scientific reports

Check for updates

OPEN

Transcriptome-guided selection of stable reference genes for expression analysis in spinach

Aboozar Soorni✉, Maryam Rezvani & Haniye Bigdeli

Accurate measurement of gene expression levels is vital for advancing plant biology research. This study explores the identification and validation of stable reference genes (RGs) for gene expression analysis in *Spinacia oleracea*. Leveraging transcriptome data from various developmental stages, we employed rigorous statistical analyses to identify potential RGs. A total of 1196 candidate genes were initially screened based on expression variability, with subsequent refinement using criteria such as low variance and stability. Among 12 commonly used candidate RGs, EF1α and H3 emerged as the most stable across diverse experimental conditions, while GRP and PPR exhibited lower stability. These findings were further validated through qRT-PCR assays and comprehensive statistical analyses, including geNorm, NormFinder, BestKeeper, and RefFinder. Our study underscores the importance of systematic RG selection to ensure accurate normalization in gene expression studies, particularly in the context of *S. oleracea* developmental stages and physiological processes like flowering. These validated RGs provide a robust foundation for future gene expression analysis in *S. oleracea* and contribute to the advancement of molecular research in plant biology.

Keywords Spinach, Reference genes, Gene expression, RNA-Seq, qRT-PCR

As the depth of plant biology studies increases, the importance of accurately measuring gene expression levels becomes more significant. Consequently, quantitative real-time polymerase chain reaction (qRT-PCR) has become a key analytical method for researchers in this field. This technique allows for the sensitive and specific measurement of gene expression, making it an essential tool in the study of *S. oleracea* (spinach) and other model organisms[1–3]. Despite the numerous advantages offered by qRT-PCR, including its high sensitivity and specificity, good reproducibility[4–6] the reliability of its results can be influenced by various factors. These factors encompass variables such as RNA concentration, reverse transcription efficiency, primer specificity, the specific experimental conditions employed during the analysis, and internal (RGs[7–9]. Among these factors, internal RGs are essential elements in qRT-PCR analysis, providing a stable foundation within the experimental setup. Their primary function is to establish a consistent baseline for measuring the expression levels of target genes. Indeed, it becomes crucial to carefully select appropriate internal RGs in order to ensure accurate data correction and standardization in qRT-PCR analysis. Through the normalization of variations caused by these factors, internal RGs enhance the reliability and robustness of analytical results. The careful selection and validation of internal RGs are fundamental requirements for ensuring the accuracy and reproducibility of qRT-PCR data analysis[10–14]. An ideal reference gene in plant biology should demonstrate consistent expression levels under various experimental conditions, irrespective of external factors, sample types, or treatment methods. It should remain unaffected by fluctuations caused by environmental cues, developmental stages, or stress responses. This stability is crucial as it guarantees that the reference gene serves as a dependable baseline for accurately evaluating the expression levels of target genes[15–19]. Recent investigations have revealed that the expression profiles of RGs may vary in different experimental contexts. This suggests that the stability of housekeeping genes depends on the specific experimental conditions and shows consistency within a limited range of cellular contexts. Therefore, it is crucial to carefully select appropriate RGs when using quantitative polymerase chain reaction (qPCR) techniques to assess gene expression dynamics in plants under well-defined experimental scenarios[20,21]. Traditionally, genes such as glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*;[22–25]), β-actin (*ACT*);[23,26–28], 18 S ribosomal RNA (*18 S* rRNA);[29,30], ubiquitin (*UBQ*);[31], tubulin α and β (*TUB*);[32], and elongation factor 1-α (*EF1A*);[33,34] have been recognized for their housekeeping functions in cellular processes and are commonly used as RGs to normalize qPCR data. However, recent studies have indicated that the stability of transcripts associated with these conventional RGs can vary under different experimental conditions[30,35,36], prompting a

Department of Biotechnology, College of Agriculture, Isfahan University of Technology, Isfahan, Iran. ✉email: soorni@iut.ac.ir

growing interest in the identification of stable reference genes. In order to select suitable RGs and assess their expression stability, a comprehensive approach is required, which involves the use of multiple methodologies and software tools. One promising approach is to leverage the abundance of transcriptome data to identify potential RG candidates[24,37,38]. Additionally, several statistical algorithms have been developed to identify RGs that maintain stable expression levels under specific experimental conditions. These algorithms include geNorm[39], NormFinder[40], BestKeeper[14], the ΔCt method[41], and RefFinder[42]. In previous research endeavors, these approaches and algorithms have been extensively employed in studying various plant species under different conditions to identify stable RGs. For example, investigations have been conducted to assess the expression stability of candidate RGs in plants like *Arabidopsis thaliana*[43], *Oryza sativa*[44], *Solanum lycopersicum*[45], and *Zea mays*[46], which were exposed to diverse environmental stresses including drought, salinity, heat, and pathogen infection. Additionally, spinach has emerged as a valuable model organism in plant biology studies due to its genetic manipulability, compact genome, and rapid growth rate. This highlights the significance of identifying suitable RGs for spinach research under different conditions[47–50]. A previous publication has highlighted the necessity for specific investigations regarding the selection of optimal RGs for spinach, particularly concerning stress response[30]. In this previous study, various housekeeping genes from spinach were evaluated for their stability under different abiotic stress conditions. Ten candidate genes, including 18 S ribosomal RNA (*18 S* rRNA), actin, ADP ribosylation factor (*ARF*), cytochrome c oxidase subunit 5 C (*COX*), cyclophilin (*CYP*), elongation factor 1-alpha (*EF1α*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), histone H3 (*H3*), 50 S ribosomal protein L2 (*RPL2*), and tubulin alpha chain (*TUBα*), were analyzed under various stress responses by qRT-PCR and NormFinder, BestKeeper, and geNorm algorithms[30]. The findings of this research identified *18 S* rRNA, *actin*, *ARF*, *COX*, *CYP*, *EF1α*, *GAPDH*, *H3*, and *RPL2* as optimal RGs for gene expression analysis of different organs and stress responses[30].

Given the intricate nature of gene expression regulation and the ever-changing landscape of plant development, it is crucial to enhance the pool of RGs that can effectively standardize gene expression data under diverse experimental conditions. This is especially important when assessing characteristics during developmental phases and critical physiological processes like flowering and bolting time in spinach. Furthermore, as transcriptome data play a significant role in identifying the most appropriate candidate genes for assessment and standardization in spinach research, additional studies should concentrate on systematically appraising the expression stability of candidate RGs pinpointed from transcriptome data in various spinach tissues at distinct developmental stages. This includes stages such as vegetative growth, reproductive growth, and senescence. It is also essential to utilize robust statistical methodologies like geNorm, NormFinder, and BestKeeper for accurate evaluation. Hence, in order to fill the existing research gap concerning stable RGs for assessing gene expression during developmental stages and important traits such as flowering/bolting time in spinach, we undertook an extensive investigation. By utilizing transcriptome data as our main source and employing advanced statistical algorithms and software tools for systematic analysis, we meticulously assessed the expression stability of multiple potential RGs. Subsequently, we chose twelve genes for qRT-PCR validation.

Results

Refining candidate RGs via RNA-Seq analysis

Our endeavor to identify RGs in spinach involved a meticulous examination of diverse transcriptome datasets. The datasets analyzed included four bioprojects, each containing a specific set of libraries focusing on different aspects of spinach biology and development. These included tissue-specific expression profiling in leaves, petioles, and roots, investigation of gene expression across various developmental stages of female flowers, and targeting genes related to bolting and flowering during the transition from vegetative to reproductive stages, as well as samples obtained from female sepal's post-pollination.

To identify potential reference genes, we employed the CV method, which allowed us to systematically screen for genes based on their expression variability. This method provided a simple yet effective way to compare and rank genes. We excluded genes with low expression levels and set a minimum mean log2(TPM) cut-off of 5 to ensure the selection of robust reference genes. Additionally, we required a SD log2(TPM) value of less than 1 to ensure low variance. We further refined the identification of RGs by applying a 0.2 CV cut-off. Through this rigorous methodology, we successfully identified a total of 1196 genes out of 25,496 (Table S1). These genes exhibited log2(TPM) values ranging from 5.02 to 13.37 and CV values ranging from 0.026 to 1.999. This comprehensive analysis provides valuable insights into the dynamics of gene expression in spinach across various developmental stages.

Gene classification and functional enrichment

The gene classification analysis (Fig. 1) uncovered a wide range of biological processes in spinach, with significant representation in multiple key categories. Metabolic processes were particularly prominent, with 163 genes involved in various metabolic activities that are crucial for the functioning of cells. Biosynthetic processes were also highly represented, with 192 genes engaged in the synthesis of essential cellular components and molecules that are vital for growth and development. Additionally, cellular processes emerged as the most major category, with a remarkable 224 genes, indicating the roles of these genes in fundamental cellular activities. Other noteworthy categories included nucleobase-containing compound metabolic processes (155 genes), carbohydrate metabolic processes (55 genes), and translation (83 genes). Furthermore, stress response mechanisms were noteworthy, with 87 genes, along with transport processes (124 genes) and developmental pathways such as multicellular organism development (26 genes), embryo development (11 genes), and flower development (12 genes).

The results of GO enrichment analysis (Fig. 2) revealed significant associations between the gene identified and various biological processes. Notably, the gene exhibited enrichment in several key cellular processes

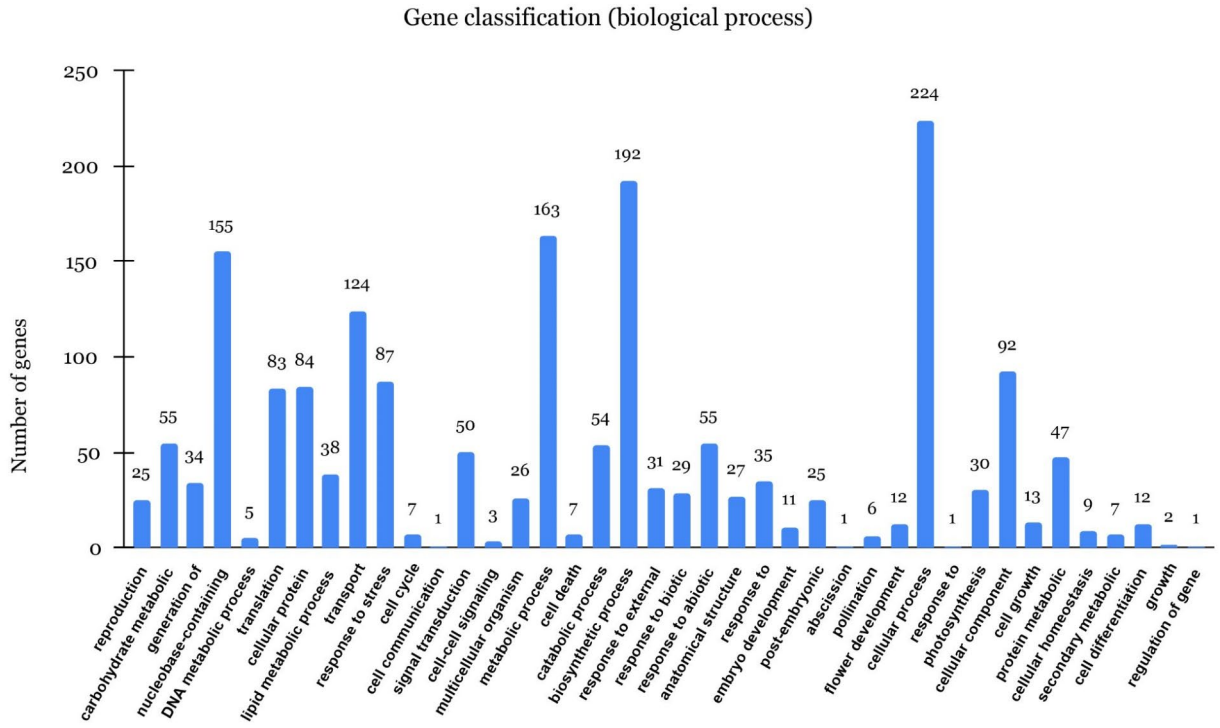

Fig. 1. Gene classification (biological process) for 1196 identified potential RGs.

essential for protein synthesis and cellular component formation. Processes such as ribosome biogenesis, ribonucleoprotein complex biogenesis, and cellular component biogenesis were highly enriched, indicating the gene's involvement in fundamental cellular structural and functional activities. Moreover, enrichment in translation and peptide biosynthetic processes underscored its crucial role in protein synthesis. Additionally, the gene showed enrichment in metabolic processes related to peptide and amide biosynthesis, highlighting its involvement in diverse biochemical pathways. Furthermore, associations with organonitrogen compound biosynthesis and cellular nitrogen compound biosynthesis suggested its participation in nitrogen metabolism, an essential aspect of cellular physiology.

Gene selection and expression analysis of candidate reference gene

Based on the results obtained from the RNA-seq data analysis, particularly focusing on the SD and CV of genes, along with an extensive literature review to ascertain gene function, a selection of genes was curated and presented in the Table 1. The SD and CV values were used as metrics to gauge the stability and consistency of gene expression across samples, with a preference for genes demonstrating lower variability. In the selection process, genes exhibiting SD values within a specific range (e.g., below 1.0) and CV values indicating relatively stable expression (e.g., below 0.1) were included for qRT-PCR assay. Typically, genes with SD values below this threshold and CV values within this range were prioritized, as they are more likely to represent reliable reference or functional genes. However, genes with excessively high SD and CV values, indicating greater variability and inconsistency in gene expression, were excluded from the list. Notably, two genes, namely *GAPDH* and *Actin*, were not in the list of identified genes. This was attributed to their significantly higher SD and CV values, which exceeded the specified thresholds.

In order to evaluate the stability of gene expression across different developmental stages, we examined the quantification cycle (Cq) values derived from our experimental data (Fig. 3). Ct values are utilized as measures of gene expression levels, where lower Cq values correspond to higher expression levels. Our analysis unveiled distinct Cq values for each gene at various developmental stages, offering insights into the expression dynamics throughout development. Among the genes investigated, we noted diverse expression patterns across the developmental stages. As depicted in the results (Fig. 3), the average Cq values of 12 RGs ranged from 18 to 32, with the majority falling between 22 and 27. Across all samples, *EF1α* and *ACT* exhibited the lowest and highest Cq values, with Cq ranges of 18.2–20.8 and 27.2–32.0, respectively. These results suggest that EF1α displays the highest expression levels, indicated by its lower Cq values, while *ACT* shows lower expression levels with higher Cq values. Additionally, genes like *GRP*, *PPR*, and *RPL27* demonstrated more variable Cq values, signifying fluctuations in expression levels across different developmental stages. Among the genes studied, including *EIF*, *GAPDH*, *UBQ*, and *EF1α*, there was a noticeable trend of displaying a narrower range of variation in their expression levels across different conditions. These genes consistently exhibited relatively stable expression

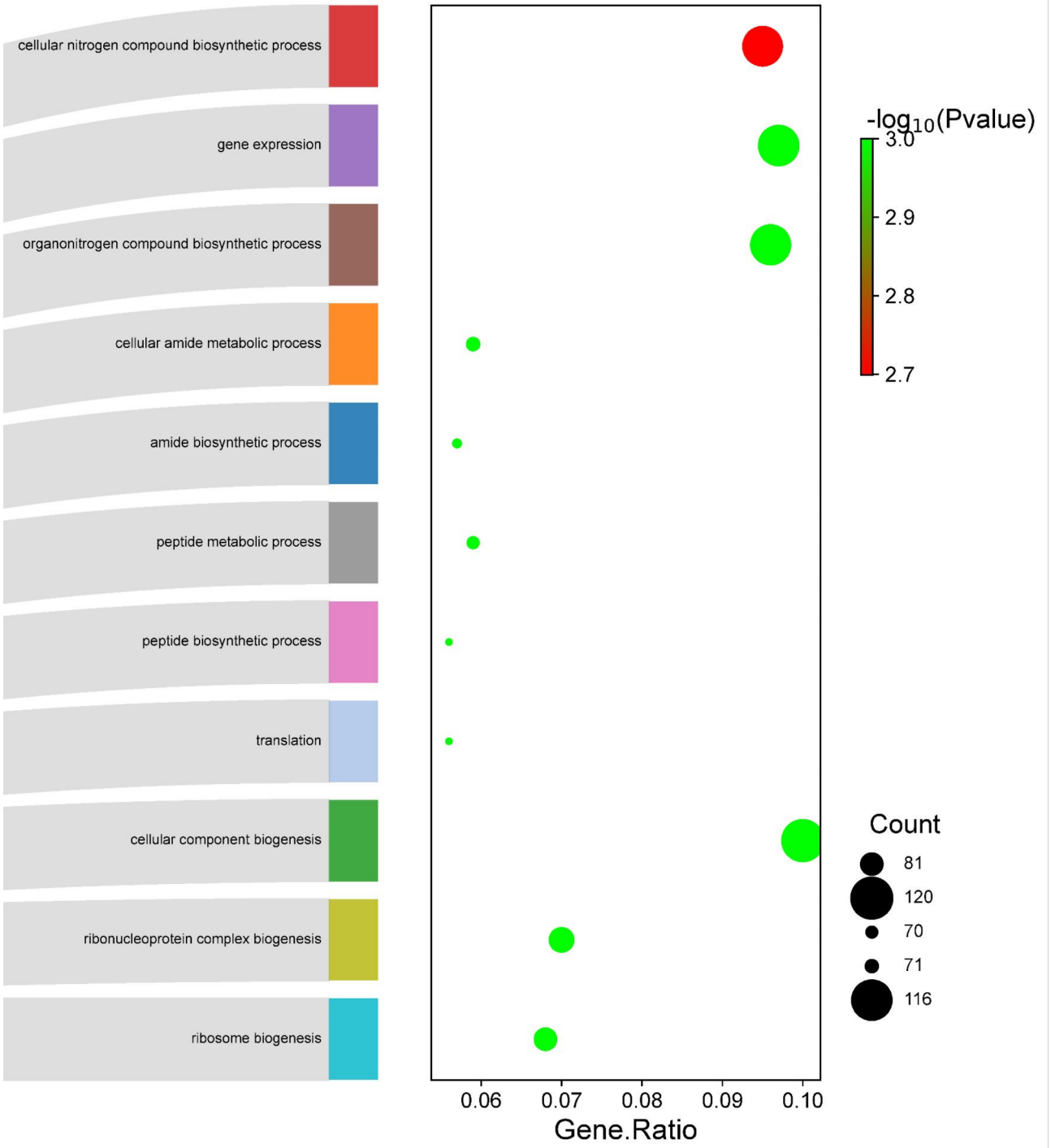

Fig. 2. GO enrichment (BP: biological process) of 1196 identified potential RGs.

patterns, with minimal fluctuations in their Cq values across various experimental conditions. In conclusion, our analysis indicates a high level of expression stability for *EF1α* and *H3*, making them potential candidates for internal RGs in gene expression studies. Their consistent expression profiles suggest their suitability for normalization purposes, facilitating accurate and reliable quantification of gene expression levels.

Expression stability of candidate genes

To provide additional evidence of the stability of the RGs, all samples underwent testing using four algorithms, namely BestKeeper, geNorm, and NormFinder. The purpose of these tests was to evaluate the stability of the

Gene	CV	SD	*r*	*p*-Value
TUBα	4.65	1.12	0.125	0.816
GRP	3.44	0.85	0.356	0.34
PPR	3.58	0.85	0.395	0.21
EIF	1.25	0.38	0.425	0.046
ACT	2.35	0.68	0.624	0.03
EF1b	4.77	1.18	0.688	0.002
GAPDH	5.93	1.11	0.721	0.002
RPL27	9.52	1.55	0.758	0.001
ARF	4.83	1.36	0.856	0.001
UBQ	2.76	0.48	0.872	0.001
EF1α	7.83	1.89	0.943	0.001
H3	6.21	1.59	0.995	0.001

Table 1. Expression stability values of 12 candidate RGs calculated by BestKeeper.

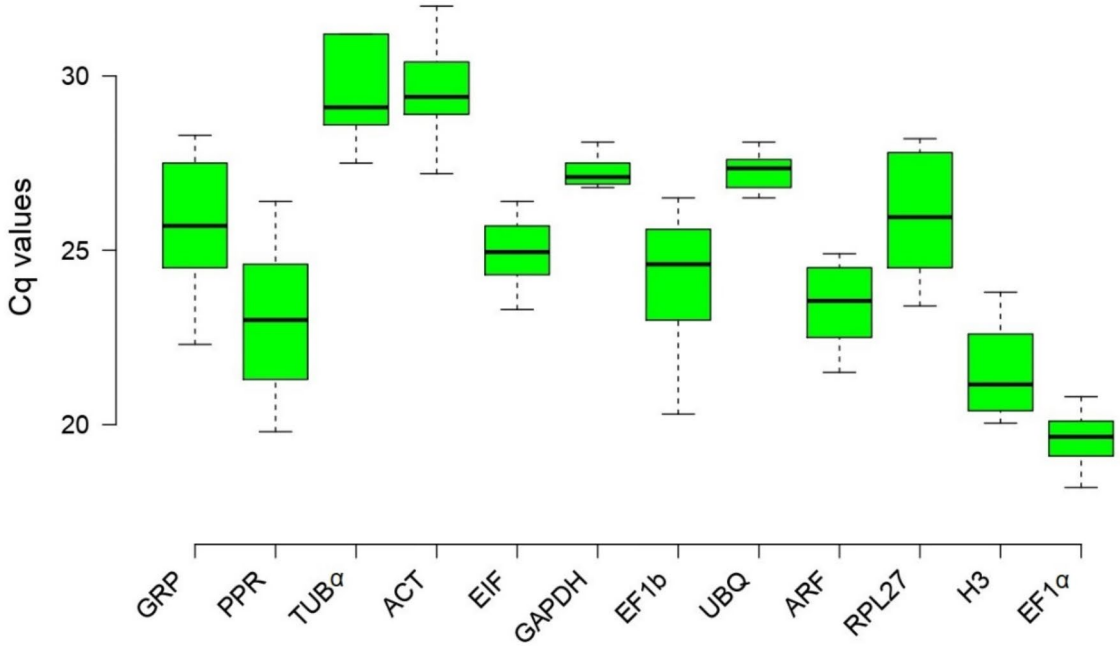

Fig. 3. Distribution of Cq values for 12 candidate RGs across all samples using qRT-PCR data.

potential RGs. Furthermore, the RefFinder tool was employed to comprehensively rank the expression stability of all potential RGs.

geNorm analysis

The expression stability of various genes was evaluated using geNorm analysis (Fig. 4), which involved assessing their respective M values. A lower M value indicates a higher level of stability in gene expression. After arranging the M values in descending order, it was observed that all genes had M values below 1.0, indicating their stability across a range of experimental conditions. Notably, among the genes analyzed, *EF1α* exhibited the highest stability with the lowest M value of 0.24, closely followed by *H3* with an M value of 0.32. In contrast, *GRP* had

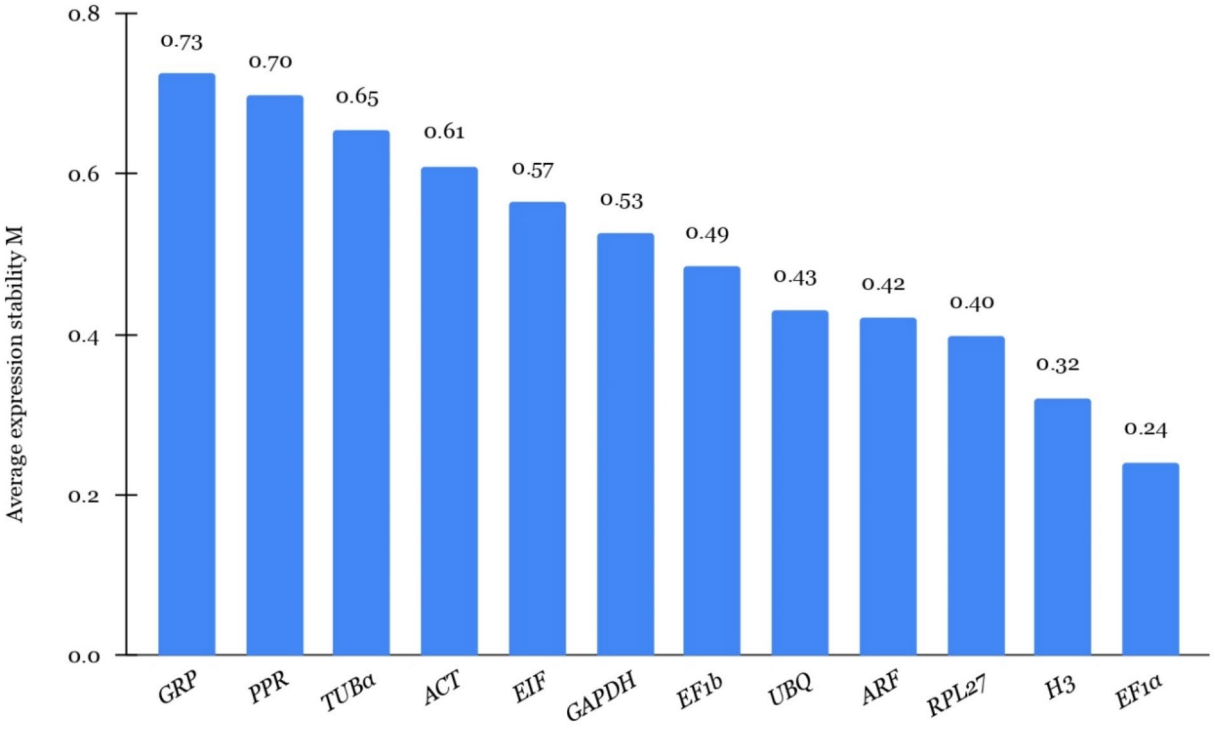

Fig. 4. geNorm analysis of the stability values (M) of 12 candidate RGs under various developmental stages.

the highest M value of 0.72, suggesting relatively lower stability compared to the other genes analyzed. These findings emphasize the significance of *EF1α* and *H3* as reliable RGs for normalization in gene expression studies conducted under different experimental settings, while also highlighting the potential variability associated with *GRP* expression.

NormFinder analysis

The expression stability of various candidate reference genes, which is crucial for reliable normalization in gene expression studies, was determined using the NormFinder analysis (Fig. 5). The obtained expression stability values from NormFinder provide valuable insights into the suitability of these genes as reference standards. Lower stability values indicate greater stability in gene expression across different experimental conditions. Among the genes examined, *EF1α* exhibited the highest stability with an expression stability value of 0.21, indicating consistent and robust expression levels. Similarly, genes such as *RPL27* (0.35) and *H3* (0.3) also showed relatively low stability values, suggesting their potential suitability as RGs in gene expression analyses. On the other hand, *GRP* displayed the highest stability value of 0.68, indicating greater variability in gene expression under the assessed experimental conditions. Comparing these results with those obtained from geNorm analysis, there is a notable consistency in identifying *EF1α* as one of the most stable genes across both methodologies. However, there are discrepancies in the rankings of other genes. This highlights the importance of using multiple algorithms to comprehensively assess the stability of candidate RGs.

BestKeeper analysis

The stability of candidate RGs was assessed using the BestKeeper software, which considered the correlation coefficient (r), coefficient of variation (CV) and the Standard Deviation (SD), and associated p-values. The Table 1 provided outlines the results of this analysis. Among the genes examined, *H3* displayed an exceptional stability in its expression, as indicated by its high correlation coefficient of 0.995. This strong correlation suggests a consistent expression pattern, which is further supported by its relatively low SD value of 1.59. Similarly, *EF1α* demonstrated a high correlation coefficient of 0.943, highlighting its stability in expression. Although *EF1α* had a slightly higher CV (7.83) and SD value ± (1.89) compared to *H3*, it still maintained a robust correlation, indicating reliable expression levels. Other genes, such as *UBQ* and *ARF*, also exhibited notable stability, with correlation coefficients exceeding 0.85 and relatively low CV and SD values, further confirming their suitability as reference genes. In contrast, genes like *TUBα* and *GRP* showed lower correlation coefficients below 0.4, indicating less stable expression profiles. This is supported by their higher CV and SD values, suggesting greater variability in expression across the samples.

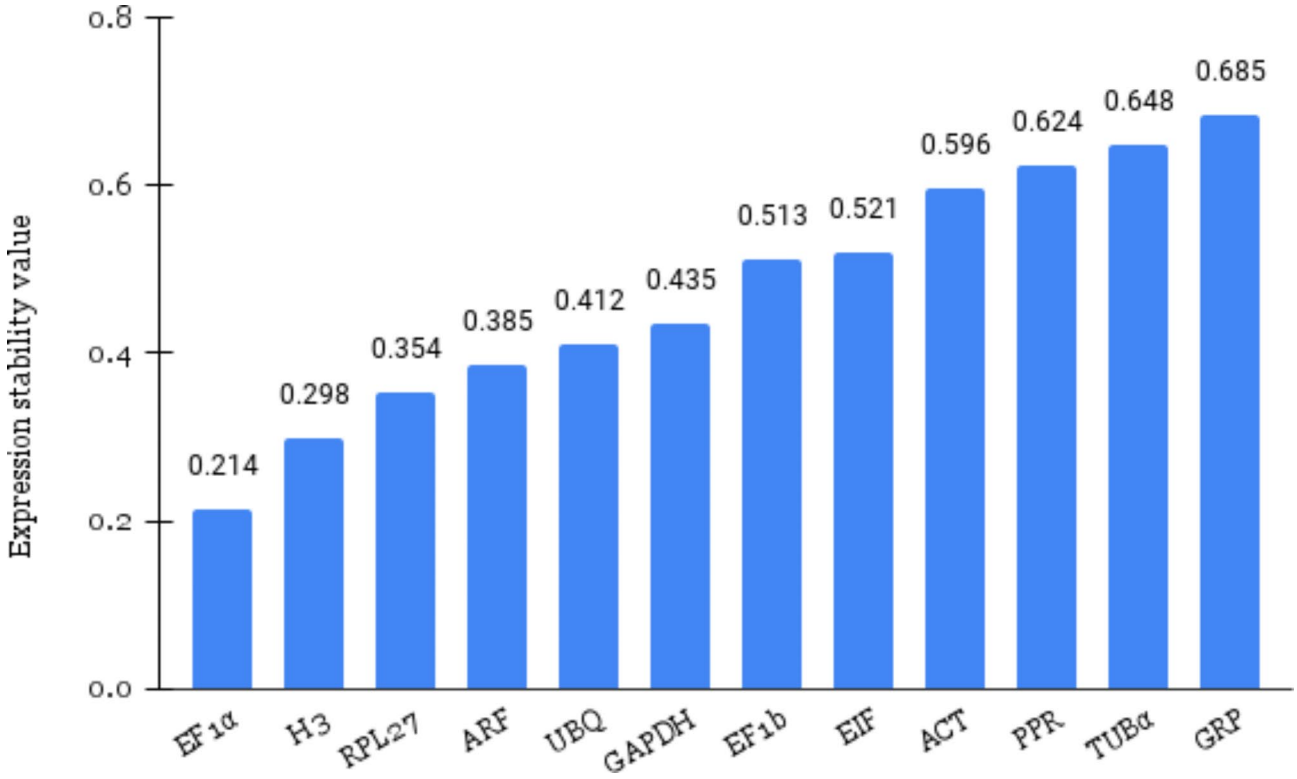

Fig. 5. Expression stability values of 12 candidate RGs calculated by NormFinder.

Rank	geNorm	NormFinder	BestKeeper	RefFinder	SV
1	*EF1α*	*EF1α*	*H3*	*EF1α*	1.36
2	*H3*	*H3*	*EF1α*	*H3*	3.42
3	*RPL27*	*RPL27*	*UBQ*	*UBQ*	3.45
4	*ARF*	*ARF*	*ARF*	*RPL27*	3.93
5	*UBQ*	*UBQ*	*RPL27*	*ARF*	4.49
6	*EF1b*	*GAPDH*	*GAPDH*	*EF1b*	5.77
7	*GAPDH*	*EF1b*	*EF1b*	*EIF*	7.34
8	*EIF*	*EIF*	*ACT*	*GAPDH*	8.61
9	*ACT*	*ACT*	*EIF*	*ACT*	8.65
10	*TUBα*	*PPR*	*PPR*	*TUBα*	9.47
11	*PPR*	*TUBα*	*GRP*	*PPR*	9.95
12	*GRP*	*GRP*	*TUBα*	*GRP*	10.24

Table 2. Ranking of expression stability for the 12 candidate RGs using RefFinder.

RefFinder analysis

The stability rankings of 12 candidates internal RGs were consolidated through the RefFinder analysis, which utilized geNorm, NormFinder, and BestKeeper. The Stability Value (SV) provided an overall measure of the stability of these genes. Among the candidates, *EF1α* was identified as the most stable gene with the lowest SV (Table 2), consistently ranking highly in all analyses. *H3* and *UBQ* also exhibited robust stability, maintaining their top positions. On the other hand, *GRP* and *PPR* showed higher SV values, indicating lower stability across the various statistical methods. These results highlight *EF1α* as a reliable internal RG, while caution is advised when considering *GRP* and *PPR* for normalization in gene expression studies.

Validation of candidate stable RGs

Based on our analysis, we identified *EF1α* and *H3* as the most stable RGs and *GRP* as one of the most unstable RGs, which were subsequently validated through an examination of the expression profiles of *FKF1* and *LHY* genes to assess their reliability (Fig. 6). Based on the slope of the standard curve, all gene primers exhibited efficient PCR amplification. While the R^2 values were slightly below the optimal threshold of 0.99, with all primers showing values greater than 0.97, this level of linearity is still considered robust. Given the inherent biological variability

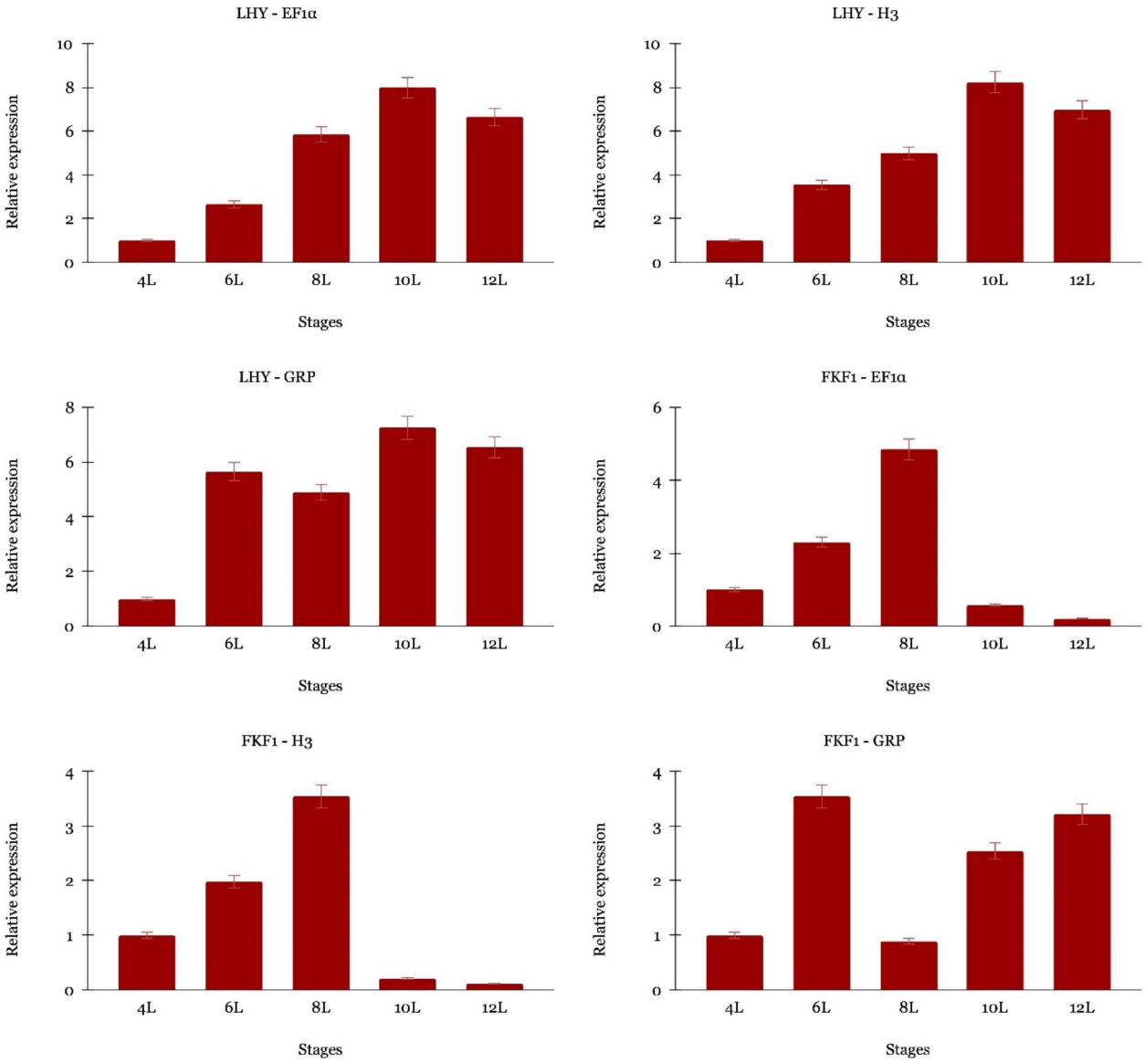

Fig. 6. Relative expression levels of *LHY* and *FKF1* across all stages normalized by the most stable genes and the most unstable gene. 4 L, 6 L, 8 L, 10 L, and 12 L represent the four, six, eight, ten and twelve leaf stages.

of our samples, we regard these values as acceptable, reflecting reliable quantification performance.The results from our experiment revealed distinct expression patterns of *FKF1* and *LHY* across various developmental stages. The results depicted in Fig. 6 indicate that the expression levels of *FKF1* and *LHY*, when standardized with *EF1α* and *H3*, demonstrated distinct patterns across diverse developmental stages according to their function. *LHY* exhibited its lowest expression during the vegetative stage, with notable upregulation at the 8-leaf stage, indicating the transition into the reproductive phase and flowering initiation, consistent with its role in the circadian regulation of flowering. *FKF1* displayed relatively stable expression during the vegetative stage but showed a marked increase at the 8-leaf stage, suggesting involvement in the early flowering phase.

Discussion

Gene expression analysis plays a crucial role in molecular biology research. To achieve this, different PCR techniques such as semiquantitative PCR, RT-qPCR, and digital PCR (dPCR) are commonly utilized. Among these techniques, dPCR stands out as it allows for the detection of absolute gene expression levels without the need for RGs[51]. However, its labor-intensive nature and limited throughput have hindered its widespread adoption in plant molecular labs. On the other hand, qRT-PCR remains a convenient and accurate method for mRNA quantitation. The key to obtaining precise quantitative results lies in the careful selection of suitable internal RGs for a specific species under specific experimental conditions. Consistent with prior research, findings indicate that there are several RGs that exhibit stable expression within specific sample sets in spinach. However, it is

important to note that no single gene remains stable across all conditions. Previous investigations in spinach[30] have utilized genes such as *18 S* rRNA, actin, and *GAPDH* as RGs, but their stability has not been thoroughly validated for gene evaluation pertaining to developmental stage and flowering/bolting time. Consequently, the objective of this study was to identify novel and stable RGs for gene expression studies in spinach using transcriptome data and qRT-PCR validation. Utilizing transcriptome data with stringent criteria for the identification of RGs offers several advantages over traditional methods of RG selection. Traditional approaches often rely on the assumption that commonly used housekeeping genes exhibit stable expression across all experimental conditions. However, this assumption may not always hold true, leading to inaccurate normalization of gene expression data. In contrast, employing transcriptome data allows for a more comprehensive assessment of gene expression stability across various experimental conditions. By applying hard criteria, such as low variance among samples and ubiquitous expression, transcriptome-based RG identification ensures the selection of genes that are truly stable and suitable for normalization[24]. This approach has been successfully employed in several studies across different organisms and experimental setups[45,52–54]. The successful identification of a total of 1196 potential RGs marks a significant achievement in our study, as it provides a robust foundation for future gene expression analysis in spinach. In our study, we conducted a thorough screening of 12 commonly used candidate RGs in spinach, leveraging transcriptome data for initial selection. Given the variations in gene expression levels across different organs and developmental stages, we employed multiple statistical tools to assess the stability of these candidate RGs under diverse circumstances. Combining commonly algorithms, including geNorm, BestKeeper, and NormFinder, allowed for a comprehensive evaluation of RG stability and enhanced the reliability of experimental data[14,39,40]. This approach aimed to mitigate the potential biases introduced by individual analysis programs, as highlighted by previous studies showing discrepancies in results among different algorithms. Our findings demonstrated varying rankings of RGs across different analytical methods, consistent with observations in other species, suggesting the influence of algorithmic principles and screening focus on analysis outcomes[31,32,54–58]. Notably, the online software RefFinder was employed to rank each candidate RG comprehensively, highlighting the importance of considering multiple algorithms for accurate RG selection[30]. Among the candidate RGs, *EF1α* and *H3* consistently demonstrated superior stability across various conditions, while *GRP*, *PPR*, and *TUBa* consistently ranked lower in stability assessments. These findings are consistent with some previous studies that have identified *EF1α* and *H3* as reliable internal RGs for standardization in various plant species, further highlighting the importance of rigorous RG selection for gene expression studies[53,59–61]. Previous research on spinach gene expression analysis has assessed *EF1α* as a potential RG. In certain algorithmic analyses, such as NormFinder, *EF1α* has been identified as one of the most stable genes across some experimental conditions such as organs/seedlings and heat stress. However, it's worth noting that in some specific experimental conditions, the stability of *EF1α* and *H3* was not consistent across all analyses[30]. The variability in the stability of *EF1α* and *H3* in different researches and different experimental conditions can be attributed to several factors. Firstly, gene expression levels may fluctuate due to biological variations inherent to specific experimental setups, such as variations in developmental stages, tissue types, or environmental conditions. Moreover, the selection of appropriate RGs depends on their suitability for normalization within the context of the specific experimental conditions being studied. Hence, factors such as RNA extraction methods, and data normalization procedures can also impact the perceived stability of *EF1α* and *H3* across different experiments. Overall, the variability in stability underscores the importance of rigorously evaluating reference gene candidates under diverse experimental conditions to ensure their reliability for accurate gene expression analysis.

Conclusion

In conclusion, our study presents a comprehensive approach to identifying stable RGs for gene expression studies in spinach, utilizing transcriptome data and qRT-PCR validation. The identification of a total of 1196 potential RGs in this study marks a significant achievement, providing a robust foundation for future gene expression analysis in spinach. Additionally, through meticulous screening of candidate RGs and rigorous evaluation using multiple statistical tools, we successfully identified *EF1α* and *H3* as the most stable RGs across various experimental conditions. Overall, our study contributes valuable insights into the dynamics of gene expression in spinach and offers a methodological framework for selecting reliable RGs in plant molecular biology research.

Materials and methods

Transcriptome datasets

Our research on identifying RGs in spinach involved the analysis of various datasets: PRJNA663445, which included seven libraries focusing on tissue-specific expression profiling in leaves, petioles, and roots of the SP75 genotype; PRJNA649901, consisting of eighteen libraries investigating gene expression across five developmental stages of spinach female flowers, from ovary formation to ovule maturation; PRJNA630139[62], which contained twelve libraries targeting bolting and flowering-related genes during the transition from the vegetative to reproductive stage in two spinach accessions; and PRJNA716151, which included six libraries obtained from female sepals post-pollination.

Read mapping, expression level calculation, and filtration

In this investigation, we rigorously conducted a series of analytical steps to ensure the reliability and accuracy of gene expression data in spinach. Initially, clean paired-end reads from each sample were aligned to the spinach genome assembly version 1[63] using the STAR v2.7.1 software[64]. Subsequently, transcripts were assembled employing StringTie v2.0.7[65], with meticulous attention to detail including the utilization of the '-e' option to enhance transcript reconstruction accuracy. The resulting transcripts read-count data matrices were generated for comprehensive analysis across all samples using a Python script, prepDE.py. To ensure robustness in subsequent

RT-qPCR analyses, we adopted a well-established method based on CV for the selection of reference genes, as outlined in previous studies[24,37]. Expression levels were quantified using Transcripts Per Million (TPM) values, which were averaged across biological replicates for statistical reliability. To maintain data integrity, genes with log2(TPM) values ≤ 5, indicative of low expression levels prone to noise, were excluded from further analysis. The CV values, calculated as (SD of log2(TPM) divided by the mean log2(TPM), were computed using Microsoft Excel. Our criteria for candidate RGs emphasized stability, with a stringent CV cutoff of 0.2 applied to identify genes exhibiting consistent expression patterns across diverse conditions. This meticulous approach ensures the selection of robust RGs essential for accurate normalization in subsequent RT-qPCR analyses, thereby enhancing the credibility and reproducibility of gene expression studies in spinach.

Function investigation

To gain deeper insights into the functionalities of the identified genes and selection of reference genes, we performed gene classification and Gene Ontology (GO) enrichment analysis on all genes selected in the previous step. This analysis involved utilizing gene functional classification tools and accessing the most recent genomic reference data for *S. oleracea* available on SpinachBase (http://spinachbase.org)[66].

Plant material, RNA isolation, and cDNA synthesis

For the purpose of sampling and RNA extraction, seeds of Viroflay accession were sown in sterilized soil in plastic pots measuring 15 cm in diameter and 25 cm in height. These pots were positioned in a growth chamber at Isfahan University of Technology, Iran, where the spring conditions were regulated with a temperature range of 24–46 degrees Celsius and a photoperiod of 12–14 h of light. The plants were grown for a duration of three months. To obtain these samples the permissions were not necessary. The formal identification of the plant material was undertaken by the herbarium of Agricultural and Natural Resources College, University of Tehran, and no voucher specimens were collected and deposited in the collection (it is not necessary as we don't describe a novel species). We also stated that the field studies were in compliance with local legislation of Iran in the experimental greenhouse and growth chamber of Isfahan University of Technology, Isfahan, and no specific licences were required. To obtain the total RNA, leaf samples were gathered at various stages, spanning from the four-leaf stage to the twelve-leaf stage. To reduce variability arising from inter-individual gene expression disparities, each sample comprised pooled material from a minimum of three plants. The RNA extraction procedure was conducted in triplicate using the DENAzist column RNA isolation kit, adhering to the manufacturer's instructions. The concentration and purity of the extracted RNA were evaluated using a NanoDrop Spectrophotometer and agarose gel. Subsequently, the RNA samples were quantified, and the construction of the cDNA library was executed. In the construction of the cDNA library for reverse transcription quantitative polymerase chain reaction (RT-qPCR), the isolated RNA samples were treated with DNase I enzyme to ensure the elimination of genomic DNA contamination. Specifically, 2 µg of total RNA was treated with 1 U of DNase I (Thermo Fisher Scientific) at 37 °C for 30 min, followed by heat inactivation at 75 °C for 10 min. Afterwards, the DNase I-treated RNA underwent reverse transcription using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT, Thermo Fisher Scientific). A reaction mixture containing 1 µg of DNase I-treated RNA, 200 U M-MLV RT, 500 µM dNTPs, 5 µM random hexamer primers, 10 mM DTT, and 20 U RNase inhibitor was incubated at 37 °C for 1 h, followed by enzyme inactivation at 70 °C for 15 min.

Gene selection and primer design

Twelve genes were chosen for validation through qRT-PCR at different stages of development. This set of genes included commonly used RGs and novel candidates. Subsequently, we utilized the Primer3 tool (https://www.primer3plus.com/index.html)[67] to design primers for the selected genes (Table 3). Our primer design took into

Gene ID	Gene name	Gene Abbreviation	SD	CV	Forward sequence	Reverce sequence	Amplification length
Spo07600	Polyubiquitin (Ubiquitin)	UBQ	0.336	0.041	ATGGAAGAACACTCGCCGATTA	CTTATCATCGGCAAGCTGCTTC	236
Spo17662	60 S ribosomal protein L27	RPL27	0.489	0.041	ACCACCACAGAATCATGTTCGA	GGCCTTCTCCTTGACATCTTCA	155
Spo00342	Eukaryotic translation initiation factor	EIF	0.424	0.043	CACCTCGTAAACCAGAACCTGA	ACAGCATCAACTACCTCTTGGG	175
Spo08258	Pentatricopeptide repeat superfamily protein	PPR	0.455	0.046	TAACTCAAACTCTCCGAACCCG	CCAAGGATGACAATGGTGAGGA	184
Spo13459	glycine-rich protein	GRP	0.459	0.045	TGGAATACCCGGGTACAATGAC	TAACCCTTACCGGCCCTACTAA	198
Spo24936	Elongation factor 1-alpha	*EF1α*	1.033	0.086	GACTCAAAGAACGACCCTGCTA	ATTGGCTTGGTGGGAATCATCT	248
Spo25393	Elongation factor 1 beta	*EF1b*	0.658	0.106	CGATGAGGAAGATGACGACGAT	TTCATCATCCCATGGCTTCACA	160
Spo20857	Tubulin alpha chain, putative	*TUBα*	1.411	0.181	CTCTGAGGTTTGATGGTGCTCT	ACACTTGGCCATCATGGAAGAT	194
Spo14438	Histone H3	H3	0.978	0.148	ACTGAGCTTTTGATCCGCAAAC	TGTCCTTAGGCATGATGGTCAC	199
Spo22714	ADP-ribosylation factor family protein	*ARF*	0.665	0.100	GTGGCGACTGGAGGATATAGTG	GAGTGAGGGTTTGCTGAGAAGA	243
Spo24687	glyceraldehyde-3-phosphate dehydrogenase	*GAPDH*	-	-	GGCTGCCATCAAGGAGGAAT	GCAATTCCAGCCTTGGCATC	129
Spo23599	*actin*	ACT	-	-	TGTTCACGACATCAGCCGAA	CGTCGGGTAGCTCGTAGTTC	138

Table 3. Spinach RGs information and their primer sequences.

account various factors, including melting temperature, self-complementarity, hairpin potential, and primer product sizes. By incorporating these criteria, we successfully developed specific and efficient primers for amplifying the target genes in our study.

Validation of the reference gene expression stability by RT-qPCR assay

To validate the expression of genes, we conducted qRT-PCR to quantitatively assess the expression of candidate genes in leaf tissues at various developmental stages. qRT-PCR experiments were carried out in triplicate using an StepOne Real-Time PCR system in a final volume of 15 µL. This volume included 7.5 µL of SYBR Green Master Mix (BioFACT, Korea), 2 µL of diluted cDNA, and 0.5 µL of each primer (10 pM), with the remainder filled with PCR-grade water. The qRT-PCR protocol involved an initial step of 5 min at 95 °C, followed by 40 cycles of 10 s at 95 °C, 20 s at the primer-specific annealing temperature, 20 s at 72 °C, and concluded with a melting curve program.

Statistical analysis

In order to thoroughly assess the stability of RGs, we conducted a rigorous analysis using different statistical methods. These methods, namely geNorm (https://genorm.cmgg.be/[39]), NormFinder (http://moma.dk/[40]), and BestKeeper (www.gene-quantification.com/bestkeeper.html[14]), were chosen based on their proven effectiveness in similar studies[24,37,57]. The experimental design involved evaluating the stability of twelve RGs across various samples from different developmental stages. For geNorm, and NormFinder, we utilized Ct values obtained from qRT-PCR. These values were then transformed into relative expression levels using the formula $2^{-\Delta CT}$, where ΔCT represents the difference between each Ct value and the lowest Ct value for that particular gene across different samples. Subsequently, these transformed values were subjected to analysis within geNorm to calculate the M values, which indicate the stability of gene expression. Lower M values corresponded to more stable expression. NormFinder, on the other hand, employed a mathematical model to assess the variance within and between groups in order to identify stable RGs. The SV determined the most suitable individual or set of genes for normalization purposes. In parallel, BestKeeper software analyzed the raw Ct data. The former calculated the standard deviations (SD) of Ct values for each RG, with higher stability associated with lower SD values. BestKeeper, on the other hand, utilized the coefficient of variance (CV) and SD of Ct values, with stability determined by the correlation coefficient between candidate genes. Lower SD and CV values were indicative of higher stability. Additionally, RefFinder, an online tool that integrates the aforementioned methods, facilitated a comprehensive ranking of RGs. The final overall ranking was determined by calculating the geometric mean of rankings from Delta Ct, geNorm, NormFinder, and BestKeeper.

RGs validation

In order to validate the reliability of the chosen RGs, the expression profiles of FLAVIN-BINDING KELCH REPEAT F-BOX 1 (*FKF1*) and LATE ELONGATED HYPOCOTYL (*LHY*) were examined at various stages, representing the photoperiod and circadian pathways, respectively. These genes play a crucial role in regulating the timing of flowering/bolting. The quantification of *FKF1* and *LHY* gene expression was determined using the $2^{-\Delta\Delta Ct}$ formula.

Data availability

All RNA-Seq data are deposited in the NCBI SRA database under the projects PRJNA663445, PRJNA649901, PRJNA630139, and PRJNA716151.

Received: 7 May 2024; Accepted: 17 September 2024
Published online: 27 September 2024

References

1. Bustin, S. A., Benes, V., Nolan, T. & Pfaffl, M. W. Quantitative real-time RT-PCR—A perspective. *J. Mol. Endocrinol.* **34**, 597–601 (2005).
2. Kubista, M. *et al.* The real-time polymerase chain reaction. *Mol. Aspects Med.* **27**, 95–125 (2006).
3. Green, M. R. & Sambrook, J. Quantification of RNA by real-time reverse transcription-polymerase chain reaction (RT-PCR). *Cold Spring Harb. Protoc.* **2018** (2018).
4. Gachon, C., Mingam, A. & Charrier, B. Real-time PCR: What relevance to plant studies?. *J. Exp. Bot.* **55**, 1445–1454 (2004).
5. Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* **1**, 1559–1582 (2006).
6. Zhao, X. *et al.* Reference gene selection for quantitative real-time PCR of Mycelia from lentinula edodes under high-temperature stress. *Biomed. Res. Int.* **2018**, 1670328 (2018).
7. Paolacci, A. R., Tanzarella, O. A., Porceddu, E. & Ciaffi, M. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol. Biol.* **10**, 11 (2009).
8. Kumar, D., Das, P. K. & Sarmah, B. K. Reference gene validation for normalization of RT-qPCR assay associated with germination and survival of rice under hypoxic condition. *J. Appl. Genet.* **59**, 419–430 (2018).
9. Sanders, R., Mason, D. J., Foy, C. A. & Huggett, J. F. Considerations for accurate gene expression measurement by reverse transcription quantitative PCR when analysing clinical samples. *Anal. Bioanal. Chem.* **406**, 6471–6483 (2014).
10. Harshitha, R. & Arunraj, D. R. Real-time quantitative PCR: A tool for absolute and relative quantification. *Biochem. Mol. Biol. Educ. Bimon Publ. Int. Union Biochem. Mol. Biol.* **49**, 800–812 (2021).
11. Zhu, X. *et al.* Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. *PLoS One* **7**, e44405 (2012).
12. Qian, J. *et al.* Selection and evaluation of appropriate reference genes for RT-qPCR Normalization of *Volvariella volvacea* gene expression under different conditions. *Biomed Res. Int.* **2018**, 6125706 (2018).
13. Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E. F. & Hellens, R. P. Protocol: A highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* **3** (2007).

14. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
15. Selvey, S. *et al.* Beta-actin–an unsuitable internal control for RT-PCR. *Mol. Cell. Probes* **15**, 307–311 (2001).
16. Chapman, J. R. & Waldenström, J. With reference to reference genes: A systematic review of endogenous controls in gene expression studies. *PLoS One* **10**, e0141853 (2015).
17. Artico, S., Nardeli, S. M., Brilhante, O., Grossi-de-Sa, M. F. & Alves-Ferreira, M. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biol.* **10**, 49 (2010).
18. Schmittgen, T. D. & Zakrajsek, B. A. Effect of experimental treatment on housekeeping gene expression: Validation by real-time, quantitative RT-PCR. *J. Biochem. Biophys. Methods* **46**, 69–81 (2000).
19. Bustin, S. A. & Nolan, T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* **15**, 155–166 (2004).
20. VeraHernández, F. P. *et al.* Reference genes for RT-qPCR normalisation in different tissues, developmental stages and stress conditions of amaranth. *Plant Biol. Stuttg.* **20**, 713–721 (2018).
21. Lucho, S. R. *et al.* Validation of reference genes for RT-qPCR studies in Stevia rebaudiana in response to elicitor agents. *Physiol. Mol. Biol. Plants Int. J. Funct. Plant Biol.* **24**, 767–779 (2018).
22. Jia, Y. *et al.* Reference gene selection and validation by qRT-PCR during flower development and in different organs of Primula forbesii. *J. Hortic. Sci. Biotechnol.* **95**, 383–394 (2020).
23. Chen, C. *et al.* Identification of reliable reference genes for quantitative real-time PCR normalization in pitaya. *Plant Methods* **15**, 70 (2019).
24. Li, Y. *et al.* Systematic identification and validation of the reference genes from 60 RNA-Seq libraries in the scallop *Mizuhopecten yessoensis*. *BMC Genom.* **20**, 288 (2019).
25. Kozera, B. & Rapacz, M. Reference genes in real-time PCR. *J. Appl. Genet.* **54**, 391–406 (2013).
26. Zhang, C., Fu, J., Wang, Y., Bao, Z. & Zhao, H. Identification of suitable reference genes for gene expression normalization in the quantitative real-time PCR analysis of sweet Osmanthus (*Osmanthus fragrans* Lour.). *PLoS One* **10**, e0136355 (2015).
27. Scharf, K.-D., Berberich, T., Ebersberger, I. & Nover, L. The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. *Biochim. Biophys. Acta* **1819**, 104–119 (2012).
28. Migocka, M. & Papierniak, A. Identification of suitable reference genes for studying gene expression in cucumber plants subjected to abiotic stress and growth regulators. *Mol. Breed.* **28**, 343–357 (2011).
29. Sen, M. K., Hamouzová, K., Košnarová, P., Roy, A. & Soukup, J. Identification of the most suitable reference gene for gene expression studies with development and abiotic stress response in Bromus sterilis. *Sci. Rep.* **11**, 13393 (2021).
30. Xie, H. *et al.* Selection and validation of reference genes for RT-qPCR analysis in *Spinacia oleracea* under abiotic stress. *Biomed Res. Int.* **2021**, 4853632 (2021).
31. You, S. *et al.* Selection and validation reference genes for qRT-PCR normalization in different cultivars during fruit ripening and softening of peach (*Prunus persica*). *Sci. Rep.* **11** (2021).
32. Duan, M. *et al.* Identification of optimal reference genes for expression analysis in radish (*Raphanus sativus* L.) and its relatives based on expression stability. *Front. Plant Sci.* **8**, 1605 (2017).
33. Rodrigues, T. B. *et al.* Validation of reference housekeeping genes for gene expression studies in western corn rootworm (*Diabrotica virgifera* virgifera). *PLoS One* **9**, e109825 (2014).
34. Deguchi, M. *et al.* Selection and validation of reference genes for normalization of qRT-PCR data to study the cannabinoid pathway genes in industrial hemp. *PLoS One* **16**, e0260660 (2021).
35. Hong, S.-Y., Seo, P. J., Yang, M.-S., Xiang, F. & Park, C.-M. Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR. *BMC Plant Biol.* **8**, 112 (2008).
36. Gutierrez, L. *et al.* The lack of a systematic validation of reference genes: A serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnol. J.* **6**, 609–618 (2008).
37. Chen, J. *et al.* Identification and validation of the reference genes in the echiuran worm *Urechis unicinctus* based on transcriptome data. *BMC Genom.* **24**, 248 (2023).
38. Smitha, P. K. *et al.* Genome wide search to identify reference genes candidates for gene expression analysis in *Gossypium hirsutum*. *BMC Plant Biol.* **19**, 405 (2019).
39. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 00341 (2002).
40. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**, 5245–5250 (2004).
41. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25**, 402–408 (2001).
42. Xie, F., Wang, J. & Zhang, B. RefFinder: A web-based tool for comprehensively analyzing and identifying reference genes. *Funct. Integr. Genom.* **23**, 125 (2023).
43. Remans, T. *et al.* Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. *Planta* **227**, 1343–1349 (2008).
44. Bevitori, R. *et al.* Selection of optimized candidate reference genes for qRT-PCR normalization in rice (*Oryza sativa* L.) during *Magnaporthe oryzae* infection and drought. *Genet. Mol. Res.* **13**, 9795–9805 (2014).
45. Pombo, M. A., Zheng, Y., Fei, Z., Martin, G. B. & Rosli, H. G. Use of RNA-seq data to identify and validate RT-qPCR reference genes for studying the tomato-Pseudomonas pathosystem. *Sci. Rep.* **7**, 44905 (2017).
46. Alves Oliveira, D., Tang, J. D. & Warburton, M. L. Reference gene selection for RT-qPCR Analysis in maize kernels inoculated with *Aspergillus flavus*. *Toxins (Basel).* **13** (2021).
47. Cai, X. *et al.* Genomic analyses provide insights into spinach domestication and the genetic basis of agronomic traits. *Nat. Commun.* **12**, 7246 (2021).
48. Hirakawa, H. *et al.* A spinach genome assembly with remarkable completeness, and its use for rapid identification of candidate genes for agronomic traits. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genom.* **28** (2021).
49. Ma, X. *et al.* The spinach YY genome reveals sex chromosome evolution, domestication, and introgression history of the species. *Genome Biol.* **23**, 75 (2022).
50. Ribera, A., Bai, Y., Wolters, A.-M.A., van Treuren, R. & Kik, C. A review on the genetic resources, domestication and breeding history of spinach (*Spinacia oleracea* L.). *Euphytica* **216**, 48 (2020).
51. Quan, P.-L., Sauzade, M. & Brouzes, E. dPCR: A technology review. *Sensors (Basel).* **18** (2018).
52. Narsai, R., Ivanova, A., Ng, S. & Whelan, J. Defining reference genes in *Oryza sativa* using organ, development, biotic and abiotic transcriptome datasets. *BMC Plant Biol.* **10**, 56 (2010).
53. Wang, X. *et al.* Identification and evaluation of reference genes for quantitative real-time PCR analysis in *Polygonum cuspidatum* based on transcriptome data. *BMC Plant Biol.* **19**, 498 (2019).
54. Jian, H., Wang, H., Qiu, X., Yan, H. & Ma, L. Identification and validation of reference genes for qRT-PCR analysis of petal-color-related genes in *Rosa praelucens*. *Genes (Basel).* **15** (2024).
55. Meng, Y., Li, N., Tian, J., Gao, J. & Zhang, C. Identification and validation of reference genes for gene expression studies in postharvest rose flower (*Rosa hybrida*). *Sci. Hortic. (Amsterdam)* **158**, 16–21 (2013).

56. Wan, Y., Hong, A., Zhang, Y. & Liu, Y. Selection and validation of reference genes of *Paeonia lactiflora* in growth development and light stress. *Biol. Plants Int. J. Funct. Plant Biol.* **25**, 1097–1105 (2019).
57. Zheng, H., Zhao, H., Zhang, X., Liang, Z. & He, Q. Systematic identification and validation of suitable reference genes for the normalization of gene expression in *Prunella vulgaris* under different organs and spike development stages. *Genes (Basel).* **13** (2022).
58. Ji, T. *et al.* Reference genes identification for qRT-PCR normalization of gene expression analysis in *Cucumis sativus* under *Meloidogyne incognita* infection and Pseudomonas treatment. *Front. Plant Sci.* **13**, 1061921 (2022).
59. Wang, C. *et al.* Identification and validation of reference genes for RT-qPCR analysis in non-heading Chinese cabbage flowers. *Front. Plant Sci.* **7**, 811 (2016).
60. Gu, C.-S. *et al.* Reference gene selection for quantitative real-time RT-PCR normalization in Iris. lactea var. chinensis roots under cadmium, lead, and salt stress conditions. *Sci. World J.* **2014**, 532713 (2014).
61. Saddhe, A. A., Malvankar, M. R. & Kumar, K. Selection of reference genes for quantitative real-time PCR analysis in halophytic plant *Rhizophora apiculata*. *PeerJ* **6**, e5226 (2018).
62. Abolghasemi, R., Haghighi, M., Etemadi, N., Wang, S. & Soorni, A. Transcriptome architecture reveals genetic networks of bolting regulation in spinach. *BMC Plant Biol.* **21**, 179 (2021).
63. Xu, C. *et al.* Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. *Nat. Commun.* **8** (2017).
64. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
65. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
66. Collins, K. *et al.* SpinachBase: A central portal for spinach genomics. *Database* **2019** (2019).
67. Untergasser, A. *et al.* Primer3–new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).

Author contributions

A.S.: Conceptualization; funding acquisition; investigation; project administration; methodology; formal analysis; validation; writing-original draft; writing-review and editing; M.R.: investigation; methodology; formal analysis, writing-review, and editing. H.B.: investigation; methodology; formal analysis, writing-review and editing.

Funding

No funding was received for conducting this study.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-73444-2.

Correspondence and requests for materials should be addressed to A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.