BMC Genomics

RESEARCH

Open Access

Identification and validation of stable reference genes via transcriptome analysis for studying sex differentiation in tung tree (*Vernicia fordii*)

Jie Cao[1,2], Wenying Li[3], Meilan Liu[1,2], Junjie Chen[1,2], Chong Ge[1,2], Lichuan Xiao[1,2] and Lin Zhang[1,2*]

Abstract

Tung oil, obtained from the seeds of the tung tree (*Vernicia fordii)*, holds significant economic value. Nevertheless, the low proportion of female flowers in tung trees limits tung oil yield. Studying the expression of genes involved in sex differentiation provides a promising strategy for increasing the female flower ratio through molecular breeding. Accurate normalization is critical for stable gene expression quantification in qRT-PCR. This study identified stable reference genes for sex differentiation research in the tung tree. Using transcriptome data from diverse floral samples of tung tree, we identified 10 candidate genes and subsequently assessed their expression stability across various tissues using multiple algorithms. Among them, *ERF3A* and *OLA1* exhibited the highest stability in floral tissues. Validation experiments demonstrated that using these stable genes for normalization yielded consistent expression patterns for target genes *AP3* and *SEP2-2*, whereas unstable references led to significant misinterpretation. Thus, *ERF3A* and *OLA1* are proposed for use as reference genes in qRT-PCR when investigating sex differentiation in the tung tree. This study lays a solid foundation for elucidating the molecular mechanisms of sex differentiation in the tung tree.

Keywords Tung tree, Sex differentiation, qRT-PCR, Reference gene

Introduction

Studying gene expression is essential for deciphering the complex regulatory networks governing gene expression in animals, plants, and microorganisms. It also facilitates the discovery and functional prediction of novel genes. Gene expression can be quantified by several established techniques, such as quantitative real-time PCR (qRT-PCR), northern blotting, and microarrays [1]. Among these, qRT-PCR has become widely adopted because of its high reproducibility, accuracy, and throughput [2]. However, reliable application of qRT-PCR requires strict quality control. This includes assessment of RNA integrity, primer validation, verification of cDNA quality and initial concentration, as well as the choice of reference

*Correspondence:
Lin Zhang
triwoodtim918@126.com
[1]State Key Laboratory of Woody Oil Resources Utilization, Central South University of Forestry and Technology, Changsha 410000, China
[2]Yuelushan Laboratory Non-wood Forests Variety Innovation Center, Central South University of Forestry and Technology, Changsha 410000, China
[3]College of Biology and Agricultural Resources, Huanggang Normal University, Huanggang, Hubei 438000, China

genes [3]. The selection of appropriate reference genes is fundamental to reliable qRT-PCR, since it governs the normalization and interpretation of gene expression data.

Reference genes, also termed housekeeping genes, are stably expressed across tissues and organs under defined conditions, with minimal influence from internal or external factors. Among the most commonly employed reference genes in plant research are *elongation factor 1 alpha* (*EF1α*), *actin* (*ACT*), *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*), and *α/β-tubulin* (*TUA*/*TUB*) [4–6]. Although traditionally assumed to show consistent expression across diverse cell types and physiological states, accumulating evidence indicates that these genes are not universally stable and can exhibit considerable plasticity in response to experimental variables. Consequently, commonly used reference genes may display substantial variability in certain contexts, driving the search for more stable alternatives [7]. For instance, *ubiquitin-conjugating enzyme 18* (*UBC18*) shows high expression stability across various tissues of *Brachypodium distachyon*, yet its stability decreases under auxin treatment or stress [8]. Similarly, *ACT* expression remains relatively consistent throughout developmental stages in *Litchi chinensis*, but becomes variable across tissues under shading conditions [9]. These examples highlight the absence of a single reference gene that performs optimally under all experimental settings. Using transcriptome databases from *Camellia oleifera*, Chang et al. screened for stable reference genes for pollen tubes [10]. When compared to traditional reference genes, the newly identified candidates demonstrated superior stability, proving more suitable for data normalization in *C. oleifera* pollen tubes. This underscores the critical need for reference genes specific to species, tissues, and experimental contexts of interest. In conclusion, the identification and validation of suitable reference genes are critical for accurate expression profiling. Proper normalization using validated reference genes under controlled conditions ensures reliable and reproducible results, thereby advancing our understanding of gene regulatory networks in plant research.

The tung tree (*Vernicia fordii*) and mu oil tree (*V. montana*) are two representative deciduous tree species within the genus *Vernicia* (Euphorbiaceae), both cultivated as economically important woody oil crops [11]. Relative to *V. montana* kernels, those of *V. fordii* exhibit a remarkably high oil content (up to 60%). This characteristic makes it a valuable source of tung oil, which is utilized in manufacturing a range of industrial products, from paints and coatings to biodiesel [12–13]. Additionally, its distinct flower morphology and vibrant colors confer significant ornamental value to this species [14]. Recent studies of *V. fordii* have focused on two key areas: (1)namely the biological characteristics and molecular mechanisms underlying female flower development, and (2) the biosynthetic pathways of tung oil. For example, Liu et al. performed transcriptomic analyses to elucidate the regulatory mechanisms of female flower development [15]. In a follow-up study, Liu et al. characterized a *HAD* (haloacid dehalogenase) gene involved in determining the specific fatty acid composition of *V. fordii* oil [16]. Furthermore, Zhang et al. conducted a genome-wide analysis to uncover the genetic basis for the exceptionally high seed oil content in this species [17]. However, the productivity of *V. fordii* is often constrained by an imbalance of female to male flowers, resulting in suboptimal fruit set. Therefore, a deeper understanding of the mechanisms controlling sex differentiation is crucial for enhancing yield [15]. As studies on sex differentiation in *V. fordii* progress, gene expression analysis will be increasingly important for deciphering the associated regulatory networks. A major challenge in this endeavor is the current lack of validated reference genes suited for studying sex differentiation in *V. fordii*. The identification and evaluation of such condition-specific reference genes are thus essential for ensuring data integrity in gene expression studies on this species.

To select candidate reference genes, we analyzed transcriptomic data from *V. fordii* under three experimental conditions: (i) 6-BA-treated male flowers [18], (ii) $AgNO_3$-treated mixed male and female flower buds [18], and (iii) male and female flowers under normal growth conditions [15]. We selected 10 candidate reference genes with high and stable expression (*OLA1*, *PNR12*, *EF1γ*, *ARD2BL*, *ERF3A*, *CWC15H*, *PAIP7*, *SKD1*, *PD2*, and *P24β3*), and subsequently analyzed their expression levels across 10 distinct tissue types of *V. fordii* via qRT-PCR. Using five algorithms (Delta-Ct, BestKeeper, geNorm, NormFinder, and RefFinder), we evaluated the candidate genes for stability across floral and non-floral organs and generated a comprehensive ranking by integrating the results. We aimed to establish a set of robust reference genes appropriate for studying sex differentiation in *V. fordii*, thereby providing a reliable basis for future research on gene regulatory mechanisms involved in this biologically and economically important trait.

Materials and methods

Plant materials and treatments

All plant materials were obtained from 13-year-old *V. fordii* 'Putaotong' trees grown at the National Tung Tree Germplasm Conservation Repository in Yongshun County, Hunan Province. Sampled tissues comprised roots, stems, leaves, fruits, and female and male flowers collected at 20 days before anthesis under standard growth conditions. Also included were male flower buds treated with either water or 640 mg/L 6-BA for 120 days, as well as mixed male and female flower buds treated

with water or $AgNO_3$ for 55 days (Table 1). For qRT-PCR analysis, three individual 13-year-old *V. fordii* 'Putaotong' trees were sampled as biological replicates for each tissue or treatment condition. Total RNA was extracted separately from each tree and analyzed independently. Following collection, samples were rinsed, snap-frozen in liquid nitrogen, and stored at − 80 °C until RNA extraction to ensure preservation of RNA integrity and suitability for subsequent molecular analyses.

Establishment of the RNA-seq database

Transcriptome data were derived from three sources (see Supplementary Table 1 for detailed sample and sequencing information): (1) 6-BA-treated male flowers of *V. fordii*; (2) male and female flowers under normal growth conditions; (3) $AgNO_3$-treated mixed male and female flower buds. The three datasets (accession numbers: SRX3843588, SRS3089151, SRS3089154, SRS3089148, SRS3089147, SRS3089150, and PRJNA1263041) were downloaded from the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/sra). The reference genome used was the *V. fordii* genome assembly version 2.0 (accessible at NCBI, BioProject: PRJNA503685, PRJNA445350, and PRJNA483508). All data were subjected to quality control using Fastp [19]. Raw reads were preprocessed by removing adapter sequences and low-quality bases (Q < 20) using Trimmomatic v0.39. Clean reads were aligned to the reference genome using HISAT2 v2.2.1 with default parameters [20]. Gene expression levels were quantified as fragments per kilobase of transcript per million mapped reads (FPKM) using StringTie v2.1.5 [21], which was also employed for novel transcript prediction. This analytical workflow enabled consistent and reproducible transcriptome profiling for downstream analyses.

Table 1 Information of samples from 10 treatments

Sample type	Samples names	Notes
Flowers	female	Normal growth of female flowers 20 days before flowering
	Male	Normal growth of Male flowers 20 days before flowering
	$AgNO_3$-Mixed	Mixed male and female flowers after 55 d of $AgNO_3$ treatment
	H_2O-Mixed	Mixed male and female flowers after 55 d of H_2O treatment
	6-BA-male	Male flowers after 120 d of 6-BA treatment
	H_2O-male	Male flowers after 120 d of H_2O treatment
Non-flowers	Roots	-
	Stems	-
	Leaves	-
	Fruits	-

Selection of candidate reference genes and primer design

Based on transcriptomic data from normal and chemically treated floral tissues of *V. fordii*, we constructed three independent datasets by selecting genes with stable expression according to criteria adapted from a previous study [10]: the FPKM ratio between any two samples within the same transcriptome ranged from 0.75 to 1.33, and the gene was present in at least two transcriptomes with an FPKM value greater than 50. To elucidate their functional relevance, these stably expressed genes were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Gene function was further annotated using the Swiss-Prot database. Through this integrated approach, we identified 10 candidate reference genes linked to sex differentiation in *V. fordii* (Table 2). These candidates provide a valuable resource for future studies aimed at characterizing gene function during sex differentiation in this species.

Total RNA extraction and cDNA synthesis

Frozen *V. fordii* tissue samples (approximately 100 mg) were ground to a powder in a liquid nitrogen-chilled mortar. Total RNA was extracted with the RNAprep Pure Plant Plus Kit (TIANGEN, Beijing, China) according to the manufacturer's protocol. RNA integrity, purity and concentration were evaluated. Integrity was verified by 1% agarose gel electrophoresis, while purity and concentration were assessed using a Nanodrop ND2000 spectrophotometer (Thermo, Waltham, USA) based on standard absorbance ratios (A_{260}/A_{280} and A_{260}/A_{230}). For cDNA synthesis, 1 μg total RNA was reverse-transcribed with the HiScript® II 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China), and the resulting cDNA was stored at − 80 °C.

qRT-PCR assay

qPCR was carried out with ChamQ Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China) on a CFX96™ Real-Time PCR Detection System (Bio-Rad, Hercules, USA). The PCR protocol was set as follows: 95 °C for 3 min, followed by 39 cycles of 95 °C for 10 s and 60 °C for 30 s. A melting curve analysis was then run using a temperature gradient (65–90 °C, 5 s per step) to verify amplification specificity. All reactions were run in accordance with standard qPCR guidelines to ensure reproducibility.

Evaluation and validation of reference gene stability

Following qRT-PCR amplification, candidate genes were assessed for expression stability during sex

Table 2 Primer sequences of ten candidate reference genes used for qRT-PCR

Gene name	Gene ID (V. fordii genome)	Description	Primer (5′-3′)	Product length (bp)
OLA1	Vf10G1658	Obg-like ATPase 1	F: GTTATTGGTGCTGAGTTACGCC R: CCTAGGCGGACATCTTTTCCAT	167 bp
PNR12	Vf01G1092	26 S proteasome non-ATPase regulatory subunit 12 homolog A	F: AAGAAGAAGCCCAAGGAAGG R: ATGGCCTTGTAGCAACGAC	155 bp
EF1γ	Vf03G0504	Elongation factor 1-gamma	F: TGGGTGAAATCAAGCAGACTG R: TTGGTGCTTCTTCCTCTTCCTC	171 bp
ARD2BL	Vf00G1767	Ankyrin repeat domain-containing protein 2B-like	F: ACTGTGCTAGTGTTGGTGATGT R: ACAAGGACTTGAGCACACTTC	169 bp
ERF3A	Vf08G0100	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A	F: GGTAAGTCTACAACTGGAGGCC R: ACCCTTAACCCTCTCCTCTTC	156 bp
CWC15H	Vf11G1193	Protein CWC15 homolog	F: GAGGAGGATGACACAGAAGCTC R: GTCATCCCACCTCCTTTTCAC	189 bp
PAIP7	Vf08G1399	Polyadenylate-binding protein-interacting protein 7	F: TTCTTCCATCCCGTCTAGAGGT R: TAAATGCCTCTCCCATTCCCAC	183 bp
SKD1	Vf02G0089	Suppressor of K(+) Transport Growth Defect 1	F: AGACGGAACTTCTTGTGCAGAT R: CCGAGCCTTCAAATCTGGTAG	155 bp
PD2	Vf00G1765	Programmed cell death protein 2	F: TTCATGTGCCCCTCAATGTCAT R: CTCCATCATGTCTAGGTGGCTC	154 bp
P24β3	Vf11G1531	Transmembrane emp24 domain-containing protein p24beta3	F: GGACGCATATCTTCTCTCTCCG R: GCAGGAGATGTCACAGTGAGAT	164 bp
AP3	Vf05G0040	APETALA3	F: AGCCCTACAACAACGACAAAG R: CCCATCCTCTGCCTTATGTCTC	155 bp
SEP2-2	Vf05G1262	SEPALLATA2-2	F: TTTGTAGCACCTCCAGCATGAT R: AAATCTTCCCCAAGGAGGTTCC	178 bp

Gene symbols (e.g., *OLA1*, *PNR12*) are assigned based on sequence homology to their respective orthologs in *Arabidopsis thaliana* or other well-annotated species

differentiation in *V. fordii* using four algorithms: Delta-CT [22], BestKeeper [23], geNorm [24], and NormFinder [25]. To integrate the outcomes from these methods, RefFinder [26, 27] was employed to compile a composite ranking, identifying the most stable reference genes for sex differentiation studies in *V. fordii*. Values represent the geometric mean of rankings from Delta-CT, geNorm, NormFinder, and BestKeeper, as calculated by RefFinder. Lower values indicate greater expression stability. Shapiro-Wilk normality tests were performed on the Ct values of each candidate gene in the floral and non-floral samples, with $\alpha = 0.05$ as the significance threshold for determining deviation from normality. The analysis was carried out using SPSS software (version 26.0, IBM Corp., Armonk, NY, USA).

Using the identified reference genes for normalization, expression of two focal genes, *VfAP3* and *VfSEP2-2*, was quantitatively analyzed across various stages of sex differentiation in *V. fordii*. The three biological replicates employed for qPCR validation were distinct from the samples used for transcriptome sequencing and reference gene stability evaluation. Quantification of target gene expression was based on the $2^{-\Delta\Delta CT}$ method, with each sample analyzed in triplicate to ensure reproducibility and statistical robustness.

Results

Screening of candidate reference genes using *V. fordii* transcriptome data

To identify reference genes specific to sex differentiation in *V. fordii*, we assembled transcriptome data from three floral samples sequenced by our research group and computed their FPKM values. Gene selection was performed using stringent criteria: the FPKM fold change between any two sample groups was required to be between 0.75 and 1.33, and the genes needed to be present in least two transcriptomes with an FPKM value greater than 50. After rigorous filtering to remove genes with unstable or low expression across the three transcriptomes, we retained 95, 155, and 130 genes from the respective datasets (Fig. 1a). KEGG enrichment analysis performed on each gene set showed a high degree of overlap, with most significantly enriched pathways shared among the three groups. In particular, genes associated with five key pathways (Proteasome, Endoplasmic Reticulum Protein Processing, Phagosome, Oxidative Phosphorylation, and Fatty Acid Metabolism) were maintained across all groups (Fig. 1b–d).

Through the selection of genes common to at least two of the three sample groups, we identified a total of 47 candidate genes, including one gene (*VF12102*) that was shared across all three groups (Fig. 1a). Hierarchical

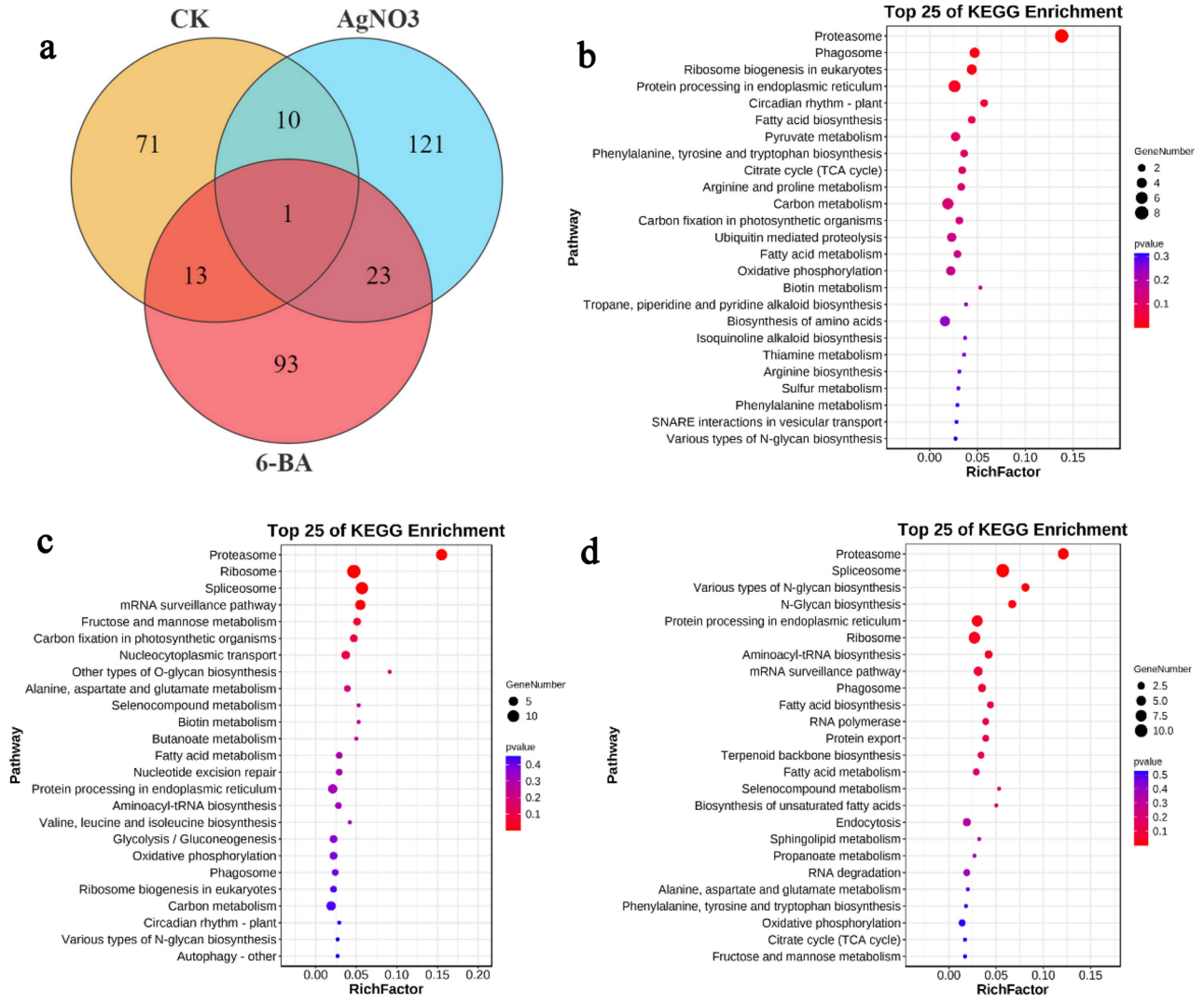

Fig. 1 Relationship and functional information of stable expression genes in male and female *V. fordii* flowers. **a** Venn diagram showing the relationships among the three stable expression gene datasets. (**b**–**d**) KEGG pathway enrichment analysis of stably expressed genes in untreated male and female floral tissues (**b**), $AgNO_3$-treated and water-treated mixed male and female floral tissues (**c**), and 6-BA-treated and water-treated male flower samples (**d**)

clustering based on the expression of these 47 genes identified a cluster of 11 genes exhibiting higher and more stable expression. Following further assessment of expression stability, *VF18742* was excluded due to its comparatively low stability, ultimately yielding 10 candidate genes for subsequent tissue-specific expression analysis in *V. fordii* (Fig. 2). These 10 genes were subsequently confirmed as reliable reference genes in studies on sex differentiation in *V. fordii*.

Expression profile of candidate reference genes

Primers for 10 candidate genes were designed (Table 2) and employed in qRT-PCR with cDNA from ten different *V. fordii* samples to accurately quantify their expression levels. The specificity and reproducibility of amplification were confirmed by single-peak melt profiles for each gene and consistent results across replicates, indicating no non-specific products or primer-dimers (Supplementary Fig. 1). These results support the suitability of the primers for subsequent qRT-PCR assays.

Expression stability of the 10 candidate genes was evaluated across non-floral tissues (roots, stems, leaves, and fruits) and floral tissues (female flowers, male flowers, $AgNO_3$-treated mixed flower buds, water-treated mixed flower buds, 6-BA-treated male flowers, and water-treated male flowers) by analyzing cycle threshold (Ct) values. Mean Ct values ranged from 23.33 to 25.55 in non-floral samples, with coefficients of variation (CV) between 0.82% and 3.25%. In floral samples, mean Ct values ranged from 23.14 to 24.96, with CVs between 1.19% and 3.32% (Fig. 3). Together, these findings indicate the consistent expression of the candidate reference genes across diverse tissue types, supporting their suitability as reliable internal controls for transcriptional analyses in *V. fordii*.

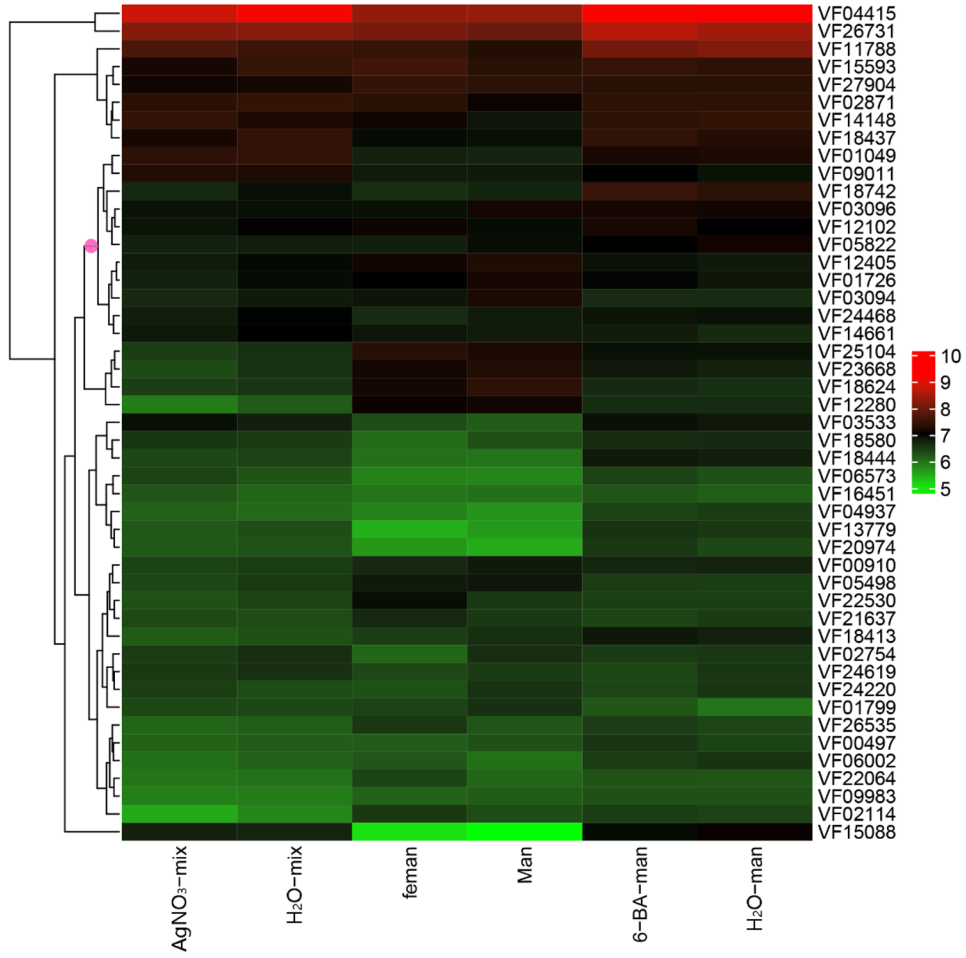

Fig. 2 Hierarchical clustering analysis of stable expression reference genes in male and female *V. fordii* flowers. In the figure, higher values represented by different colors indicate higher expression levels

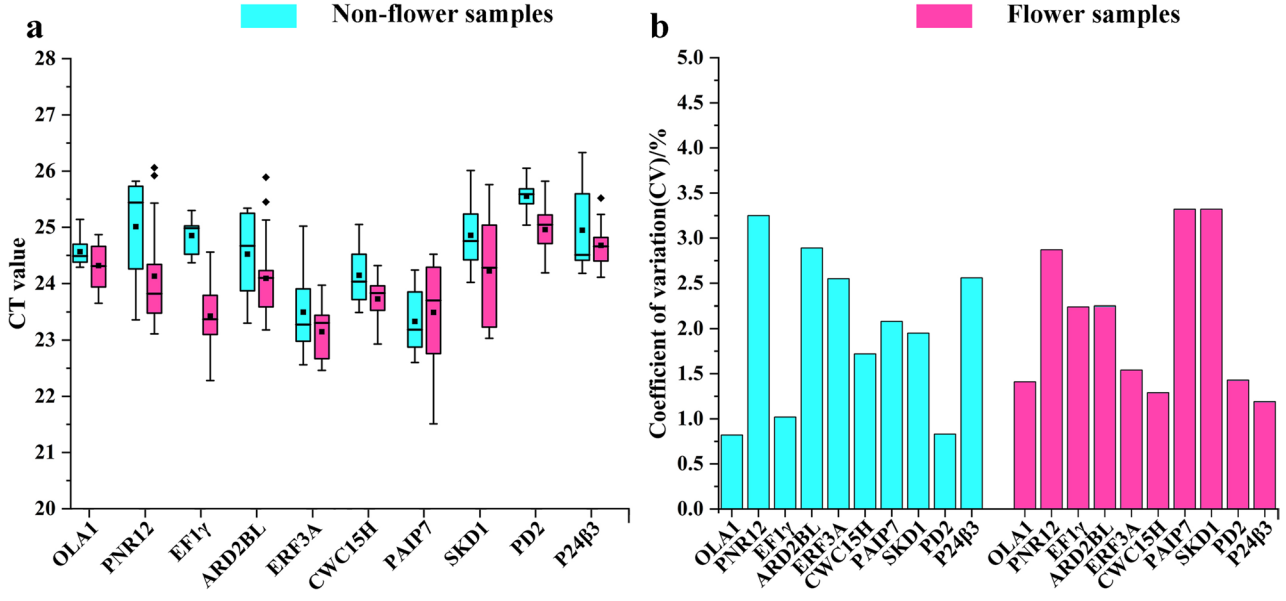

Fig. 3 Distribution of Ct values and coefficient of variation for for all candidates in 10 *V. fordii* samples. **a** Distribution of Ct values for all candidates in 10 *V. fordii* samples. **b** Coefficient of variation for all candidates in 10 *V. fordii* samples

Expression stability of candidate genes

The stability of expression for the 10 candidate reference genes was evaluated across non-floral samples, floral samples, and a combined dataset using multiple algorithms: Delta-CT, BestKeeper, NormFinder, geNorm, and RefFinder (Table 3; Fig. 4).

In non-floral samples, Delta-CT and NormFinder identified *CWC15H*, *OLA1*, and *SKD1* as the most stable reference genes. These results were corroborated by geNorm, which also highlighted *ERF3A*, *CWC15H*, and *SKD1* as highly stable. In contrast, BestKeeper ranked *OLA1*, *PD2*, and *EF1γ* as the most stable, revealing a discrepancy with the other methods. Nevertheless, all four tools consistently identified *ARD2BL* and *PNR12* as the least stable genes. To integrate these outcomes, we employed RefFinder, a comprehensive platform synthesizing results from all four algorithms. It ranked *CWC15H* and *OLA1* as the most stable, followed by *SKD1*, while *ARD2BL* and *PNR12* were confirmed as the least stable, providing a consolidated stability ranking for non-floral tissues.

In floral samples, Delta-CT, NormFinder, and geNorm identified *ERF3A*, *OLA1*, and *EF1γ* as the most stable reference genes, whereas *SKD1* and *PNR12* showed lower stability. BestKeeper, however, identified *P24β3*, *CWC15H*, and *OLA1* as the most stable, and *PAIP7* and *SKD1* as the least stable, again highlighting methodological discrepancies. The RefFinder-integrated ranking resolved these variations, identifying *ERF3A* and *OLA1* as the most stable, followed by *EF1γ*, with *SKD1* and *PNR12* as the least stable, thus offering a reliable consensus for floral tissue studies.

In the combined dataset, Delta-CT and NormFinder identified *OLA1*, *ERF3A*, and *PD2* as the most stable genes. Similarly, geNorm highlighted *OLA1*, *ERF3A*, and *CWC15H*, showing minimal divergence from the former methods. BestKeeper also aligned closely, ranking *OLA1*, *CWC15H*, and *PD2* as the most reliable. In contrast, Delta-CT, NormFinder, and geNorm consistently indicated *PAIP7* and *PNR12* as the least stable, while BestKeeper identified *EF1γ* and *PNR12* as such, indicating only minor discrepancies. Collectively, these analyses consistently identified *OLA1* and *ERF3A* as stably expressed, whereas *PNR12* exhibited relatively unstable expression across sample types. To assess whether the Ct values of candidate reference genes met the normality assumption required by BestKeeper, we performed Shapiro-Wilk tests on the Ct values of ten candidate reference genes across floral and non-floral sample subsets. As shown in Supplementary Table 2, significant deviations from normality ($p < 0.05$) were observed for multiple genes in both sample types.

Correlation between RNA-seq and qRT-PCR data for gene expression

We assessed the relationship between transcript abundance, represented by FPKM values derived from transcriptome data, and expression levels measured via qRT-PCR Ct values for a set of ten candidate reference genes [28]. Transcript abundance, indicated by Log_2FPKM values (6–8), and expression levels, measured by qRT-PCR Ct values (21–27), showed distinct ranges that together indicate a broad distribution across samples (Fig. 5). A statistically significant inverse correlation was observed between Log_2(FPKM) and Ct values ($p < 0.01$), consistent with the theoretical expectation that higher transcript abundance tends to require fewer amplification cycles for detection. However, the coefficient of determination (R^2=0.0035) was extremely low, indicating that FPKM explains only a negligible fraction of the variance in Ct values. Thus, while a simple linear model ($Ct = -0.2075 \times \log_2(FPKM) + 25.466$) can be fitted, its predictive power is very limited, and it should not be relied upon for accurately estimating Ct values from transcriptomic

Table 3 The expression stability of candidate reference genes was determined and ranked by RefFinder

Rank	Non-flower		Flower		Total	
	Gene	Stability (RefFinder)	Gene	Stability (RefFinder)	Gene	Stability (RefFinder)
1	CWC15H	1.41	ERF3A	1.41	OLA1	1.00
2	OLA1	2.45	OLA1	2.06	ERF3A	2.00
3	SKD1	3.22	EF1γ	3.22	CWC15H	3.31
4	ERF3A	3.98	P24β3	3.34	PD2	3.41
5	PAIP7	4.47	PD2	4.23	P24β3	4.47
6	EF1γ	4.79	CWC15H	4.92	ARD2BL	6.24
7	PD2	5.66	ARD2BL	6.48	SKD1	7.2
8	P24β3	6.29	PAIP7	8.24	EF1γ	7.97
9	ARD2BL	9.00	SKD1	9.24	PAIP7	8.45
10	PNR12	10.00	PNR12	9.46	PNR12	10.00

Values represent the geometric mean of rankings from Delta CT, geNorm, NormFinder, and BestKeeper, as calculated by RefFinder. Lower values indicate greater expression stability

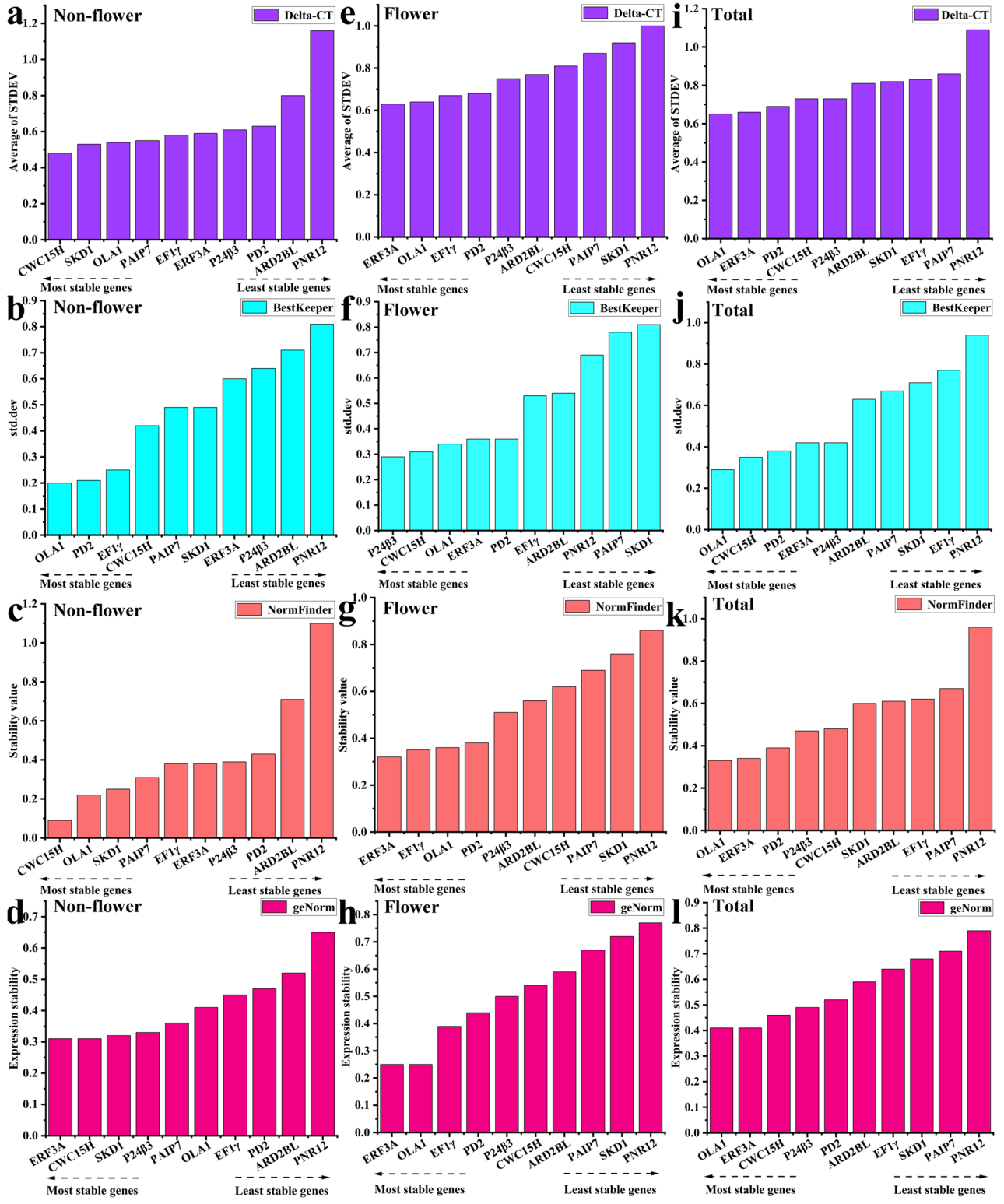

Fig. 4 Ranking of candidate reference gene stability by Delta-CT, BestKeeper, NormFinder, and geNorm Assessment of 10 putative reference gene stability in non-floral samples (**a**), floral samples(**e**), total samples (**i**) by Delta-CT analysis. Assessment of 10 putative reference gene stability in non-floral samples(**b**), floral samples(**f**), total samples (**j**) by Bestkeeper analysis. Assessment of 10 putative reference gene stability in non-floral samples(**c**), floral samples(**g**), total samples (**k**) by Normfinder analysis. Assessment of 10 putative reference gene stability in non-floral samples(**d**), floral samples(**h**), total samples (**l**) by geNorm analysis. Note: The stability scores calculated by these four software programs for reference genes indicate that lower scores correspond to higher stability, while higher scores correspond to lower stability

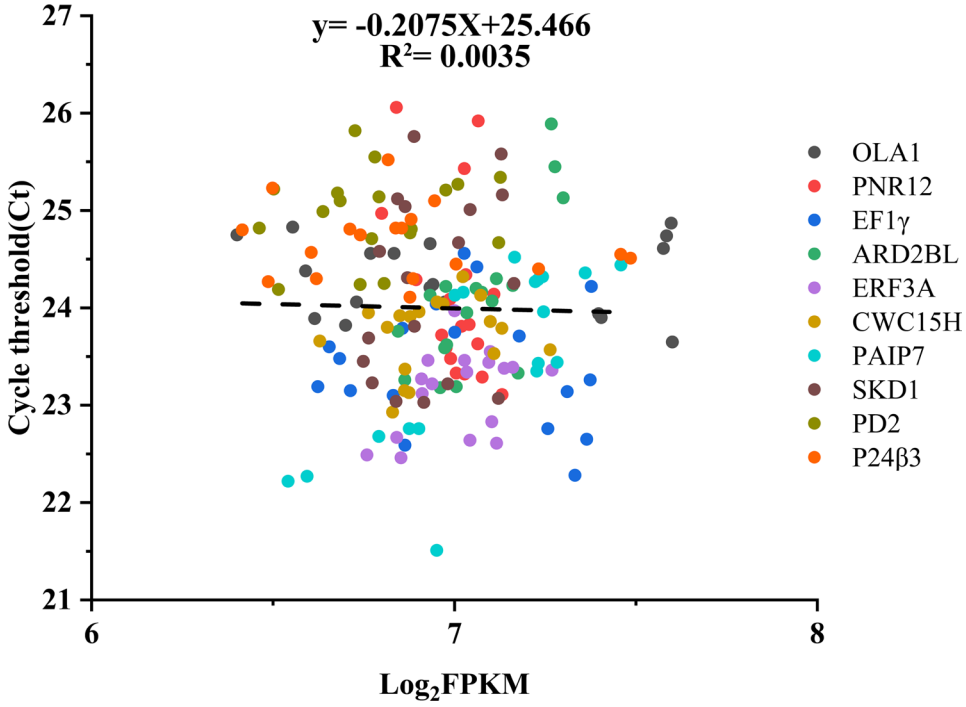

Fig. 5 Scatter plot of Ct values versus log_2FPKM values. Note: Scatter plot showing the relationship between log_2(FPKM) values from transcriptome data and Ct values from qRT-PCR for ten candidate reference genes. A statistically significant inverse correlation was observed ($p < 0.01$), but the extremely low R^2 value (0.0035) indicates limited predictive power. Each point represents an individual gene–sample combination

data. This weak correlation likely reflects inherent technical differences between RNA-seq and qRT-PCR, as well as gene-specific factors affecting amplification efficiency. Therefore, the empirical formula is presented merely as a descriptive observation rather than a practical predictive tool for reducing dependency on experimental validation.

Validation of reference genes

Based on the RefFinder stability ranking (Table 3), we selected *ERF3A* and *OLA1*, the most reliable reference genes in floral samples, as well as *SKD1* and *PNR12*, which exhibited the lowest stability, to normalize the expression of target genes (Fig. 6). Using *AP3* and *SEP2-2*, known regulators of floral sex differentiation, as target genes, we compared their expression patterns in HM (water-treated male flowers) and BM (6-BA-treated male flowers). When *ERF3A* or *OLA1* was employed as reference, the expression patterns of *AP3* and *SEP2-2* were consistent across HM and BM samples. Conversely, employing the unstable reference genes *SKD1* and *PNR12* resulted in a marked underestimation of *AP3* and *SEP2-2* transcript levels in both sample types. These results highlight the crucial impact of reference gene selection on normalization accuracy and demonstrate that the use of unstable reference genes can significantly distort the interpretation of gene expression data.

Discussion

The accuracy of qRT-PCR normalization relies on stable reference genes, yet commonly used housekeeping genes (e.g., *ACT*, *GAPDH*, *EF1α*) often exhibit variable expression across different tissues, developmental stages, or treatments [29–36], highlighting the need for condition-specific validation [37]. Transcriptome-wide approaches have proven effective for identifying stable reference genes in various plant species [38–40]. Although *V. fordii* has been extensively studied in genomics and floral development [15, 17, 41], systematic evaluation of reference genes, particularly across specialized tissues, remains limited [42, 43]. In this study, we leveraged transcriptome data from diverse floral samples to identify and validate optimal reference genes for non-floral and floral tissues, providing a reliable normalization basis for future gene expression studies in this species.

The advancement of multi-omics research has led to a substantial increase in transcriptomic data for *V. fordii*, providing a valuable resource for systematically analyzing gene expression under a range of tissue-specific and treatment conditions. In this study, we leveraged transcriptome datasets previously generated by our group, which included normal female and male flowers [15], 6-BA-treated male flowers [18], and exogenous $AgNO_3$-treated mixed floral samples [18]. By applying stringent screening criteria to these three floral transcriptome datasets, we identified three distinct gene sets. KEGG

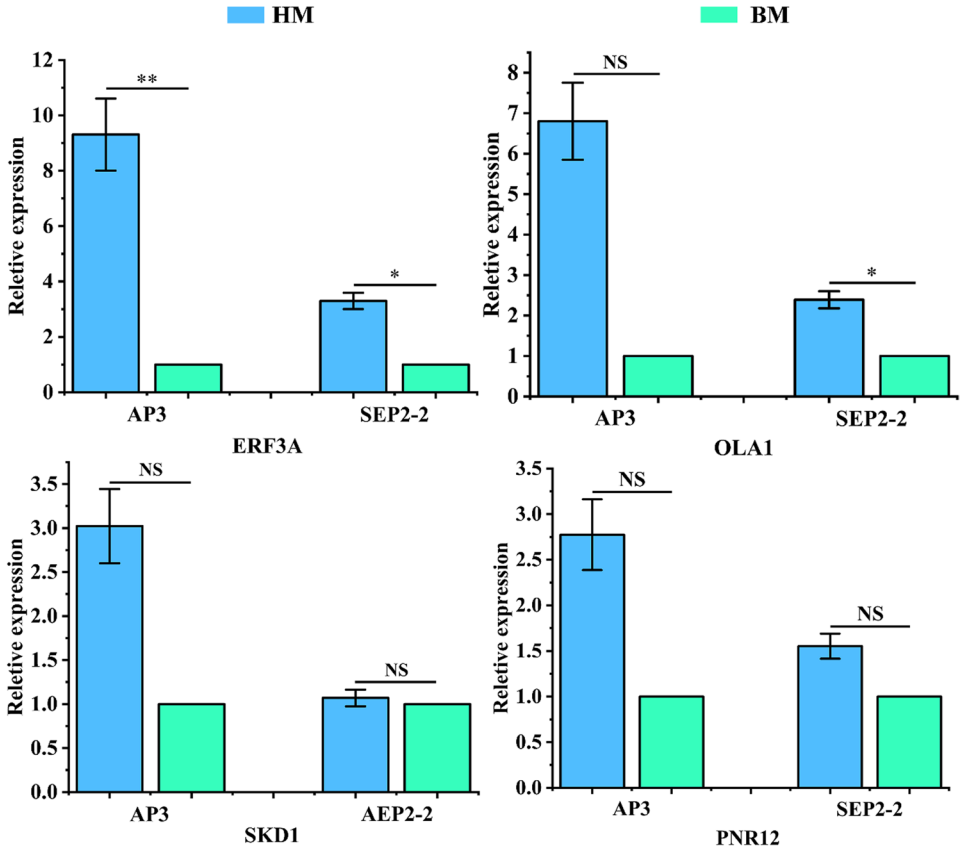

Fig. 6 *AP3* and *SEP2-2* expression in *V. fordii* flower buds. Note: Validation of reference gene suitability. Abbreviations: HM, water-treated male flowers; BM, 6-BA-treated male flowers. Expression levels of target genes *AP3* and *SEP2-2* were normalized using stable (*ERF3A*, *OLA1*) and unstable (*SKD1*, *PNR12*) reference genes in HM and BM. Error bars were derived from the standard error of three biological replicates. Statistical significance was determined using independent-sample t-tests with Bonferroni correction for multiple comparisons (two tests; corrected $\alpha = 0.025$). The symbols "*" and "**" denote significant differences at $p < 0.05$ and $p < 0.01$, respectively, while "NS" indicates no significant difference after correction

enrichment analysis revealed that pathways such as the Proteasome, Protein Processing in Endoplasmic Reticulum, Phagosome, Oxidative Phosphorylation, and Fatty Acid Metabolism were consistently enriched across all datasets (Fig. 1b–d). These findings converge with those reported in maize [44] and moso bamboo [45], suggesting that these pathways potentially underpin key processes in *V. fordii*, including the development of male and female flowers, responses to external stimuli, and floral morphogenesis. The persistent enrichment underscores the importance of these pathways in the biological processes and regulatory mechanisms that drive floral development and differentiation in *V. fordii*.

Substantial discrepancies were observed in the identification of the most stable genes by BestKeeper compared to the other three methods in non-floral and floral samples. However, in total samples, BestKeeper results aligned more closely with those from Delta-CT, NormFinder, and geNorm. This divergence can be explained by the fact that BestKeeper employs Pearson correlation analysis, which assumes normal distribution and homoscedasticity of the data [46]. In non-floral and floral samples, the expression data of highly stable genes exhibited uneven distribution and variance, leading to divergent BestKeeper outcomes. In total samples, the expression distribution of stable genes was more uniform, resulting in greater consensus across tools. For low stability genes, which generally showed poor expression characteristics, results from BestKeeper were largely consistent with those from the other algorithms. This alignment can be attributed to the fact that both Delta-CT and NormFinder are primarily based on the analysis of Ct values [22, 25]. In non-floral and total samples, geNorm produced similar but not identical rankings for the most stable genes compared to Delta-CT and NormFinder, while consistently identifying the same least stable genes. The similarity in results arises because geNorm's statistical output is derived from relative expression quantities [24], which are themselves calculated from Ct values. Consequently, its analytical results show consistency with those of Delta-CT and NormFinder, both of which directly utilize Ct values. The inconsistency between BestKeeper and the other algorithms can be attributed to their different statistical foundations. BestKeeper is a

parametric tool assuming normal distribution of input data. However, our Shapiro-Wilk test revealed significant deviation from normality ($p < 0.05$) in several sample sets (Supplementary Table S2), which likely biased the BestKeeper output. Conversely, geNorm and NormFinder, which are based on rank correlations or linear mixed models, are less sensitive to such distribution violations, providing more reliable rankings for this dataset. Moreover, BestKeeper is highly sensitive to outliers because Pearson correlation and SD can be easily distorted by extreme values [46]. Delta-CT and NormFinder are comparatively robust, and NormFinder's grouping structure further helps reduce the influence of outliers [22, 24, 25].

To integrate the results from different algorithms, this study employed RefFinder to comprehensively synthesize the gene stability rankings from all four methods, thereby providing a more robust stability evaluation. According to RefFinder's integrated assessment, *CWC15H* exhibited the highest stability in non-floral samples. *CWC15* functions as a spliceosome-associated protein that ensures efficient mRNA splicing, a critical pre-translational step for protein biosynthesis; it also positively regulates primiRNA transcription [47]. Although *CWC15* is functionally important and evolutionarily conserved in plants, its expression patterns differ markedly between female and male gametophytes [48], which may be associated with its high stability in non-floral tissues and relatively lower stability in floral samples. In floral samples, *ERF3A* was identified as the most stable gene. Previous studies have shown that *MIR159* and *ERF3A* interact to regulate the development of floral organs such as anthers and carpels [49]. This flower-specific functional relevance may contribute to the high stability of *ERF3A* in floral tissues. In the total sample set, *OLA1* exhibited the highest stability and remained consistently stabe across all sample types. *OLA1* is involved in regulating the signaling hub in plastid nucleoids for DNA replication and ribosome biogenesis during chloroplast development [50, 51], which may partially explain its relatively stable expression across diverse tissues and developmental stages.

A negative correlation was observed between Log_2FPKM and Ct values in this study, with a very low R^2 (0.0035). This can be attributed to three factors: (1) the narrow expression range resulting from the selection of stably expressed genes (FPKM ratio 0.75–1.33); (2) inherent technical discrepancies between RNA-seq and qRT-PCR platforms; and (3) biological heterogeneity across different sample types. Thus, the identification of suitable reference genes does not depend on a predictive relationship between FPKM and Ct values. Accordingly, while we caution against using this weak correlation for cross-platform prediction, we do not consider it evidence of fundamental data incomparability.

An assessment of reference gene stability was performed by profiling the expression of two functional genes (*AP3* and *SEP2-2*) across normalizations with each of the four candidate genes. The results revealed that the expression patterns of these target genes were highly dependent on the choice of reference gene for normalization. *AP3* and *SEP2-2* encode MADS-box transcription factors that play central regulatory roles in floral organ formation, particularly in promoting stamen development [52, 53]. Application of 6-BA has been reported to promote pistil development while suppressing stamen development [54, 55]. Consistent with these findings, our research shows that exogenous 6-BA inhibits stamen development and promotes pistil development in *V. fordii* [18]. When *ERF3A* and *OLA1* were employed for normalization, both *AP3* and *SEP2-2* showed a decrease in expression levels after 6-BA treatment of male flower buds, aligning with our transcriptome data. In contrast, normalization with *SKD1* or *PNR12* led to underestimation of *AP3* and *SEP2-2* expression levels. These observations highlight the necessity of employing suitable reference genes for the accurate quantification of gene expression. Based on these results, *ERF3A* and *OLA1* are recommended as the most reliable and specific reference genes for studying sex differentiation in *V. fordii*, thereby ensuring the validity and reproducibility of qRT-PCR data in related research.

Conclusion

Based on transcriptome data and qRT-PCR validation, this study systematically evaluated ten candidate reference genes for sex differentiation research in *V. fordii*. The results clearly demonstrate that no single reference gene is uniformly stable and applicable across all tissue types. Instead, the most suitable reference gene depends on the specific experimental samples. *CWC15H* ranked as the most stable gene in non-floral tissues (roots, stems, leaves, and fruits), while *ERF3A* and *OLA1* showed the highest stability in floral tissues. For studies involving all sample types, *OLA1* exhibited the best overall stability. In contrast, *PNR12* was consistently ranked as the least stable candidate. The critical importance of this selection was validated by normalizing the expression of two target genes (*AP3* and *SEP2-2*) involved in floral development; using unstable reference genes (*SKD1*, *PNR12*) led to significant underestimation and distortion of expression patterns compared to using stable ones (*ERF3A*, *OLA1*). In conclusion, this work provides the first set of validated, condition-specific reference genes for *V. fordii*. We recommend using *CWC15H* for non-floral organ studies, and *ERF3A* or *OLA1* for research on floral tissues and sex differentiation. This foundation is critical for obtaining precise gene expression data via qRT-PCR in subsequent studies elucidating the regulatory mechanisms of plant sex differentiation.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12890-w.

Supplementary Material 1: Fig. 1.

Supplementary Material 2: Table 1.

Supplementary Material 3: Table 2.

Acknowledgements
Not applicable.

Authors' contributions
Conceptualization, Lin Zhang and Meilan Liu; Wenying Li and Jie Cao; Formal analysis, Jie Cao and Wenying Li; Investigation, Lichuan Xiao and Chong Ge; Project administration, Lin Zhang; Resources, Chong Ge and Junjie Chen; Supervision, Lin Zhang; Validation, Wenying Li, Junjie Chen and Lichuan Xiao; Writing–original draft, Jie Cao.

Funding
This research work was supported by the National Natural Science Foundation of China (32230073, 32401637) and the Graduate Research Innovation Project of Hunan Province (Grant No.: CX20220705).

Data availability
The raw transcriptome sequencing data used in this study were downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive database under accession numbers SRX3843588, SRS3089151, SRS3089154, SRS3089148, SRS3089147, SRS3089150 and PRJNA1263041. The reference genome used was the V. fordii genome assembly version 2.0 (accessible at NCBI, BioProject: PRJNA503685, PRJNA445350, and PRJNA483508). Detailed information on all public transcriptomes used for analysis, including their source databases, NCBI accession numbers, and web links, has been provided in the 'Establishment of the RNA-seq Database' section of the manuscript.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 12 December 2025 / Accepted: 23 April 2026
Published online: 28 April 2026

References

1. Wang W, Zhang X, Xu X, Xu X, Fu L, Chen H. 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. 2023;14:1284007. https://doi.org/10.3389/fpls.2023.1284007.
2. Wang J, Liu L, Tang Q, Sun K, Zeng L, Wu Z. Evaluation and selection of suitable qRT-PCR reference genes for light responses in tea plant (*Camellia sinensis*). Sci Hortic. 2021;289:110488. https://doi.org/10.1016/j.scienta.2021.110488.
3. Marroufi A, Van Bockstaele E, De Loose M. Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. BMC Cell Mol Biol. 2010;11:1–12. https://doi.org/10.1186/1471-2199-11-15.
4. Chen L, Zhong HY, Kuang JF, Li JG, Lu WJ, Chen JY. Validation of reference genes for RT-qPCR studies of gene expression in banana fruit under different experimental conditions. Planta. 2011;234:377–90. https://doi.org/10.1007/s00425-011-1410-3.
5. Liu M, Wu Z, Jiang F. Selection and validation of garlic reference genes for quantitative real-time PCR normalization. Plant Cell Tiss Org. 2015;122:435–344. https://doi.org/10.1007/s11240-015-0780-9.
6. Liu J, Li P, Lu L, Xie L, Chen X, Zhang B. Selection and evaluation of potential reference genes for gene expression analysis in *Avena fatua*. Plant Protect Sci. 2018;55(1):61–71. https://doi.org/10.17221/20/2018-PPS.
7. Huggett J, Dheda K, Bustin S, Zumla A. Real-time RT-PCR normalisation; strategies and considerations. Genes Immun. 2005;6(4):279–84. https://doi.org/10.1038/sj.gene.6364190.
8. 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. 2013;158:16–21. https://doi.org/10.1016/j.scienta.2013.04.019.
9. Zhong HY, Chen JW, Li CQ, Chen L, Wu JY, Chen JY, et al. Selection of reliable reference genes for expression studies by reverse transcription quantitative real-time PCR in litchi under different experimental conditions. Plant Cell Rep. 2011;30:641–53. https://doi.org/10.1007/s00299-010-0992-8.
10. Chang Y, Hu S, Xu J, Gong H, Guo X, Song Q, et al. Identification of reference genes provides insights into the determinants of self-incompatibility in *Camellia oleifera*. Sci Hortic. 2023;321:112301. https://doi.org/10.1016/j.scienta.2023.112301.
11. Li W, Dong X, Zhang X, Cao J, Liu M, Zhou X, et al. Genome assembly and resequencing shed light on evolution, population selection, and sex identification in *Vernicia montana*. Hortic Res. 2024;11(7):uhae141. https://doi.org/10.1093/hr/uhae141.
12. Cui P, Lin Q, Fang D, Zhang L, Li R, Cheng J, et al. Tung tree (*Vernicia fordii*, Hemsl.) genome and transcriptome sequencing reveals co-ordinate up-regulation of fatty acid β-oxidation and triacylglycerol biosynthesis pathways during eleostearic acid accumulation in seeds. Plant Cell Physiol. 2018;59(10):1990–2003. https://doi.org/10.1093/pcp/pcy117.
13. Li X, Zhang J, Lin S, Xing Y, Zhang X, Ye M, et al. (+)-Catechin, epicatechin and epigallocatechin gallate are important inducible defensive compounds against *Ectropis grisescens* in tea plants. Plant Cell Environ. 2022;45(2):496–511. https://doi.org/10.1111/pce.14216.
14. Li W, Liu M, Dong X, Cao H, Wu Y, Shang H, et al. Flower biology and ontogeny of the tung tree (*Vernicia fordii* Hemsl). Trees-Struct Funct. 2020;34(6):1363–681. https://doi.org/10.1007/s00468-020-02041-3.
15. Liu M, Li W, Zhao G, Fan X, Long H, Fan Y, et al. New insights of salicylic acid into stamen abortion of female flowers in tung tree (*Vernicia fordii*). Front genet. 2019a;10:316. https://doi.org/10.3389/fgene.2019.00316.
16. Liu M, Long H, Li W, Shi M, Cao H, Zhang L. Boosting C16 fatty acid biosynthesis of *Escherichia coli*, yeast and tobacco by tung tree (*Vernicia fordii* Hemsl.) beta-hydroxyacyl-acyl carrier protein dehydratase gene. Ind Crop Prod. 2019b;127:46–54. https://doi.org/10.1016/j.indcrop.2018.10.067.
17. Zhang L, Liu M, Long H, Dong W, Pasha A, Esteban E, et al. Tung tree (*Vernicia fordii*) genome provides a resource for understanding genome evolution and improved oil production. Genom Proteom Bioinf. 2019;17(6):558–75. https://doi.org/10.1016/j.gpb.2019.03.006.
18. Liu M. Studies on the characteristics of female flower development and its hormone regulation mechanism in tung tree [Doctoral dissertation, Central South University of Forestry and Technology]. 2019. CNKI. https://doi.org/10.27662/d.cnki.gznlc.2019.000130
19. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90. https://doi.org/10.1093/bioinformatics/bty560.
20. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907. https://doi.org/10.1038/s41587-019-0201-4.
21. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5. https://doi.org/10.1038/nbt.3122.
22. Silver N, Best S, Jiang J, Thein SL. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol Cell Biol. 2006;7:1–9. https://doi.org/10.1186/1471-2199-7-33.
23. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: bestkeeper–excel-based tool using pair-wise correlations. Biotechnol Lett. 2004;26:509–15. https://doi.org/10.1023/b:bile.0000019559.84305.47.
24. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3:1–12. https://doi.org/10.1186/gb-2002-3-7-research0034.

25. Andersen CL, Jensen JL, Ørntoft TF. 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. 2004;64(15):5245–50. https://doi.org/10.1158/0008-5472.CAN-04-0496.
26. 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. 2012;80:75–84. https://doi.org/10.1007/s11103-012-9885-2.
27. Xie F, Wang J, Zhang B. RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes. Funct Integr Genomics. 2023;23(2):125. https://doi.org/10.1007/s10142-023-01055-7.
28. Li Y, Zhang L, Li R, Zhang M, Li Y, Wang H, et al. Systematic identification and validation of the reference genes from 60 RNA-Seq libraries in the scallop *Mizuhopecten yessoensis*. BMC Genomics. 2019;20(1):288. https://doi.org/10.1186/s12864-019-5661-x.
29. Gachon C, Mingam A, Charrier B. Real-time PCR: what relevance to plant studies? J Exp Bot. 2004;55(402):1445–54. https://doi.org/10.1093/jxb/erh181.
30. Bustin SA, Benes V, Nolan T, Pfaffl MW. Quantitative real-time RT-PCR–a perspective. J Mol Endocrinol. 2005;34(3):597–601. https://doi.org/10.1677/jme.1.01755.
31. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. Nat protoc. 2006;1(3):1559–82. https://doi.org/10.1038/nprot.2006.236.
32. Umadevi P, Suraby EJ, Anandaraj M, Nepolean T. Identification of stable reference gene for transcript normalization in black pepper-*Phytophthora capsici* pathosystem. Physiol Mol Biol Pla. 2019;25:945–52. https://doi.org/10.1007/s12298-019-00653-9.
33. Wu H. Plant salt tolerance and Na+ sensing and transport. Crop J. 2018;6(3):215–25. https://doi.org/10.1016/j.cj.2018.01.003.
34. Yang Q, Li J, Shen J, Xu Y, Liu H, Deng W, et al. Metabolic resistance to acetolactate synthase inhibiting herbicide tribenuron-methyl in *Descurainia sophia* L. mediated by cytochrome *P450* enzymes. J Agric Food Chem. 2018;66(17):4319–27. https://doi.org/10.1021/acs.jafc.7b05825.
35. Li MY, Wang F, Jiang Q, Wang GL, Tian C, Xiong AS. Validation and comparison of reference genes for qPCR normalization of celery (*Apium graveolens*) at different development stages. Front Plant Sci. 2016;7:313. https://doi.org/10.3389/fpls.2016.00313.
36. Die JV, Román B, Nadal S, González-Verdejo CI. Evaluation of candidate reference genes for expression studies in *Pisum sativum* under different experimental conditions. Planta. 2010;232:145–53. https://doi.org/10.1007/s00425-010-1158-1.
37. Ji T, Ma S, Liang M, Wang X, Gao L, Tian Y. Reference genes identification for qRT-PCR normalization of gene expression analysis in *Cucumis sativus* under *Meloidogyne incognita* infection and *Pseudomonas* treatment. Front Plant Sci. 2022;13:1061921. https://doi.org/10.3389/fpls.2022.1061921.
38. Dos Santos KCG, Desgagné-Penix I, Germain H. Custom selected reference genes outperform pre-defined reference genes in transcriptomic analysis. BMC Genomics. 2020;21:1–9. https://doi.org/10.1186/s12864-019-6426-2.
39. Zhao H, Gao Z, Wang L, Wang J, Wang S, Fei B, et al. Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). Gigascience. 2018;7(10):giy115. https://doi.org/10.1093/gigascience/giy115.
40. Chen J, Wang Y, Yang Z, Liu D, Jin Y, Li X, et al. Identification and validation of the reference genes in the echiuran worm *Urechis unicinctus* based on transcriptome data. BMC Genomics. 2023;24(1):248. https://doi.org/10.1186/s12864-023-09358-6.
41. Dong X, Li W, Li C, Akan OD, Liao C, Cao J, Zhang L. Integrated transcriptomics and metabolomics revealed the mechanism of catechin biosynthesis in response to lead stress in tung tree (*Vernicia fordii*). Sci Total Environ. 2024;930:172796. https://doi.org/10.1016/j.scitotenv.2024.172796.
42. Cao H, Cao F, Klasson KT. Characterization of reference gene expression in tung tree (*Vernicia fordii*). Ind Crop Prod. 2013;50:248–55. https://doi.org/10.1016/j.indcrop.2013.07.030.
43. Cao H, Shockey JM. Comparison of TaqMan and SYBR Green qPCR methods for quantitative gene expression in tung tree tissues. J Agric Food Chem. 2012;60(50):12296–303. https://doi.org/10.1021/jf304690e.
44. Lin F, Jiang L, Liu Y, Lv Y, Dai H, Zhao H. Genome-wide identification of housekeeping genes in maize. Plant Mol Biol. 2014;86:543–54. https://doi.org/10.1007/s11103-014-0246-1.
45. Liu Y, Zhu C, Lin Z, Li H, Di X, Yue X, et al. Systematic identification and validation of the reference genes from 447 transcriptome datasets of moso bamboo (*Phyllostachys edulis*). Hortic Plant J. 2024;11(3):1353–63. https://doi.org/10.1016/j.hpj.2023.11.007.
46. Nazet U, Schröder A, Spanier G, Wolf M, Proff P, Kirschneck C. Simplified method for applying static isotropic tensile strain in cell culture experiments with identification of valid RT-qPCR reference genes for PDL fibroblasts. Eur J Orthodont. 2020;42(4):359–70. https://doi.org/10.1093/ejo/cjz052.
47. Zhou B, Yu H, Xue Y, Li M, Zhang C, Yu B. The spliceosome-associated protein CWC15 promotes miRNA biogenesis in *Arabidopsis*. Nat Commun. 2024;15(1):2399. https://doi.org/10.1038/s41467-024-46676-z.
48. Slane D, Lee CH, Kolb M, Dent C, Miao Y, Franz-Wachtel M, et al. The integral spliceosomal component *CWC15* is required for development in *Arabidopsis*. Sci Rep. 2020;10(1):13336. https://doi.org/10.1038/s41598-020-70324-3.
49. Anand S, Lal M, Bhardwaj E, Shukla R, Pokhriyal E, Jain A, et al. *MIR159* regulates multiple aspects of stamen and carpel development and requires dissection and delimitation of differential downstream regulatory network for manipulating fertility traits. Physiol Mol Biol Pla. 2023;29(10):1437–56. https://doi.org/10.1007/s12298-023-01377-7.
50. Chen J, Deng F, Deng M, Han J, Chen J, Wang L, et al. Identification and characterization of a chloroplast-targeted *Obg GTPase* in *Dendrobium officinale*. DNA Cell Biol. 2016;35(12):802–11. https://doi.org/10.1089/dna.2016.3413.
51. Chen J, Wang L, Jin X, Wan J, Zhang L, Je BI, et al. *Oryza sativa ObgC1* acts as a key regulator of DNA replication and ribosome biogenesis in chloroplast nucleoids. Rice. 2021;14(1):65. https://doi.org/10.1186/s12284-021-00498-5.
52. Ma YY, Meng Q, Tan XM, Yang L, Zhang KL, Xu ZQ. Functional identification of the different regions in B-class floral homeotic MADS-box proteins *IiAP3* and *IiPI* from *Isatis indigotica*. Physiol Plant. 2022;174(3):e13713. https://doi.org/10.1111/ppl.13713.
53. Li J, Jia D, Chen X. *HUA1*, a regulator of stamen and carpel identities in *Arabidopsis*, codes for a nuclear RNA binding protein. Plant Cell. 2001;13(10):2269–81. https://doi.org/10.1105/tpc.010201.
54. Luo Y, Pan BZ, Li L, Yang CX, Xu ZF. Developmental basis for flower sex determination and effects of cytokinin on sex determination in *Plukenetia volubilis* (Euphorbiaceae). Plant Reprod. 2020;33:21–34. https://doi.org/10.1007/s00497-019-00382-9.
55. Wu G, Tian X, Qiu Q, Zhang Y, Fan X, Yuan D. Dynamic cytological and transcriptomic analyses provide novel insights into the mechanisms of sex determination in *Castanea henryi*. Front Plant Sci. 2023;14:1257541. https://doi.org/10.3389/fpls.2023.1257541.

Publisher's Note