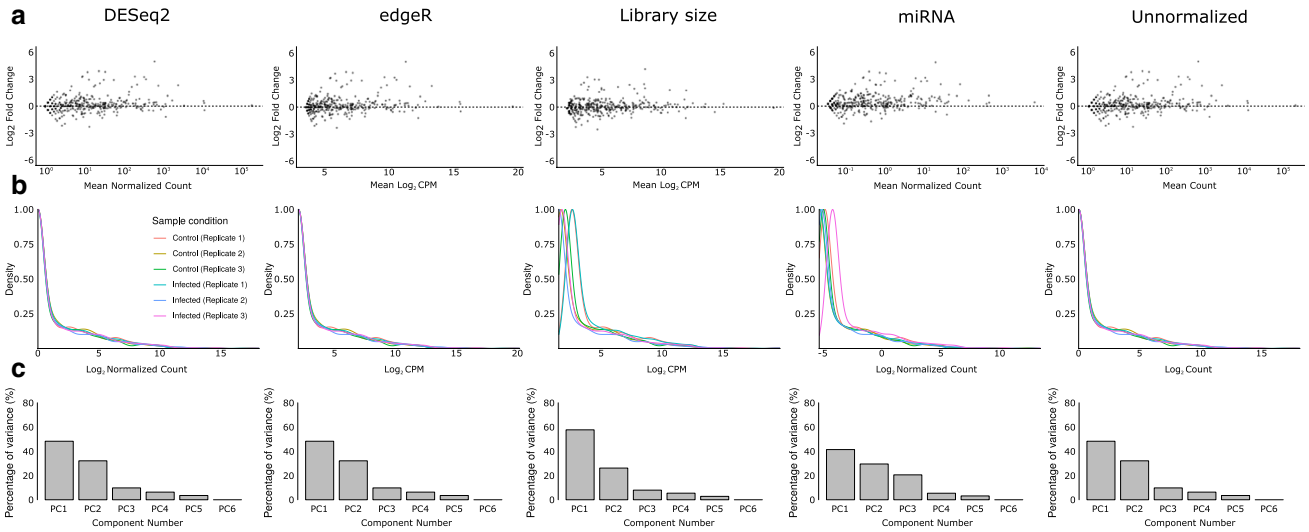


Fig. 3 Normalization of count data using default or miRNA normalization with either DESeq2 or edgeR. **a** MA plots of DESeq2 show log2fold change in function of the mean of normalized counts over all samples. MA plots of edgeR show log2 fold change in function of average logCPM (average of logarithmic counts per million over all samples). **b** Density plots of DESeq2 show density in function of the logarithmic transformed normalized counts. Density plots of edgeR show density in function of log2CPM (logarithmic counts per million). **c** Scree plots

would be advisable to use caution when considering using miR390-5p as a reference gene in experimental conditions/organisms other than those used in this work (Ding et al. 2016; Lu et al. 2018). On the other hand, to the best of our knowledge, miR7694-3p has not been shown to be differentially expressed under either biotic or abiotic conditions, and hence could be used in other studies. The combination of miR7694-3p, miR1868 and miR1849 was also tested and validated for its suitability as a reference gene set under nematode infection and viral infection. However, this does not guarantee their suitability as reference miRNAs in other conditions such as fungal or bacterial infection, as these conditions may have a differing effect on the behavior of miRNAs, and hence experimental confirmation will still be needed under other stress conditions. Nevertheless, since none of these miRNAs have been described thus far as being differentially expressed under abiotic or biotic stress conditions, they have great potential to be used as candidate reference genes in experimental conditions beyond those described in this work.

In conclusion, miR390-5p and miR7694-3p have been found to be suitable miRNA reference genes for rice, an important staple crop and model organism, under either infection with the root-knot nematode *Meloidogyne graminicola* or treatment with BABA. Furthermore, we have demonstrated the potential of miR7694-3p, miR1868 and miR1849 to be used as candidate reference miRNAs in experimental conditions beyond those described in this work.

Author contribution statement BV designed the experiment, performed the data analysis and drafted the manuscript. LDS performed the wet-lab activities. TK and TDM read and corrected the draft. All authors read and approved the manuscript.

Acknowledgements This work was funded by Fonds Wetenschappelijk Onderzoek-Vlaanderen (Project: G007417N).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Andersen CL, Jensen JL, Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64VN-r:5245–5250. https://doi.org/10.1158/0008-5472.can-04-0496

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bridge J, Plowright RA, Peng D (2005) Nematode parasites of rice. In: Sikora R, Coyne D, Hallmann J, Timper P (eds) Plant parasitic nematodes in subtropical and tropical agriculture. CABI, Wallingford, pp 87–130

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

Buswell W, Schwarzenbacher RE, Luna E et al (2018) Chemical priming of immunity without costs to plant growth. New Phytol 218:1205–1216. https://doi.org/10.1111/nph.15062

Chugh P, Dittmer DP (2012) Potential pitfalls in microRNA profiling. Wiley Interdiscip Rev RNA 3:601–616. https://doi.org/10.1002/wrna.1120

Ding Y, Ye Y, Jiang Z et al (2016) MicroRNA390 is involved in cadmium tolerance and accumulation in rice. Front Plant Sci 7:235. https://doi.org/10.3389/fpls.2016.00235

Djami-Tchatchou AT, Sanan-Mishra N, Ntushelo K, Dubery IA (2017) Functional roles of microRNAs in agronomically important plants—potential as targets for crop improvement and protection. Front Plant Sci. https://doi.org/10.3389/fpls.2017.00378

Dobin A (2013) STAR for miRNA. https://groups.google.com/forum/#!topic/rna-star/RBWvAGFooMU. Accessed 14 May 2019

FAO (2017) Food and Agriculture Organization of the United Nations. http://www.fao.org/home/en. Accessed 28 May 2019

Fausto AKS, da Silva TF, Romanel E, Vaslin MFS (2017) MicroRNAs as reference genes for quantitative PCR in cotton. Public Libr Sci one 12:1–21. https://doi.org/10.1371/journal.pone.0174722

Gene Quantification Platform (2019) Real Time PCR amended, A useful new approach? Statistical problems? https://www.gene-quantification.de/avery-rel-pcr-errors.pdf. Accessed 12 Jun 2019

Gheysen G, Mitchum MG (2011) How nematodes manipulate plant development pathways for infection. Curr Opin Plant Biol 14:415–421. https://doi.org/10.1016/J.PBI.2011.03.012

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

Hellemans J, Mortier G, De Paepe A et al (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8:R19. https://doi.org/10.1186/gb-2007-8-2-r19

Ji H, Kyndt T, He W et al (2015) β-Aminobutyric acid-induced resistance against root-knot nematodes in rice is based on increased basal defense. Mol Plant-Microbe Interact 28:519–533. https://doi.org/10.1094/MPMI-09-14-0260-R

Jisha KC, Puthur JT (2016) Seed priming with BABA (β-amino butyric acid): a cost-effective method of abiotic stress tolerance in *Vigna radiata* (L.) Wilczek. Protoplasma 253:277–289. https://doi.org/10.1007/s00709-015-0804-7

Kozera B, Rapacz M (2013) Reference genes in real-time PCR. J Appl Genet 54:391–406. https://doi.org/10.1007/s13353-013-0173-x

Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. Nucleic Acids Res 47:D155–D162. https://doi.org/10.1093/nar/gky1141

Lawrence M, Huber W, Pagès H et al (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:1–21. https://doi.org/10.1186/s13059-014-0550-8

Lu Y, Feng Z, Liu X et al (2018) MiR393 and miR390 synergistically regulate lateral root growth in rice under different conditions. BMC Plant Biol 18:261. https://doi.org/10.1186/s12870-018-1488-x

Mestdagh P, Van Vlierberghe P, De Weer A et al (2009) A novel and universal method for microRNA RT-qPCR data normalization. Genome Biol. https://doi.org/10.1186/gb-2009-10-6-r64

Millar AA, Waterhouse PM (2005) Plant and animal microRNAs: similarities and differences. Funct Integr Genomics 5:129–135. https://doi.org/10.1007/s10142-005-0145-2

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

Pritchard CC, Cheng HH, Tewari M (2012) MicroRNA profiling: approaches and considerations. Nat Rev Genet 13:358–369. https://doi.org/10.1038/nrg3198

Reversat G, Boyer J, Sannier C, Pando-Bahuon A (1999) Use of a mixture of sand and water-absorbent synthetic polymer as substrate for the xenic culturing of plant-parasitic nematodes in the laboratory. Nematology 1:209–212. https://doi.org/10.1163/156854199508027

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25. https://doi.org/10.1186/gb-2010-11-3-r25

Schmittgen TD, Zakrajsek BA (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods 46:69–81. https://doi.org/10.1016/S0165-022X(00)00129-9

Selvey S, Thompson EW, Matthaei K et al (2001) β-Actin—an unsuitable internal control for RT-PCR. Mol Cell Probes 15:307–311. https://doi.org/10.1006/MCPR.2001.0376

Slaughter A, Daniel X, Flors V et al (2012) Descendants of primed *Arabidopsis* plants exhibit resistance to biotic stress. Plant Physiol 158:835–843. https://doi.org/10.1104/pp.111.191593

Vandesompele J, De Preter K, Pattyn I et al (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:34. https://doi.org/10.1186/gb-2002-3-7-research0034

Varkonyi-Gasic E, Wu R, Wood M et al (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3:1–12. https://doi.org/10.1186/1746-4811-3-12

Wang ST, Sun XL, Hoshino Y et al (2014) *MicroRNA319* positively regulates cold tolerance by targeting *OsPCF6* and *OsTCP21* in rice (*Oryza sativa* L.). PLoS One 9:1–12. https://doi.org/10.1371/journal.pone.0091357

Yang C, Li D, Mao D et al (2013) Overexpression of microRNA319 impacts leaf morphogenesis and leads to enhanced cold tolerance in rice (*Oryza sativa* L.). Plant Cell Environ 36:2207–2218. https://doi.org/10.1111/pce.12130

Yang J, Zhang F, Li J et al (2016) Integrative analysis of the microRNAome and transcriptome illuminates the response of susceptible rice plants to rice stripe virus. PLoS One 11:e0146946. https://doi.org/10.1371/journal.pone.0146946

Zaleski C, Jha S, Gingeras TR et al (2012) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Zhang Y-C, Yu Y, Wang C-Y et al (2013) Overexpression of microRNA OsmiR397 improves rice yield by increasing grain size and promoting panicle branching. Nat Biotechnol 31:848–852. https://doi.org/10.1038/nbt.2646