scientific reports

Check for updates

OPEN

Reference genes selection and validation in yam by real-time quantitative polymerase chain reaction

Linan Xing[1,4], Yanfang Zhang[1,4], Yanping Xing[2], Mingran Ge[1], Huiqin Miao[3] & Xiuwen Huo[1]✉

Gene expression pattern analysis plays a crucial role in omics research, as it helps us understand the regulatory mechanisms of gene expression and the associated biological processes. Real-time quantitative polymerase chain reaction (q-PCR) is an efficient method for studying gene expression. Reliable q-PCR results depend on proper data normalization, which requires the use of stable reference genes. Yam, valued as both food and medicine, has significant commercial and economic potential. Its cultivation is being actively promoted in the central and western regions of Inner Mongolia. However, studies have shown that no reference gene is universally stable under all conditions. Therefore, screening and validating suitable reference genes are essential for accurately studying gene expression patterns in yam. This study evaluated nine potential reference genes based on transcriptome data of Chinese yam (*Dioscorea opposita* Thunb.), under various conditions, including yam tuber development, different tissues, High temperature, Low temperature, salt stress (NaCl), drought stress (PEG), abscisic acid stress (ABA), and methyl jasmonate treatment (MeJA). Using geNorm, NormFinder, BestKeeper, and RefFinder, we ranked these nine reference genes according to their expression stability. The results revealed that *EIF* was not suitable as a reference gene. Conversely, *UBQ* and *PP2A* were identified as the most stable and ideal reference genes for expression analysis across different conditions. Our findings provide a theoretical foundation for selecting reference genes in yam and will support more accurate gene expression studies within the genus *Dioscorea*.

Keywords Gene expression, Internal reference gene, Normalization, RT-qPCR, Yam

Chinese yam (*Dioscorea opposita* Thunb.) is a monocot belonging to the Dioscoreaceae family[1]. It has been an underutilized crop that is being neglected globally[2]. The yam tubers are highly nutritious, containing secondary metabolites that offer health benefits such as regulating blood sugar levels and controlling cholesterol, fat absorption and hypertension[3–5]. Rich in resistant starch, polysaccharides, and steroidal saponins, Chinese yam has become a popular vegetable in Southeast Asia, contributing to an improved and diversified the local diet[6]. In China, yam has been used as both food and traditional medicine for thousands of years, underscoring its economic value and potential for the development of health-related products. In recent years, the cultivation area of Chinese yam has expanded significantly, with the planting area of Huaishan yam in China exceeding 200,000 hectares annually. The planting area of Huaishan yam in China has reached more than 200,000 hm^2. However, challenges such as soil microbial imbalance, soil nutrient depletion, high soil temperatures, and the accumulation of pathogenic bacteria hinder yam cultivation, becoming major factors restricting its production and industrialization. With the expansion of northern yam species to the West, the planting area of these varieties has increased rapidly in regions such as Shaanxi, Gansu and Ningxia, Xinjiang and western Inner Mongolia[7]. The western region of Inner Mongolia offers rich land resources, sandy soil, abundant sunlight, and significant temperature differences between day and night, making it an ideal location for yam cultivation. However, factors such as the short frost-free period, drought, and saline-alkali soil conditions negatively impact yam production in this area. As a result, breeding and screening yam varieties suited for these challenging conditions have become urgent priorities. Additionally, studying the genetic resistance of yam to abiotic stresses and analyzing their gene expression are essential steps in developing varieties suitable for local production.

[1]College of Horticulture and Plant Protection, Inner Mongolia Agricultural University, Hohhot 010019, Inner Mongolia, China. [2]College of Life and Science, Inner Mongolia Agricultural University, Hohhot 010019, Inner Mongolia, China. [3]Library, Inner Mongolia Agricultural University, Hohhot 010019, Inner Mongolia, China. [4]Linan Xing and Yanfang Zhang contributed equally to this work. ✉email: huoxiuwen@imau.edu.cn

With the advancement of molecular biology, the research on plant functional gene expression has become a prominent focus.. Real-time quantitative PCR (RT-qPCR) is a highly effective method for quantitatively analyzing target gene expression. It has emerged as a reliable technique for detecting and measuring gene expression levels in molecular biology research. While genome and transcriptome sequencing can also detect gene expression, RT-qPCR is widely favored for its high sensitivity, flexibility, and efficiency[8–10].

The quantitative process is influenced by factors such as the quality of RNA and cDNA, primer specificity, and PCR amplification efficiency[11]. Consequently, the quantitative analysis of target genes requires stable internal reference genes for normalization[12–14]. Internal reference genes are expected to exhibit stable expression across various conditions. However, their expression levels can vary significantly depending on environmental factors, plant tissues, and physiological states[15,16]. Therefore, selecting stable reference genes specific to each species and treatment method is essential to improve the accuracy of gene expression calibration and standardization[17]. Genes that maintain basic cellular functions and encode essential proteins are frequently used as reference genes, as their products are necessary for cellular activity and are typically expressed under a wide range of conditions. These genes include the 18S ribosomal RNA (*18S*) gene[18], ubiquitin (*UBQ*)[19], actin(*ACT*)[20], and glycero-aldehyde-3-phosphate dehydrogenase (*GAPDH*) gene[21]. *GADPH* is not an optimal reference gene for ABA treatment in *Isodon rubescens*. However, it is suitable for use in the roots of *Avena sativa* L. under drought stress and the leaves under salt stress[22,23]. For *Iris lactea* var. *Chinensis*, the most suitable internal reference genes under cadmium stress are *EIF-5A* and *UBC*. Previous studies have identified *TIP41*, *CYP*, *PGK*, *GAP*, and *PP2Acs* as the most suitable internal reference genes under abiotic stresses, such as salt, drought, and cold[24]. Among different varieties and tissues of *Zanthoxylum bungeanum*, *UBQ* and *UBA* expression are the most stable during different development stages of fruit[25]. Screening for stable reference genes using q-PCR analysis has been conducted in many plants, including sweet potato[26], *Lilium henryi* Baker[27], *Broussonetia papyrifera*[28], and *Allium sativum* L. (garlic) [29]. The evaluation of reference genes is based on Ct values from qRT-PCR, with widely used software including using the delta-Ct method[30], geNorm[31], NormFinder[32], BestKeeper[33], and RefFinder[34]. Despite the maturity of reference gene screening methods, there are limited studies on reference genes specific to yam and its tissues under various treatments. This lack of research significantly restricts the effective application of reference genes in yam studies.

This study aims to identify reliable internal reference genes for analyzing target gene expression in yam. Nine candidate reference genes were selected based on yam transcriptome data for gene standardization. To assess the suitability of the selected candidate genes and ensure the accuracy of the results, we used both the most stable and the most unstable internal reference genes identified in this study. We then measured the relative expression levels of *WRKY* genes in yam across different developmental stages, tissues, and stress conditions using RT-qPCR. This research provides a theoretical foundation for the subsequent screening of resilient yam varieties and further gene function studies.

Materials and methods

Experimental material

Dioscorea opposita variety 'Dahechangyu' was used as the experimental material in this study. The yam plants were grown at College of Horticulture and Plant Protection, Inner Mongolia Agricultural University, Hohhot, China, was used as experimental material in the present study. The experimental fields were managed according to standard agricultural practices. Tubers were sampled at intervals of 90, 105, 120, 135, 150 and 165 days after planting. In October 2023, the roots, stems, and leaves of the yam were collected at the peak of tuber expansion. Yam tubers that had been released from dormancy and were of similar size were selected and planted into plastic pots (top diameter 10 cm, bottom diameter 8 cm), which were filled with a substrate composed of peat and perlite (1:1 ratio). The plants were cultured in a chamber at 25 °C with a 16/8 h day/night photoperiod and 65–75% relative humidity. When the yam seedlings reached approximately 15 cm in height(about one month after planting), the following treatments were applied: For high temperature stress, seedlings were grown in the chamber at 36 °C; for cold stress, seedlings were grown in the chamber at 4 °C; for salt stress, seedlings were grown in water containing 200 mmol/L NaCl; for drought stress, seedlings were grown in water containing 20% PEG. Yam leaves were collected at 0, 12, 24, and 48 h after each treatment. For abscisic acid stress (ABA) and methyl jasmonate treatment (MeJA) stress, seedlings were grown under normal growth conditions and sprayed with ABA/MeJA (100 mmol/L). The leaves of yam were collected at 0, 1, 6, 12, and 24 h after treatment. All samples were immediately frozen in liquid nitrogen and stored at − 80 °C. All treatments were conducted in triplicate.

Total RNA isolation and cDNASynthesis

Total RNA was extracted using an RNA extraction kit (9769; TaKaRa, Beijing). The quality of the RNA samples was assessed using a NanoDrop 2000c (Thermo Fisher Scientific, USA) and gel electrophoresis. First-strand cDNA synthesis was performed using cDNA Synthesis SuperMix (RR047A; TaKaRa, Beijing) and stored at − 20 °C.

Selection of yam sequences, primer design and PCR optimization

The candidate reference genes were selected based on our transcriptome data[35] and previous reports of reference genes used in studies of various plant species[36–38]. Primers were designed with Primer Premier 5.0, with melting temperatures between 51–60 °C, sizes of 17–20 bp, and approximately 50% GC content (Table 1). Before qRT-PCR, the primer pairs were tested using a standard PCR reaction on a Mastercycler Gradient (Eppendorf) to determine the optimal conditions. Amplicons of the expected size were verified by 1% agarose gel electrophoresis and ethidium bromide staining. The amplification efficiency (E) of the candidate genes was calculated using a

Gene name	Tentative annotation	Primer sequence	Tm (°C)	Amplicon length (bp)	Slope	E (%)	R^2
F-box	F-box/kelch-repeat protein At1g22040-like	F:TACTGGCTGACATGCTAAAG	51.8	133	− 3.0177	113.79	0.9999
		R:TTGGTCTCAGGGTCATAGAT					
ACT	Actin	F:CCATACTTCGCTTGGATCTT	51.3	154	− 3.4615	94.98	0.999
		R:TGTTCGTAATCAAGAGCGAT					
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	F:CAGCTCATTTGAAGGGTGG	54.5	156	− 3.0462	108.93	0.9827
		R:CAAGAGGAGCAAGGCAGTT					
PP2A	Phosphoprotein phosphatase	F:ATGCCAATGTCTGGAAGTA	51.3	139	− 3.188	104.17	0.9944
		R:CGGTCAAGAGCACGAAT					
18S	18S ribosomal RNA gene	F:CCATAAACGATGCCGACCAG	58.4	109	− 3.467	94.98	0.9999
		R:AGCCTTGCGACCATACTCCC					
GUSB	Beta-glucuronidase	F:GCCGAGCGGATGTAAGA	54.5	141	− 3.222	104.17	0.9974
		R:TGTTGTGAGTTGCCCTGT					
UBQ	Ubiquitin	F:GGGCTTTCAAGGTCGTC	52.9	100	− 2.9803	96.64	0.9998
		R:TGAAGGGTTTGCTCATCC					
CHC1	clathrin heavy chain 1	F:CAGACACCACCATTACTTCA	51.6	101	− 3.2126	104.17	0.9977
		R:TTTTGGTTCACAACAAGACG					
EIF	eukaryotic translation initiation factor	F:TCATCTTTACCATGCCTACG	51.2	122	− 2.7736	96.39	0.9997
		R:AACAGGACAGATGAAGCAAT					
WRKY46a	WRKY transcription factor	F:GAAGGCACATCTAAATGAACAGC	55.2	148			
		R:AGTAGAGTTAAGCCACACCGAAG					
WRKY51	WRKY transcription factor	F:GAGCACTCTAATGGTGGTTTG	56.1	160			
		R:AGAGAGCTGGGAGACGAGATG					

Table 1. Primer sequences and PCR amplification characteristics information on the candidate reference genes of yam.

standard curve (a fivefold dilution series of cDNA was used as the template) by q-PCR. The formula for efficiency is: E (%) = $(10^{-1}/\text{slope}-1)\times 100\%$[39].

qPCR analysis of reference genes

q-PCR was performed using a CFX Connect™ real-time system (Bio-Rad, Singapore) followding the instructions provided by TaKaRa Biotech (9769; TaKaRa, Beijing, China) for the Super Real Fluorescence Quantitative Premix Reagent-Enhanced Kit. The reaction mixture cinsisted of 10 µL of 2 × SYBR Green RT-qPCR Master Mix, 8.2 µL of ddH_2O, 1 µL of cDNA, and 0.4 µL each of forward and reverse amplification primers (10 µmol/L), for a final volume of 20 µL. The PCR conditions included denaturation at 95 °C for 30 s, followed by 40 cycles of 5 s at 95 °C and 30 s at 60 °C. The melting curve was set to the instrument's default. All samples were tested with three technical replicates and three independent biological replicates.

Stability analysis and validation of selected rreference gGenes

Delta-Ct, BestKeeper, geNorm, NormFinder and RefFinder software programs were used to analyze the expression stability of the candidate genes of in yam across different developmental stages, different tissues and different abiotic stresses. The geNorm and NormFinder analyses employed the $2^{-\Delta CT}$ method for calculations, while BestKeeper analyzed the Ct values. RefFinder was employed for a comprehensive evaluation, determining the most stable reference gene(s). The geometric mean of all rankings was used to calculate the overall expression stability ranking, aiding in the selection of the most suitable reference genes for yam. Based on the selected reference genes, the expression patterns of *WRKY* during yam development and under abiotic stress were analyzed to further validate stability of the reference genes. The samples and procedure are the same as qPCR analysis of reference genes, and relative expression level was calculated using the $2^{-\Delta\Delta CT}$ method.

Results

Primer specificity and amplification efficiency of candidate reference genes

Agarose gel electrophoresis results showed that the fragments for the nine candidate reference genes (*F-box*, *ACT*, *GAPDH*, *PP2A*, *18S*, *GUSB*, *UBQ*, *CHCI* and *EIF*) (Fig. 1A) and two validation genes (Fig. 1B) were between 100 and 200 bp, which was consistent with the expected fragment sizes. Each product was well amplified with no primer dimers, indicating that the selected reference gene primers were suitable for use in subsequent experiments. Additionally, qRT-PCR was performed using cDNA templates from different samples, and the melting curve was used for evaluation. The melting curves of the nine candidate reference genes all showed single peaks, indicating specific primers and reproducible amplification curves between samples (Fig. 2).

The efficiency of primer amplification (E) was calculated based on the slope (K) and correlation coefficient (R^2) derived from the standard curve(Supplementary materials). The amplification of the candidate genes

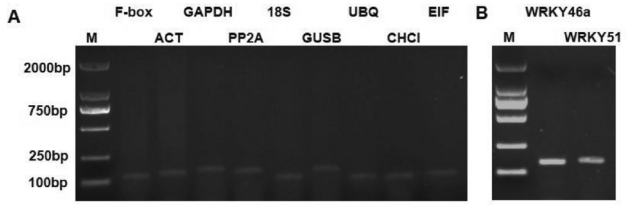

Fig. 1. Specificity analysis of nine candidate reference gene primers of Yam. M: 2000 marker; (**A**) PCR of nine gene primers for internal reference gene screening; (**B**) Two genes for internal reference gene validation.

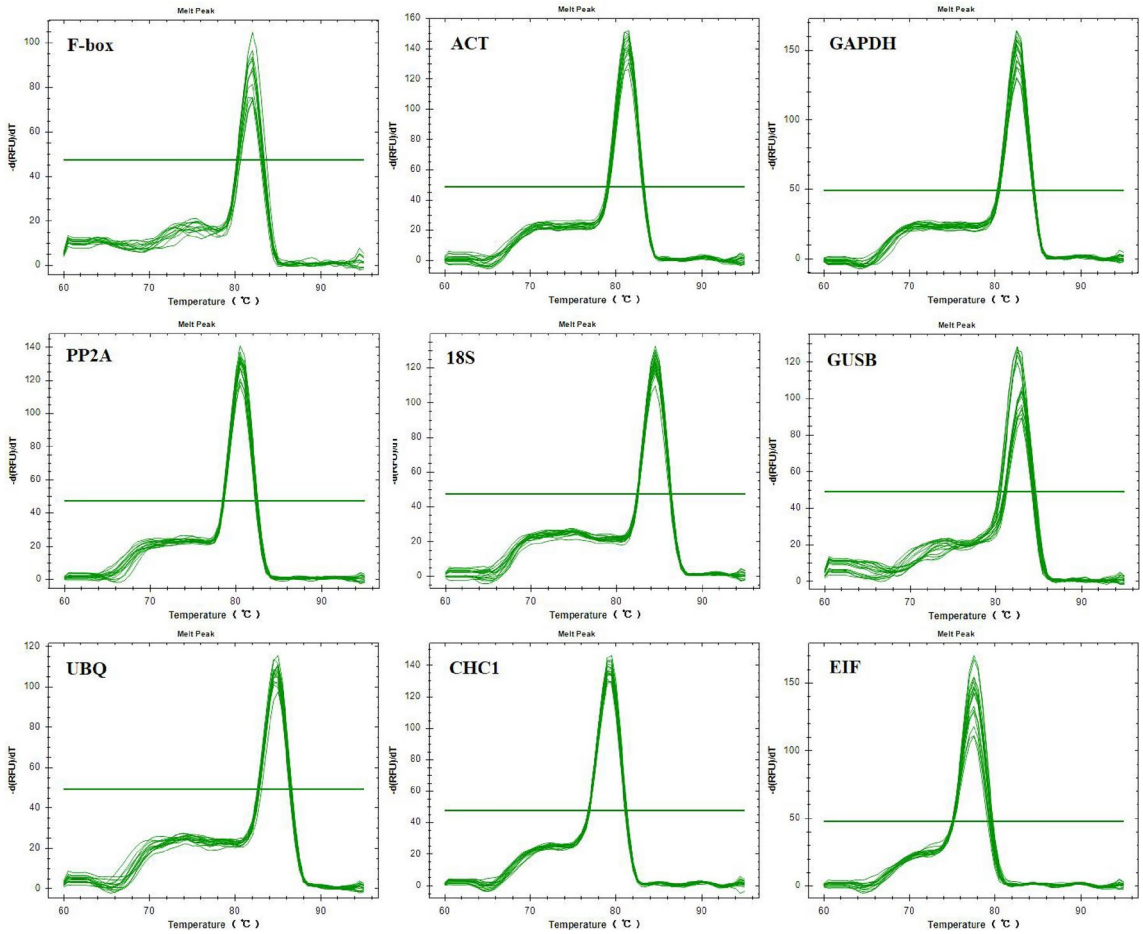

Fig. 2. Melting curves of qPCR products of nine candidate reference genes of yam.

ranged from 94.98% to 113.79% with all R^2 values exceeding 0.98 (Table 1). These results indicate strong primer amplification efficiency, meeting the stringent criteria for qRT-PCR.

Analysis of candidate gene expression abundance

To assess the expression consistency of the nine potential reference genes across various tissues and different treatment conditions, the transcriptional abundances were represented as Ct values. The expression levels of candidate reference genes were determined by measuring the cycle threshold (Ct) values using RT- qPCR. The Ct value is inversely proportional to gene expression abundance, meaning a lower Ct value indicates higher gene expression. The Ct values of the nine candidate reference genes ranged from 6.45 to 34.53 across different tuber expansion periods, tissues, and abiotic stress conditions. The smallest variation in gene expression was observed for *CHCI* (2.79) and *UBQ* (3.34), while *EIF* (11.41) was the gene with the most variable levels of expression. During different tuber expansion periods of yam, the Ct values of the candidate genes ranged from 6.45 to 32.38. In different tissues, *18S* had the lowest mean Ct value, while *EIF* had the highest. For different abiotic stresses, the coefficient of variation for *18S* was relatively low, whereas *EIF* exhibited the highest coefficient of variation (Table 2). These finding suggest that none of the selected genes exhibited a constant level of

Treaments		Ct mean values								
		F-box	*ACT*	*GAPDH*	*PP2A*	*18S*	*GUSB*	*UBQ*	*CHCI*	*EIF*
Different stages of tuber development	90d	25.86	19.75	18.54	25.88	7.58	25.63	24.13	22.49	31.09
	105d	22.33	17.49	15.59	21.33	6.45	26.11	24.28	22.29	29.16
	120d	29.03	23.98	23.32	27.39	12.48	24.15	22.80	21.13	28.47
	135d	30.46	26.27	26.21	29.05	14.86	23.99	23.12	22.06	23.12
	150d	32.38	28.39	25.65	30.02	13.59	24.55	24.40	23.01	27.27
	165d	27.23	26.33	25.81	28.41	12.89	24.23	24.31	23.55	28.59
Different tissues	Root	25.23	22.69	22.03	25.42	12.38	24.85	24.63	22.30	33.93
	Stem	25.13	21.00	19.60	25.13	11.06	22.28	24.24	21.43	30.83
	Leaf	25.84	21.54	19.23	25.67	13.11	21.59	24.47	22.93	34.53
Different abiotic stresses	High temperature	24.33	18.96	18.47	22.94	10.28	22.60	23.05	21.60	27.70
	Low temperature	23.39	18.72	17.98	24.33	10.28	21.77	22.53	21.35	26.95
	NaCl	23.11	18.58	17.75	23.08	9.81	22.25	22.79	21.33	28.06
	PEG	25.69	21.88	20.75	26.65	11.60	24.43	24.64	22.95	30.08
	ABA	26.53	21.90	20.65	25.37	12.29	26.65	25.24	24.23	31.76
	MeJA	26.18	21.82	20.58	25.17	14.19	26.18	25.87	23.97	33.51

Table 2. Internal reference gene Ct value.

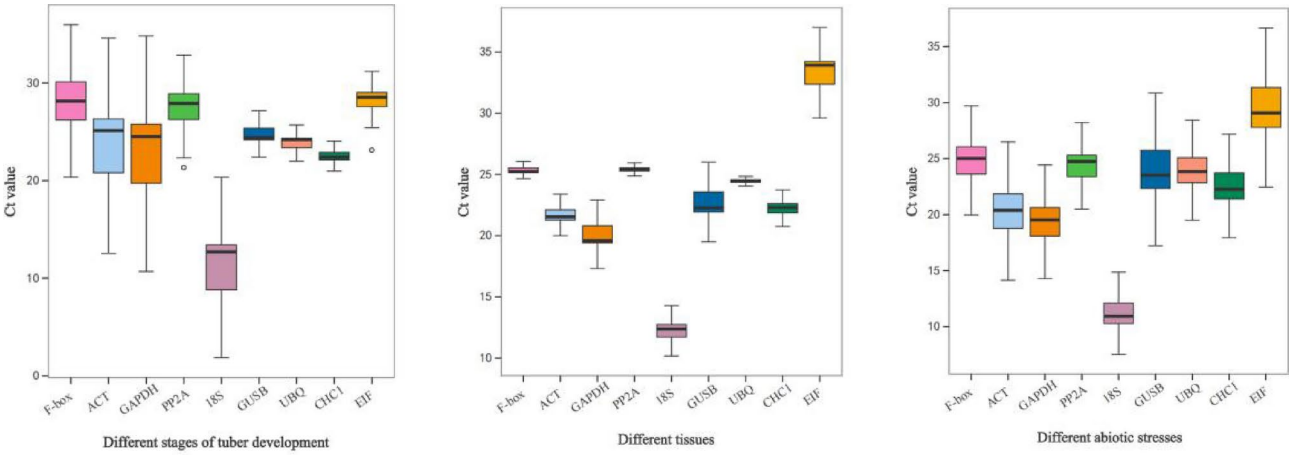

Fig. 3. Ct value distribution of nine candidate reference genes in different developmental stages, different tissues and abiotic stress of yam. The boxes indicate the 25th and 75th percentiles. The line across the box and each box indicate the median and mean Ct values, respectively.

expression across the different *D.opposita* samples tested. The ranking of the expression levels was as follow: *18S* > *GAPDH* > *ACT* > *CHCI* > *UBQ* > *GUSB* > *PP2A* > *F-box* > *EIF* (Fig. 3). Overall, the results indicated that the gene expression levels showed significant divergence across all samples.

Estimation of stability by delta-Ct

Since the nine candidate reference genes exhibited wide variations in expression levels across the different samples, statistical methods were employed to rank the stability of gene expression and determine the number of reference genes required for accurate gene expression profiling under specific experimental conditions. The stability of each reference gene was evaluated by ΔCt, with the gene exhibiting the lowest mean standard deviation (mSD) considered the most stable. The results of the ΔCt analysis showed that *UBQ* was the most stable gene across different developmental stages of tuber expansion as well as among the various tissues. *GUSB* demonstrated the highest stability in High-temperature treated samples, while *ACT* was the most stable in Low-temperature treated samples. *GAPDH* proved to be the most stable in NaCl-treated samples, and *18S* showed the highest stability in PEG-treated samples (Fig. 4). *EIF* was identified as the least stable reference gene across all samples, including different tissues, High-temperature treated, Low-temperature treated, and NaCl-treated conditions.

geNorm analysis

The geNorm analysis evaluates the expression stability of candidate genes by calculating the M value (Expression Stability) for nine candidate reference genes under various developmental stages of tuber growth, different

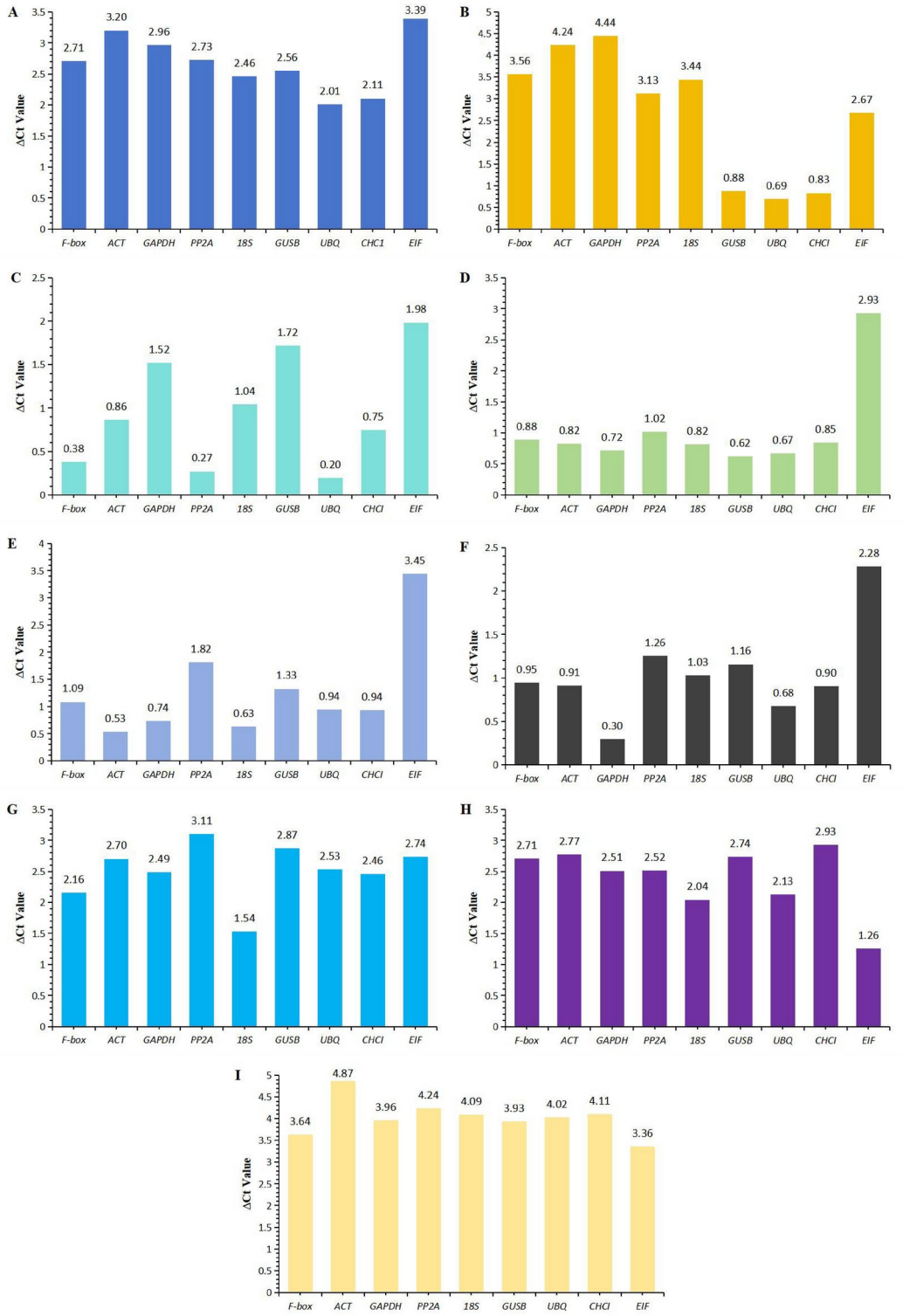

Fig. 4. Stability evaluation of nine reference genes analyzed using ΔCt. (**A**) All samples; (**B**) Different stages of tuber developmental; (**C**) Different tissues; (**D**) High temperature-treated; (**E**):Low temperature-treated; (**F**) NaCl-treated; (**G**) PEG-treated; (**H**): ABA-treated; (**I**) MeJA-treated.

tissues, and abiotic stress conditions. Developed by Vandesompele et al. in 2002, geNorm is widely used to assist in screening reference genes[31]. A smaller M value indicates greater stability in gene expression. Genes with an M value below 1.5 are considered suitable for use as a reference genes. Figure 5 shows the changes in M values under different conditions and tissues. The expression stability values of all reference samples were assessed, and all nine candidate reference genes had M values below 1.5. In the different developmental stages, the lowest

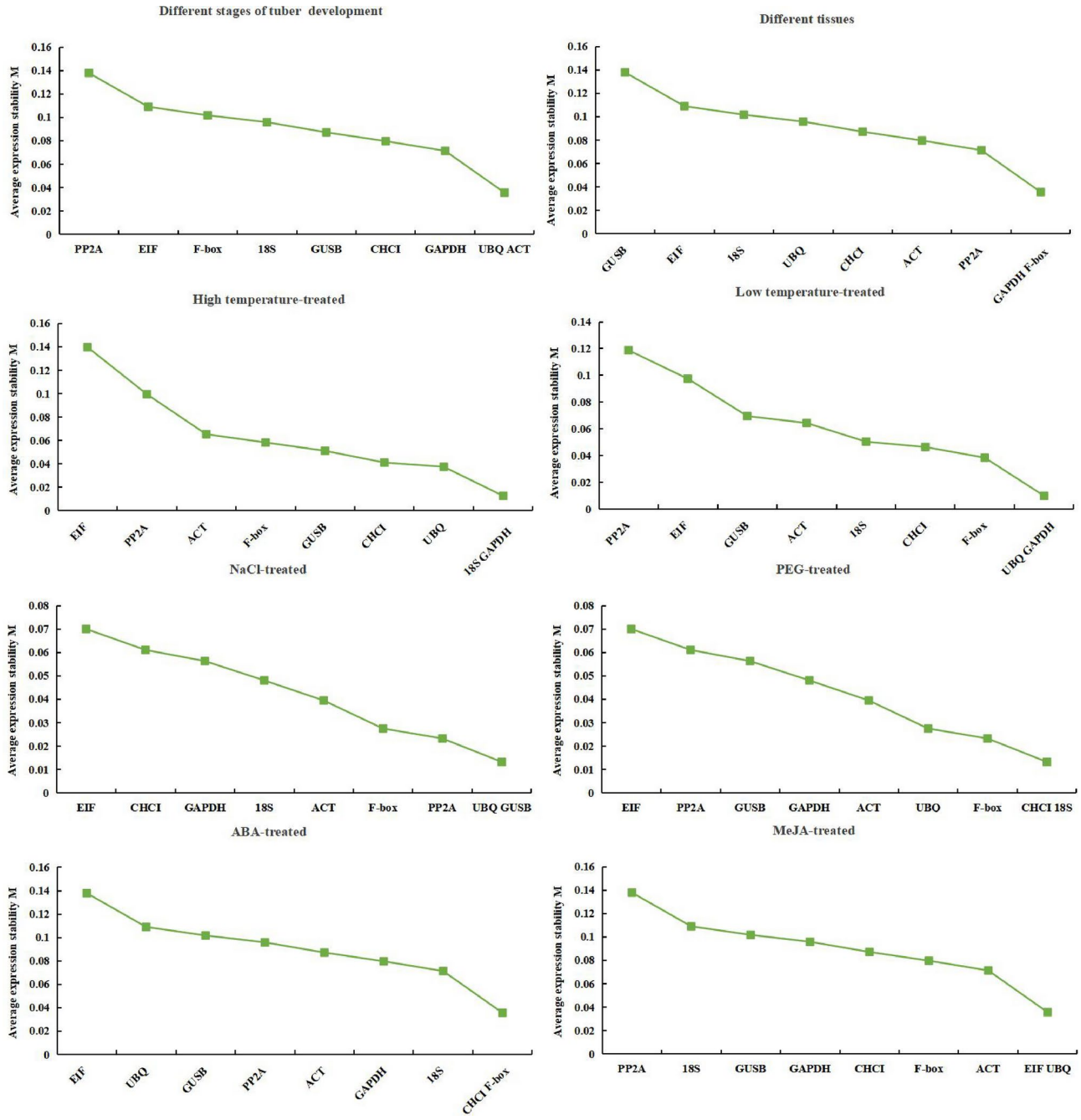

Fig. 5. Mean expression stability values (M) of candidate reference genes. Note: The most stable genes are listed on the right and the most unstable genes are listed on the left.

M values for *UBQ* and *ACT* indicated the highest stability, while the highest M of *PP2A* indicated the lowest stability. In different tissues, *GAPDH* and *F-box* exhibited the most stable expression, while *GUSB* showed the greatest variability. For High-temperature treatment, *GAPDH* and *18S* were the most stable reference genes. In Low-temperature treated samples, the combination of *GAPDH* and *UBQ* was the most suitable, while *PP2A* was the least stable. *F-box* was identified as the most stable endogenous reference gene in ABA treatment and tissues. On the other hand, *EIF* ranked the lowest in expression stability across high-temperature, NaCl, PEG, and ABA treatments, making it the most unstable reference gene. Thus, *EIF* is not suitable as an internal reference gene for gene expression analysis.

Additionally, the paired variant V value of the standardized factor can be calculated by the software after introducing a new reference gene. geNorm determines the optimal number of candidate reference genes by calculating the ratio $Vn/n+1 = 0.15$ to achieve more accurate and reliable results. If $Vn/Vn+1 < 0.15$, then the ideal number of reference genes is n. Conversely, if $Vn/Vn+1 > 0.15$, then the number should be $n+1$. As shown

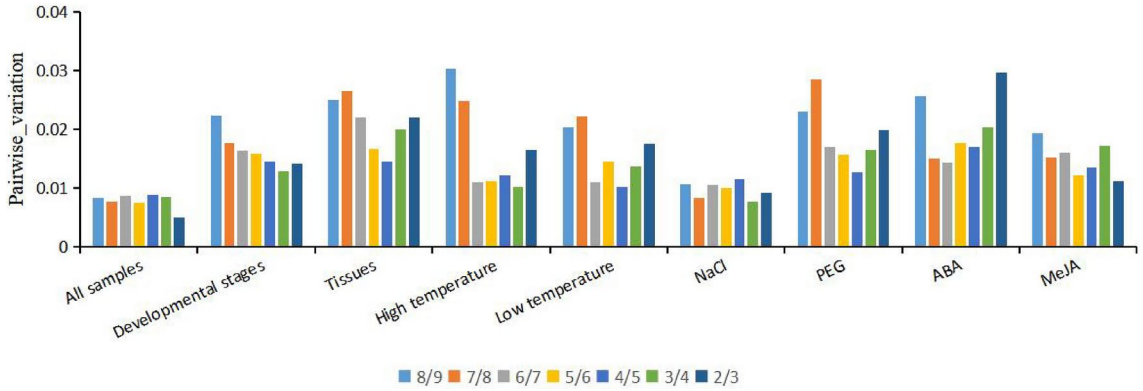

Fig. 6. Pairwise variation (V) analysis of candidate reference genes.

	All samples		Tuber at different developmental stages		Tissues		High temperature-treated		Low temperature-treated		NaCl-treated		PEG-treated		ABA		MeJA	
Rank	Gene	SV	Gene	SV	Gene	SV	Gene	SV	Gene	SV	Gene	SV	Gene	SV	Gene	SV	Gene	SV
1	*F-box*	0.20	*ACT*	0.02	*CHCI*	0.09	*GAPDH*	0.01	*F-box*	0.01	*GUSB*	0.02	*UBQ*	0.02	*ACT*	0.02	*CHCI*	0.01
2	*CHCI*	0.23	*UBQ*	0.02	*GAPDH*	0.11	*18S*	0.01	*CHCI*	0.01	*F-box*	0.02	*GUSB*	0.04	*PP2A*	0.05	*ACT*	0.02
3	*GAPDH*	0.26	*GAPDH*	0.02	*ACT*	0.13	*UBQ*	0.04	*18S*	0.04	*PP2A*	0.03	*F-box*	0.04	*GUSB*	0.06	*UBQ*	0.04
4	*ACT*	0.26	*CHCI*	0.06	*18S*	0.13	*ACT*	0.05	*ACT*	0.06	*UBQ*	0.04	*ACT*	0.04	*CHCI*	0.07	*F-box*	0.04
5	*18S*	0.28	*F-box*	0.09	*PP2A*	0.14	*CHCI*	0.05	*GUSB*	0.07	*18S*	0.05	*18S*	0.10	*GAPDH*	0.08	*GAPDH*	0.05
6	*UBQ*	0.30	*GUSB*	0.09	*F-box*	0.14	*GUSB*	0.06	*GAPDH*	0.07	*ACT*	0.06	*CHCI*	0.10	*F-box*	0.10	*EIF*	0.07
7	*GUSB*	0.38	*EIF*	0.11	*EIF*	0.20	*F-box*	0.10	*UBQ*	0.08	*CHCI*	0.06	*GAPDH*	0.15	*UBQ*	0.11	*GUSB*	0.11
8	*PP2A*	0.39	*18S*	0.14	*UBQ*	0.24	*PP2A*	0.20	*EIF*	0.18	*GAPDH*	0.07	*PP2A*	0.20	*18S*	0.11	*18S*	0.14
9	*EIF*	0.44	*PP2A*	0.20	*GUSB*	0.29	*EIF*	0.27	*PP2A*	0.18	*EIF*	0.09	*EIF*	0.20	*EIF*	0.23	*PP2A*	0.17

Table 3. Stable values of nine candidate reference genes in different treatments and tissues by NormFinder.

in Fig. 6, the Vn/Vn + 1 values for different treatments were all less than 0.15, indicating that two reference genes are sufficient for accurate normalization of target gene expression.

NormFinder analysis

NormFinder evaluates the stability of candidate genes by calculating an expression stability value (SV), where a lower SV indicates greater stability (Table 3). In the different developmental stages of the tuber, *ACT* and *UBQ* exhibited the most stable expression, while *PP2A* was the most unstable. *CHCI* showed the most stable expression in different tissues, while *GUSB* exhibited the greatest instability. Under High-temperature treatment, *GAPDH* and 18S were the most stable gene, while *EIF* was the most unstable. In Low temperature-treated samples, *F-box* was the most stable, whereas *EIF* and *PP2A* were the least stable. In NaCl-treated samples, *GUSB* and *F-box* were the most stable, while *EIF* was the least stable. In PEG-treated samples, *UBQ* was stable, while *EIF* again showed the least stability. The NormFinder analysis of all samples revealed that *EIF* was the least stable across conditions. These results indicated that NormFinder and geNorm produced almost identical stability rankings for the candidate genes, although some minor differences were noted.

BestKeeper analysis

In the BestKeeper analysis, candidate reference genes were ranked based on the standard deviations (SD) of the original Ct values under various coefficient of variation (CV) settings. Lower SD and CV values indicate better stability of reference genes. As shown in Fig. 7, *UBQ* exhibited the highest stability across different developmental stages, tissues, and under High-temperature conditions. *ACT* showed the highest stability under cold stress, while *18S* was relatively stable under PEG stress. Under all treatments, the coefficient of variation for *EIF* was greater than 1, indicating a large degree of variation. These results demonstrate that the stability of gene transcripts varies considerably under different experimental conditions, highlighting the need for careful selection of reference genes.

ReFinder analysis

To address discrepancies in the results from the four analytical methods, the RefFinder online tool was used for comprehensive ranking. RefFinder integrates the outputs of multiple algorithms to identify the optimal reference genes under different experimental conditions. As shown in Table 4, RefFinder identified *UBQ* as the

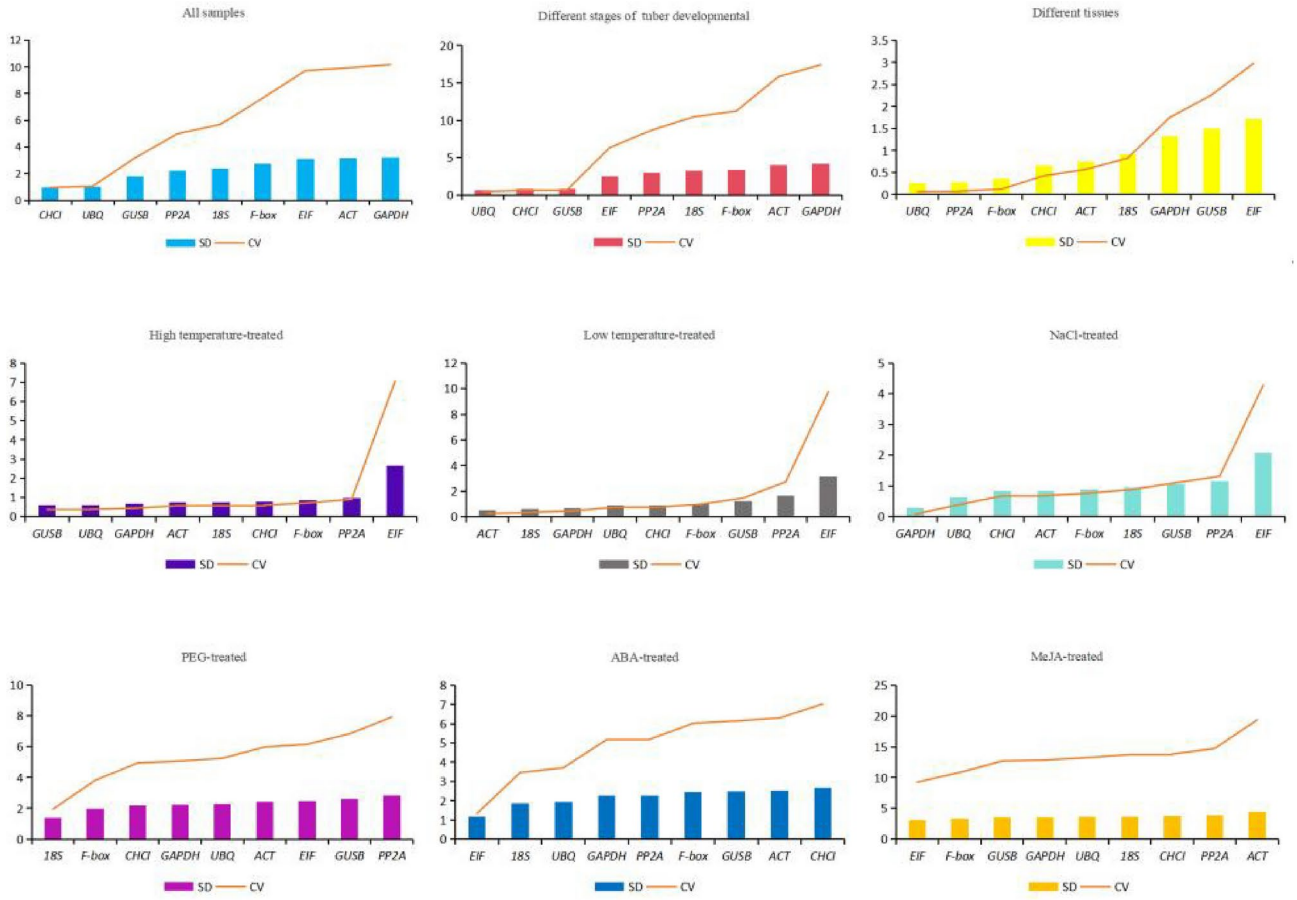

Fig. 7. Stability analysis of nine reference genes based on BestKeeper.

	Total		Different stages of tuber development		Different tissues		Different treatments	
Rank	Gene	Rank	Gene	Rank	Gene	Rank	Gene	Rank
1	*UBQ*	1.78	*PP2A*	2.11	*PP2A*	1.68	*UBQ*	1.32
2	*CHCI*	2.71	*CHCI*	2.63	*UBQ*	1.86	*CHCI*	2.06
3	*F-box*	3.31	*UBQ*	3.25	*F-box*	3.08	*F-box*	3.72
4	*ACT*	3.44	*18S*	3.71	*ACT*	3.31	*GAPDH*	3.94
5	*GAPDH*	3.83	*ACT*	4.12	*CHCI*	3.46	*ACT*	4.33
6	*18S*	4.09	*GAPDH*	4.90	*18S*	5.48	*PP2A*	4.76
7	*PP2A*	4.92	*F-box*	5.23	*GAPDH*	7.00	*18S*	5.73
8	*GUSB*	8.00	*GUSB*	5.86	*GUSB*	8.00	*GUSB*	7.24
9	*EIF*	9.00	*EIF*	7.35	*EIF*	9.00	*EIF*	9.00

Table 4. Comprehensive stability analysis of reference genes based on RefFinder in different experiment conditions.

most stable reference gene across all treatments. The stability ranking for the nine candidate reference genes was as follows: *UBQ* > *CHCI* > *F-box* > *ACT* > *GAPDH* > *18S* > *PP2A* > *GUSB* > *EIF*. For different developmental stages and tissues, *PP2A* emerged as the optimal reference gene, with *UBQ* consistently ranking in the top three. Under various abiotic stress treatments, *UBQ* was the most stable gene. In contrast, *EIF* was consistently identified as the least stable gene under all conditions.

Stability verification of candidate reference genes

To validate the reliability of the selected reference genes, the expression of *WRKY* was analyzed using the identified stable and unstable reference genes. *WRKY* is a key plant-specific transcription factor families that regulates various aspects of plant growth, development, physiological processes, and responses to abiotic

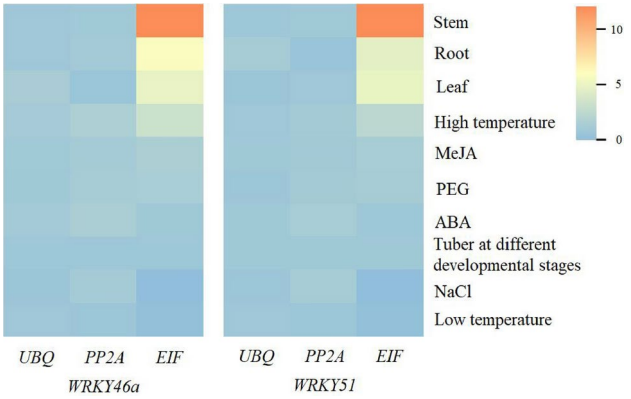

Fig. 8. Stability verification of candidate reference genes with *WRKYs*.

stresses. The stable reference genes *UBQ* and *PP2A*, as well as the unstable reference gene *EIF*, were used to measure the expression levels of *WRKY46a* and *WRKY51* in yam under different developmental stages, across various tissues, and during abiotic stresses (Fig. 8). Results showed consistent expression levels of *WRKY* when stable reference genes (*UBQ* and *PP2A*) were used. In contrast, the expression levels varied significantly when the unstable reference gene (*EIF*) was used, underscoring its unreliability. These findings confirm the stability and accuracy of the selected reference genes, validating the robustness of this study's experimental results.

Discussion

Identifying stable reference genes across different stages of plant development, tissues, and abiotic stress conditions is a critical prerequisite for studying gene functions and regulatory mechanisms in plant growth and development. With advancements in transcriptome data analysis, it is now possible to screen candidate reference genes systematically[40]. q-PCR is regarded as an effective tool for molecular biology research[41]. It is instrumental in evaluating gene expression levels and elucidating plant responses to various stressors. Reference genes serve as the cornerstone for accurate normalization in qPCR, enabling researchers to understand biosynthetic pathways and regulatory mechanisms. Over the years, many suitable reference genes have been identified in both animal and plant species[42]. However, there are currently no documented reference genes for *Dioscorea opposita* (commonly known as Chinese yam). Among its varieties, Dahechangyu yam is a widely cultivated type in both northern and southern China. Identifying and validating stable reference genes for yam will provide a robust foundation for further studies on gene expression regulation in this species.

It is reported that the optimal reference genes can vary depending on the species and specific conditions[43]. Among the nine reference genes analyzed, *EIF* exhibited the highest mean Ct value, while *18S* had the lowest. Since lower Ct values indicate higher gene expression levels, *18S* was expressed at a significantly higher level compared to the other genes. The smallest variation in gene expression was observed for *CHCI* (3.10) and *UBQ* (3.34), whereas *EIF* (11.41) displayed the highest variability across all samples. In order to explore the changes in gene expression levels in yam, geNorm, NormFinder and BestKeeper were used to evaluate the stability of reference gene transcripts[28]. The results demonstrated that there were some differences in the stability of the reference genes across the different software. The reason for this may be due to the differences in algorithms between each software[44] and a similar trend was seen in reports from other species, such as in *Rubus*[45] and *Gleditsia microphylla*[46]. In this study, geNorm ranked *UBQ* as the most stable reference gene across different developmental stages, as well as under Low-temperature, salt, and MeJA treatments. Conversely, *EIF* consistently ranked as the least stable gene in seven of the eight treatments, making it the most unreliable choice as an internal reference gene for gene expression analysis.

To enhance the detection accuracy of target genes and address the variability among different software tools, Vandesompele[31] recommends using the V value (Pairwise variation value) generated from geNorm as the threshold. This approach minimizes errors by incorporating multiple reference genes. The geNorm program typically sets the threshold at $Vn/n+1 = 0.15$. While 0.15 is considered an ideal value, it is influenced by factors such as the number of genes analyzed and the type of test samples. Consequently, significant differences in the accuracy of screening results have been observed[47]. According to both NormFinder and BestKeeper, *EIF* was identified as the most unstable reference genes, though the rankings for the most stable genes differed between the two programs.

To minimize errors caused by differences in software analysis and human selection, integrated analysis was used to evaluate the results from different software tools. This approach provided a means to better reflect the stability of each reference gene and reduce discrepancies in software evaluations[48]. *UBQ*, *F-box*, and *ACT* were most frequently ranked as the best genes for normalization by the four programs, with *UBQ* occurring 17 times, *F-box 8* times, and *ACT* 7 times. In contrast, *EIF* was identified as the least stable, being ranked as the most unstable 22 times. Under most experimental conditions, *UBQ* displayed high expression stability, consistent with findings from an earlier study on the ephemeral plant *Arabidopsis pumila*[49]. The *UBQ* gene belongs to the ubiquitin gene family, which plays a role in several essential biological processes, including cell cycle regulation, growth and apoptosis, signal transduction, and immune responses. Due to its high sequence homology and

conserved nature, *UBQ* has been used as a reference gene in rice[50] and black fungus[51]. In contrast, *EIF* was the most unstable gene and unsuitable as an internal reference for gene expression determination, a finding consistent with similar studies in eggplant[52]. The stability evaluation in this study revealed that no single reference gene was suitable for all experimental conditions. The most stable reference genes varied across groups and even between organs within the same group. Therefore, we recommend selecting the most appropriate reference genes for specific conditions rather than relying on a universal reference gene for normalization, despite the additional effort and time required.

WRKY transcription factors (TFs) play a critical regulatory role in plant stress signaling by controlling the expression of various stress-related genes, thereby enhancing plant stress resistance[53]. At the same time, it also plays an important regulatory role in plant growth and development[54]. To verify the accuracy and reliability of the selected reference genes, a normalization procedure was performed on the relative expression levels of the target genes *WRKY46a* and *WRKY51* under different experimental conditions. As shown in Fig. 8, the expression patterns of the target genes differed significantly depending on whether stable or unstable reference genes were used for normalization. When stable reference genes were used either individually or in combination there was minimal variation in expression levels. In contrast, the use of unstable reference genes resulted in pronounced changes in the expression patterns, highlighting the importance of selecting appropriate reference genes for accurate normalization. Meanwhile, the unstable internal reference genes changed significantly. Although the expression trends observed with individual stable reference genes were consistent with those derived from combinations of stable reference genes, slight differences in expression levels were still evident. In contrast, the use of inappropriate reference genes resulted in entirely inconsistent expression profiles compared to those obtained with stable reference genes, with significant changes observed. This highlights the substantial impact of using unstable reference genes on qRT-PCR results and underscores that the relative expression levels of target genes depend on both the selection and number of reference genes used. Similar findings have been reported in various species, including *Codonopsis pilosula*[55], *Scutellaria baicalensis*[56], and *Schima superba*[57].

Conclusion

In this study, we systematically identified and screened ideal reference genes for qRT-PCR analysis of target gene expression in the *Dioscorea* genus under diverse experimental conditions. Our findings revealed that two consistently stable genes, *UBQ* and *PP2A*, are suitable for accurately normalizing target gene expression in yam across all tested conditions. *EIF* is unsuitable for accurately normalizing internal reference gene. Furthermore, these genes were successfully employed for gene expression analysis in the study of WRKY transcription factors. These results not only provide a solid theoretical basis for functional genomics research but also establish a foundation for exploring the molecular mechanisms underlying biological functions in the *Dioscorea* genus.

Data availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files], transcriptome data from yam was collected from our laboratory previous studies53, and the genes information used is in the supplementary file.

Received: 30 October 2024; Accepted: 26 February 2025
Published online: 06 June 2025

References

1. Zhou, Y. et al. Integrated mRNA and miRNA transcriptome analysis reveals a regulatory network for tuber expansion in Chinese yam (*Dioscorea opposita*). *BMC Genomics* **21**(1), 117 (2020).
2. Epping, J. & Laibach, N. An underutilized orphan tuber crop-Chinese yam: A review. *Planta* **252**(4), 58 (2020).
3. Zhang, Z., Wang, X., Liu, C. & Li, J. The degradation, antioxidant and antimutagenic activity of the mucilage polysaccharide from *Dioscorea opposita*. *Carbohydr. Polym.* **5**(150), 227–231 (2016).
4. Zhao, L. et al. Overexpression of DoBAM1 from Yam (*Dioscorea opposita* Thunb.) Enhances Cold Tolerance in Transgenic Tobacco. *Genes (Basel)* **13**(12), 2296 (2022).
5. Amat, N. et al. Aqueous extract of *dioscorea opposita* thumb. normalizes the hypertension in 2K1C hypertensive rats. *BMC Complement. Altern. Med.* **21**(14), 36 (2014).
6. Conti, M. V. et al. Potential role of neglected and underutilized plant species in improving women's empowerment and nutrition in areas of sub-Saharan Africa. *Nutr. Rev.* **77**(11), 817–828 (2019).
7. Wei, B. H. Some opinions on research and utilization of Huai Yam in China. *China Rural Sci. Technol.* **04**, 44–47 (2012).
8. Wong, M. L. & Medrano, J. F. Real-time PCR for mRNA quantitation. *Biotechniques* **39**(1), 75–85 (2005).
9. Bustin, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. *J. Mol. Endocrinol.* **29**(1), 23–39 (2002).
10. Shi, L. et al. Reference genes selection and validation for *Cinnamomum burmanni* by real-time quantitative polymerase chain reaction. *Int. J. Mol. Sci.* **25**(6), 3500 (2024).
11. Delporte, M., Legrand, G., Hilbert, J. L. & Gagneul, D. Selection and validation of reference genes for quantitative real-time PCR analysis of gene expression in *Cichorium intybus*. *Front. Plant Sci.* **18**(6), 651 (2015).
12. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; Strategies and considerations. *Genes Immun.* **6**(4), 279–284 (2005).
13. Kozera, B. & Rapacz, M. Reference genes in real-time PCR. *J. Appl. Genet.* **54**(4), 391–406 (2013).
14. Deng, Y., Li, Y. & Sun, H. Selection of reference genes for RT-qPCR normalization in blueberry (*Vaccinium corymbosum* × *angustifolium*) under various abiotic stresses. *FEBS Open Bio* **10**(8), 1418–1435 (2020).
15. Liu, Q. et al. Reference gene selection for quantitative real-time reverse-transcriptase PCR in annual ryegrass (*Lolium multiflorum*) subjected to various abiotic stresses. *Molecules* **23**(1), 172 (2018).
16. Yi, S. et al. Selection of suitable reference genes for gene expression normalization studies in *Dendrobium huoshanense*. *Genes (Basel)* **13**(8), 1486 (2022).
17. Zhang, P. et al. Selection and validation of qRT-PCR internal reference genes to study flower color formation in *Camellia impressinervis*. *Int. J. Mol. Sci.* **25**(5), 3029 (2024).

18. Sun, H. et al. Identification and validation of stable reference genes for RT-qPCR analyses of *Kobresia littledalei* seedlings. *BMC Plant Biol.* **24**(1), 389 (2024).
19. Li, T. et al. Selection and validation of appropriate reference genes for qRT-PCR analysis in *Isatis indigotica Fort*. *Front. Plant Sci.* **28**(8), 1139 (2017).
20. Thomas, C. et al. Molecular characterization and spatial expression of the sunflower *ABP1* gene. *Plant Mol. Biol.* **52**(5), 1025–1036 (2003).
21. Sikand, K., Singh, J., Ebron, J. S. & Shukla, G. C. Housekeeping gene selection advisory: Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β-actin are targets of miR-644a. *PLoS One* **7**(10), e47510 (2012).
22. Lian, C. et al. Validation of suitable reference genes by various algorithms for gene expression analysis in *Isodon rubescens* under different abiotic stresses. *Sci. Rep.* **12**(1), 19599 (2022).
23. Tajti, J., Pál, M. & Janda, T. Validation of reference genes for studying different abiotic stresses in oat (*Avena sativa* L.) by RT-qPCR. *Plants (Basel)* **10**(7), 1272 (2021).
24. 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. WorldJ.* **2014**, 532713 (2014).
25. Fei, X., Shi, Q., Yang, T., Fei, Z. & Wei, A. Expression stabilities of ten candidate reference genes for RT-qPCR in *Zanthoxylum bungeanum* Maxim. *Molecules* **23**(4), 802 (2018).
26. Park, S. C. et al. Stable internal reference genes for the normalization of real-time PCR in different sweetpotato cultivars subjected to abiotic stress conditions. *PLoS One.* **7**(12), e51502. https://doi.org/10.1371/journal.pone.0051502 (2012).
27. Jin, G. et al. Screening and validation of optimal miRNA reference genes in different developing stages and tissues of *Lilium henryi* Baker. *Sci. Rep.* **14**(1), 1545 (2024).
28. Chen, M. et al. Screening and validation of appropriate reference genes for real-time quantitative PCR under PEG, NaCl and $ZnSO_4$ treatments in *Broussonetia papyrifera*. *Int. J. Mol. Sci.* **24**(20), 15087 (2023).
29. Wang, Q., Guo, C., Yang, S., Zhong, Q. & Tian, J. Screening and verification of reference genes for analysis of gene expression in garlic (*Allium sativum* L.) under cold and drought stress. *Plants (Basel)* **12**(4), 763 (2023).
30. Silver, N., Best, S., Jiang, J. & Thein, S. L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol. Biol.* **6**(7), 33 (2006).
31. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**(7), RESEARCH0034 (2002).
32. 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**(15), 5245–5250 (2004).
33. 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**(6), 509–515 (2004).
34. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* **80**, 75–84 (2012).
35. Shao, Y. et al. Enhanced carbohydrate metabolism and salinity tolerance of tobacco plants overexpressing a vacuolar H+-ATPase subunit B2 (DoVHAb2) gene from Yam (*Dioscorea opposita* Thunb.). *Journal of Biobased Materials and Bioenergy* **15**(4), 504–513 (2021).
36. Jain, M., Nijhawan, A., Tyagi, A. K. & Khurana, J. P. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* **345**(2), 646–651 (2006).
37. Nicot, N., Hausman, J. F., Hoffmann, L. & Evers, D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J. Exp. Bot.* **56**(421), 2907–2914 (2005).
38. Remans, T. et al. Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. *Planta* **227**(6), 1343–1349 (2008).
39. Zhao, X. et al. Identification and validation of reference genes for qRT-PCR studies of gene expression in *Dioscorea opposita*. *Biomed. Res. Int.* **2016**(1), 3089584 (2016).
40. Lu, L. et al. Selection and verification of reliable internal reference genes in stem development of herbaceous peony (*Paeonia lactiflora* Pall.). *Physiol. Mol. Biol. Plants* **29**(6), 773–782 (2023).
41. Zheng, W. et al. Identification of key modules and candidate genes associated with endometriosis based on transcriptome data via bioinformatics analysis. *Pathol. Res. Pract.* **244**, 154404 (2023).
42. Hu, A. et al. Selection and validation of appropriate reference genes for RT-qPCR analysis of *Nitraria sibirica* under various abiotic stresses. *BMC Plant Biol.* **22**(1), 592 (2022).
43. Rebouças, E. D. L. et al. Real time PCR and importance of housekeeping genes for normalization and quantification of mRNA expression in different tissues. *Braz. Arch. Biol. Technol.* **56**, 143–154 (2013).
44. De Spiegelaere, W. et al. Reference gene validation for RT-qPCR, a note on different available software packages. *PLoS One* **10**(3), e0122515 (2015).
45. Wu, Y. et al. Selection and validation of candidate reference genes for gene expression analysis by RT-qPCR in *Rubus*. *Int. J. Mol. Sci.* **22**(19), 10533 (2021).
46. Yang, J. et al. Identification of reference genes for RT-qPCR analysis in *Gleditsia microphylla* under abiotic stress and hormone treatment. *Genes (Basel)* **13**(7), 1227 (2022).
47. 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.* **7**(8), 112 (2008).
48. Wang, W. et al. Systematic identification of reference genes for qRT-PCR of *Ardisia kteniophylla* A. DC under different experimental conditions and for anthocyanin-related genes studies. *Front. Plant Sci.* **14**, 1284007 (2023).
49. Jin, Y., Liu, F., Huang, W., Sun, Q. & Huang, X. Identification of reliable reference genes for qRT-PCR in the ephemeral plant *Arabidopsis pumila* based on full-length transcriptome data. *Sci. Rep.* **9**(1), 8408 (2019).
50. Auler, P. A. et al. Selection of candidate reference genes and validation for real-time PCR studies in rice plants exposed to low temperatures. *Genet. Mol. Res.* **16**(2), 104238 (2017).
51. Jia, D. H. et al. Validation of reference genes for quantitative gene expression analysis in *Auricularia cornea*. *J. Microbiol. Methods* **163**, 105658 (2019).
52. Bao, Y. et al. Overexpression of a Ramie (*Boehmaeria nivea* L. Gaud) Group I WRKY gene, BnWRKY49, increases drought resistance in *Arabidopsis thaliana*. *Plants (Basel)* **13**(3), 379 (2024).
53. Zhu, Z. et al. The WRKY gene family in the halophyte Limonium bicolor: Identification, expression analysis, and regulation of salt stress tolerance. *Plant Cell Rep.* **43**(7), 167 (2024).
54. Yang, J. et al. Selection of suitable reference genes for qRT-PCR expression analysis of *Codonopsis pilosula* under different experimental conditions. *Mol. Biol. Rep.* **47**(6), 4169–4181 (2020).
55. Wang, W. et al. Selection and evaluation of reference genes for qRT-PCR of *Scutellaria baicalensis Georgi* under different experimental conditions. *Mol. Biol. Rep.* **48**(2), 1115–1126 (2021).
56. Lin, H., Qiu, B. Y., Zhu, H. S. & Wen, Q. F. Screening and evaluation of reference genes in eggplant seedlings by real-time fluorescence quantitative PCR under low temperature and weak light stress. *Fujian Agric. Sci. Technol.* **55**(08), 1–15 (2024).
57. Yao, J. et al. Reference gene selection for qPCR analysis in *Schima superba* under abiotic stress. *Genes (Basel)* **13**(10), 1887 (2022).

58. Gao, Y. et al. Transcriptome analysis reveals the accumulation mechanism of starch in chinese yam (*Dioscorea opposite* Thunb.) tubers using RNA-Seq and Iso-Seq. *Sci. Adv. Mater.* **14**(2), 305–318 (2022).

Acknowledgements
Thank you for the funding from the National Natural Science Foundation, Grant number 32260759.

Author contributions
L N X and Y F Z conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft. X W H conceived and designed the experiments,authored or reviewed drafts of the paper, and approved the final draft. M R G, Y P X and H Q M analyzed the data, prepared figures and/or tables.

Funding
This research was funded by the National Natural Science Foundation of China, Grant number 32260759.

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/s41598-025-92244-w.

Correspondence and requests for materials should be addressed to X.H.

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.