Evaluation of Reference Genes for Expression Normalization in Psoralea Corylifolia L. under Abiotic Stress

Yuan Zhong  
Jiangsu Health Vocational College

Yanan Gai  
Jiangsu Province and Chinese Academy of Sciences

Jiajia Gao  
China Pharmaceutical University

Weifen Nie  
China Pharmaceutical University

Zhenzhen Bao  
Jiangsu Health Vocational College

Weiqi Wang  
Jiangsu Health Vocational College

Xiaoqing Xu  
Jiangsu Health Vocational College

Jie Wu  
Jiangsu Health Vocational College

Yuedong He (✉ yuedong_he@yeah.net)  
Jiangsu Health Vocational College/Hunan Agricultural University

Research Article

Keywords: P. corylifolia, RT-qPCR, Reference gene, Normalization, Abiotic stress

DOI: https://doi.org/10.21203/rs.3.rs-469808/v1

License: ☺️ ① This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background: *Psoralea corylifolia* L., a traditional Chinese herbal medicine, has a wide range of pharmacological activities for the treatment of various diseases. To date, there have been no published reports of the systematic evaluation and selection of the best reference genes for biological analysis of *P. corylifolia*. Reverse transcription quantitative polymerase chain reaction (RT-qPCR) is a method for quantifying gene expression. Selecting the appropriate reference genes is essential to ensure accurate normalization of RT-qPCR results.

Results: In this study, 10 candidate reference genes, namely actin (*ACT*), glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), protein phosphatase 2A subunit A3 (*PP2A*), nuclear cap-binding protein subunit 2 (*NCBP2*), cyclophilin 2 (*CYP2*), elongation factor 1-alpha (*EF-1α*), protein phosphatase 2A-2 (*PP2A2*), thioredoxin-like protein YLS8 (*YLS8*), polypyrimidine tract-binding protein (*PTBP*), and TIP41-like family protein (*TIP41*), were screened from the transcriptome database of *P. corylifolia*. The plants were treated under different abiotic stress conditions (H₂O₂, NaCl, CuSO₄, high-temperature, low-temperature, PEG, and UV). Next, their expression stability was evaluated using four algorithm tools: BestKeeper, NormFinder, geNorm, and RefFinder. The results revealed that *TIP41*, *CYP2*, and *YLS8* were the top three genes with the most stable expression levels under various abiotic stress conditions.

Conclusion: We successfully selected appropriate reference genes for *P. corylifolia*, providing a reliable basis for further studies on the quantitative gene expression profile of the main active components in *P. corylifolia*.

Background

*P. corylifolia* L. belongs to the Leguminosae family [1, 2], and is broadly distributed in tropical and subtropical regions, particularly in China and South Africa [3]. The active ingredients in *P. corylifolia* are mainly coumarins, such as psoralen and isopsoralen, which display a wide range of biological activities including anti-inflammatory, anti-oxidative, anti-bacterial, immunomodulation, and tumor-suppressive activities [4–9]. Owing to its extensive clinical effects [10, 11], *P. corylifolia* has earned a high reputation throughout the history of traditional Chinese medicine. In recent years, it has also been found to be effective in treating diabetes, psoriasis and other diseases [12–16].

Compared with synthetic compounds, natural active components in plants have been pre-screened by the natural environment, which ensures their stability, bioavailability, and biological activity. Understanding their biosynthetic pathways is important for the effective synthesis of these natural products *in vitro* [17]. Moreover, owing to habitat destruction as well as illegal and indiscriminate collection, *P. corylifolia* has become a rare and endangered medicinal herb [18]. Therefore, there is an urgent need to study the biosynthetic pathways of active components in *P. corylifolia*.

Advances in sequencing and bioinformatics technologies, coupled with the growing knowledge of metabolic pathways in plant genomes, have led to significant elucidation of the biosynthetic pathways of natural products. Two-step reverse transcription-quantitative PCR (RT-qPCR), is a commonly used technique for quantifying gene expression levels [19–22]. However, the accuracy of this method is determined by various biological and technical factors, including sample type and amount, primer design, RNA degradation, cDNA quality, and PCR efficiency [23–27]. Hence, a normalization process is needed to eliminate the variations caused by these factors. Generally, normalization is conducted by adding an internal control gene, which is also known as a reference gene or housekeeping gene because of its relatively constant expression level in plants [25, 27, 28]. Some reference genes have been frequently used for normalization analysis in plants, such as glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), and elongation factor 1-alpha (*EF-1α*).
dehydrogenase (GAPDH), 18 S ribosomal RNA (18 S rRNA), α-tubulin (TUB α), β-actin (ACT β), and elongation factor 1α (EF-1α). However, recent studies have reported variabilities in the expression of these reference genes under different processing environments [29–32]. Selecting incorrect reference genes will inevitably misguide the analysis. To date, there has been no systematic evaluation or validation of reference genes for RT-qPCR analysis of *P. corylifolia*. Consequently, identification of a reliable reference gene is essential for elucidating the gene expression profiles of *P. corylifolia*.

Therefore, in this study, we aimed to discover a reliable reference gene for *P. corylifolia*. For this purpose, 10 candidate reference genes, namely ACT, GAPDH, PP2A, NCBP2, CYP2, EF-1α, PP2A2, YLS8, PTBP, and TIP41, were selected from the *P. corylifolia* database. To determine the appropriate reference genes under different environmental conditions, the following abiotic stress conditions were applied: oxidative stress (H₂O₂), salt stress (NaCl), mental stress (CuSO₄), high-temperature (42°C) and low-temperature (4°C) stress, osmotic stress (polyethylene glycol, PEG), and ultraviolet (UV) rays. Finally, the raw data were analyzed using three Excel-based software packages, BestKeeper [33], NormFinder [34], and geNorm [35]. RefFinder was used to comprehensively evaluate the results. This study is the first to explore the selection of appropriate reference genes in *P. corylifolia* under different abiotic stress conditions. It also provides a credible normalization standard and basis for further analysis of the gene expression of bioactive components in *P. corylifolia*.

**Material And Methods**

**Plant materials and stress treatment**

One-year-old plants of *P. corylifolia* were collected from from Ningguo City, Anhui Province, China (longitude: 118.95E, latitude: 30.62N), and was identified as *Psoralea corylifolia* Linn. by Zhang Ning, school of pharmacy, Jiangsu Health Vocational College, and deposited in the herbarium of the medicinal botanical garden (ID: JSJK-PC-012). The planting field of *P. corylifolia* was a private land and the landowner has allowed us to use it for study. Then, these accessions were transplanted into plastic basin composed of vermiculite, perlite, and peat moss at a ratio of 1:1:1. Next, the plants were grown in a greenhouse at a temperature of 25 °C, a long photoperiod of 16 hours light and 8 hours darkness, 40–70% relative humidity, and 3000 lux light intensity until treated. To induce oxidative stress, the plants were exposed to 200 ml of H₂O₂ for 24 h. For the salt stress treatment, the plants were exposed to approximately 200 mL (600 mM) NaCl for 7 days. Heavy metal stress was induced by treatment with 500 mM CuSO₄ for 24 h. For high- and low-temperature stress, the plants were stored in a constant-temperature incubator at 42°C and 4°C, respectively, for 48 h. To induce osmotic stress, the plants were cultured in 200 mL of 25% PEG 6000 for one week. For UV stress, the plants were exposed to UV light for 24 h, during which irrigation was carried out with 100 mL distilled water. The control group received no treatment, but was sprinkled with distilled water over time. All groups had three biological replicates. The collected fresh leaf samples were cleaned, sterilized, frozen in liquid nitrogen to prevent RNA degradation, and then stored in a refrigerator at -80°C [36].

**Selection of candidate reference genes and primers design**

Based on previous studies in various plants, 10 common reference genes (ACT, GAPDH, PP2A, NCBP2, CYP2, EF-1α, PP2A2, YLS8, PTBP, and TIP41) were selected as candidates reference genes in this study. Information on the reference gene sequence of *P. corylifolia* was obtained from the TAIR database and is listed in Table 1. The potential unigenes of the local blast were screened by a program in the Bioedit Sequence Alignment Editor (v7.0.9), and the value of FPKM (fragments per kilobase of exon model per million mapped fragments) was employed to
characterize the expression level of unigene, which could eliminate the influence of gene length and sequencing volume differences on the calculated gene expression level (Table S1).

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Homolog locus</th>
<th>Primers 5'-3' (F/R)</th>
<th>Length</th>
<th>PCR efficiency (E)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>Actin – 11</td>
<td>AT3G18780</td>
<td>CACATGCCATCCTTCGTCTT/ GGCAATGTAAGCCAGCTTTT</td>
<td>146</td>
<td>1.852</td>
<td>0.999</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
<td>AT1G42970</td>
<td>CCTGGATGAAGAGTTCGGAA/ CCTGTGCTCGTTGGGACTAT</td>
<td>144</td>
<td>1.839</td>
<td>0.999</td>
</tr>
<tr>
<td>PP2A</td>
<td>Protein phosphatase 2A subunit A3</td>
<td>AT1G13320</td>
<td>TTTTGCCGGATTTTAGATCC/ GGAGCCATTCCCATTATCAC</td>
<td>125</td>
<td>1.530</td>
<td>0.997</td>
</tr>
<tr>
<td>NCBP2</td>
<td>Nuclear cap-binding protein subunit 2</td>
<td>AT5G44200</td>
<td>ATGGGCTTGGACAAGAACAC/ TGAAACCCCCAATCAAATC</td>
<td>149</td>
<td>1.704</td>
<td>0.999</td>
</tr>
<tr>
<td>CYP2</td>
<td>Cyclophilin 2</td>
<td>AT4G33060</td>
<td>AATACTCAAAGACGCCAGG/ GACCTGGCTGCTGACGTATT</td>
<td>105</td>
<td>1.870</td>
<td>0.999</td>
</tr>
<tr>
<td>EF-1α</td>
<td>Elongation factor 1-alpha</td>
<td>AT1G07920</td>
<td>TCTTCTTGCTTTACCCCTCG/ CAATTTCATCATATCGGGCC</td>
<td>101</td>
<td>1.757</td>
<td>0.999</td>
</tr>
<tr>
<td>PP2A2</td>
<td>Protein phosphatase 2A-2</td>
<td>AT1G59830</td>
<td>TCTAGGCGATACCCAGGT/ CAAGGGCAAATAATCGAAC</td>
<td>108</td>
<td>1.638</td>
<td>0.916</td>
</tr>
<tr>
<td>YLS8</td>
<td>Thioredoxin-like protein YLS8</td>
<td>AT5G08290</td>
<td>GCATCTGTGGCGGAGACTAT/ GAAGAACATCACGTCGAGG</td>
<td>117</td>
<td>1.861</td>
<td>0.999</td>
</tr>
<tr>
<td>PTBP</td>
<td>Polypyrimidine tract-binding protein</td>
<td>AT3G01150</td>
<td>TATGGTTCGTGGAAAAACC/ CACGCCCTCGATAGTTACCA</td>
<td>115</td>
<td>1.860</td>
<td>0.999</td>
</tr>
<tr>
<td>TIP41</td>
<td>TIP41-like family protein</td>
<td>AT4G34270</td>
<td>CCTGGAGATGAGGGTTG/ AACGATGCTGCTTCAGGAAA</td>
<td>134</td>
<td>1.863</td>
<td>0.999</td>
</tr>
</tbody>
</table>

The results of the melt curve analysis of the PCR products are depicted in Fig. S1, and most of them presented a single peak curve. The specificity of the primer sequences was verified. However, compared with that of other reference genes, the melting curve of PP2A showed an apparent fluctuation trend, indicating that the specificity of its primer may be poor.

**Total RNA extraction and RT-qPCR analysis**

RNA was extracted from approximately 100 mg of frozen samples using a Spectrum Plant Total RNA kit (Sigma, USA). The purity of the extracted RNA was analyzed using a NanoDrop spectrophotometer 2000 (Thermo Scientific, USA). Only samples with OD_{260/280} and OD_{260/230} ratios between 1.8–2.1 and 1.6–2.2, respectively, can be used for
Reverse transcription was performed using HiScript II Reverse Transcriptase (Vazyme, Nanjing, China) in a 20 µL volume system containing 1 µg RNA. The reverse cDNA was diluted 10-fold as a template and amplified in PCR instrument under the following conditions: one cycle of 95°C for 5 min, 40 cycles of 95°C for 10 s, 58–60°C for 20 s, and 72°C for 20 s. The relative expression levels of the genes were calculated according to the formula $2^{\Delta\Delta Ct}$ [37].

### Gene expression stability analysis

The cycle threshold (Ct) value indicates the cycle number when the generated fluorescence signals in each response tube reach a detectable level. Hence, the Ct value directly reflects the expression level of a candidate reference gene. Correspondingly, a lower Ct value indicated a higher gene expression level. In this study, the expression stability of the candidate reference genes was compared using Bestkeeper, NormFinder and geNorm. These kinds of software use different algorithms to rank the expression level of each candidate gene.

BestKeeper is an algorithm software based on Excel, in which the raw Ct value can be inputted directly for evaluating and sequencing the stability of the corresponding reference genes under different stresses by calculating two parameters: coefficient of variation (CV) and standard deviation (SD). A low CV ± SD value indicates a high gene expression stability.

In contrast, in NormFinder and geNorm, the raw Ct value needs to be converted by the formula $2^{-\Delta Ct}$ ($\Delta Ct =$ each corresponding Ct value - the minimum Ct value). geNorm shows the gene expression level by the parameter (M), where the threshold value of M is 1.5 [38]. The lower the M value, the higher the expression stability of the candidate. Furthermore, the pairwise variation ($V_n/V_{n+1}$) analysis of geNorm can be utilized to optimize the number of reference genes for accurate normalization. In general, when the value of $V_n/V_{n+1}$ is less than 0.15, no additional genes are required for normalization [39]. NormFinder measures the stability of gene expression by considering the value of the inter-group and intra-group variance M. Similar to that in other two algorithms, the lowest value of parameter M reflects the most stable gene expression. Through comprehensive analysis of the results of these software, reference genes that show the most stable expression under different stress conditions were accurately screened.

### Statistical analysis

Three technical and biological replicates were implemented to obtain qPCR data, and the raw data are listed in Table S2. All data are presented as the mean ± standard error of the mean (SEM). Graphs were generated using Origin 2019 (OriginLab Corporation, Northampton, MA, USA). All data were analyzed using the geNorm, NormFinder and BestKeeper software according to the specification methods.

### Results

#### Expression profile of the candidate reference genes

The mean Ct values of the 10 candidate reference genes ranged from 9.95-33.16, and most were distributed between 18 and 24. PP2A presented the highest expression level, with the lowest Ct values. TIP41 had the least variability, with the lowest SD value of 0.72, followed by YLS8, CYP2, PP2A, and ACT, suggesting that these genes possessed stable expression levels under different stress conditions. In contrast, PP2A2 showed the highest mean Ct value (33.16 ± 2.44), which indicated that this reference was not worthy of consideration (Fig. 1).
Expression stability of the reference genes

BestKeeper analysis

The results of BestKeeper analysis are shown in Table 2. As shown in the table, the CV ± SD value increases from top to bottom (the lowest value is shown in the first row of the table). In all groups, TIP41 showed the highest stability with the lowest CV ± SD value of 2.45 ± 0.56, followed by CYP2 and YLS8. Accordingly, under H2O2, NaCl, CuSO4, high temperature, UV, and wild type stress, TIP41 showed the most stable expression level, with CV ± SD values as low as 2.35 ± 0.54, 1.54 ± 0.36, 0.95 ± 0.21, 1.82 ± 0.40, and 1.87 ± 0.45, respectively. With the lowest CV ± SD value of 1.64 ± 0.31 and 1.55 ± 0.29, respectively, YLS8 and GAPDH showed the highest expression stability under low-temperature and PEG stress. It is worth noting that the high CV ± SD value of PP2A2 under each condition indicated that its gene expression was not stable, which corresponded with the result of the Ct value. However, PP2A showed the lowest stability with the lowest CV ± SD value under most stress treatments, which was contrary to the outcome of the Ct value. Therefore, further evaluation is of great significance to obtain accurate results in reference gene selection.

Table 2

Expression stability rank of 10 candidate reference genes using BestKeeper in *P. corylifolia* (coefficient of variation (CV) ± standard deviation (SD)).
<table>
<thead>
<tr>
<th>Rank</th>
<th>H$_2$O$_2$</th>
<th>NaCl</th>
<th>CuSO$_4$</th>
<th>high-temperature</th>
<th>low-temperature</th>
<th>PEG</th>
<th>UV</th>
<th>WT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TIP41</td>
<td>TIP41</td>
<td>TIP41</td>
<td>TIP41</td>
<td>TIP41</td>
<td>YLS8</td>
<td>GAPDH</td>
<td>TIP41</td>
<td>TIP41</td>
</tr>
<tr>
<td></td>
<td>2.35 ± 0.54</td>
<td>1.54 ± 0.36</td>
<td>0.95 ± 0.21</td>
<td>1.82 ± 0.40</td>
<td>1.64 ± 0.31</td>
<td>1.55 ± 0.29</td>
<td>1.87 ± 0.45</td>
<td>1.86 ± 0.42</td>
<td>2.45 ± 0.56</td>
</tr>
<tr>
<td>2</td>
<td>CYP2</td>
<td>PTBP</td>
<td>EF-1α</td>
<td>EF-1α</td>
<td>PTBP</td>
<td>TIP41</td>
<td>GAPDH</td>
<td>PTBP</td>
<td>CYP2</td>
</tr>
<tr>
<td></td>
<td>3.03 ± 0.68</td>
<td>1.93 ± 0.41</td>
<td>1.58 ± 0.34</td>
<td>1.88 ± 0.39</td>
<td>1.74 ± 0.37</td>
<td>1.69 ± 0.39</td>
<td>2.23 ± 0.48</td>
<td>2.24 ± 0.48</td>
<td>3.23 ± 0.73</td>
</tr>
<tr>
<td>3</td>
<td>EF-1α</td>
<td>CYP2</td>
<td>PTBP</td>
<td>GAPDH</td>
<td>EF-1α</td>
<td>EF-1α</td>
<td>CYP2</td>
<td>EF-1α</td>
<td>YLS8</td>
</tr>
<tr>
<td></td>
<td>3.38 ± 0.74</td>
<td>1.97 ± 0.44</td>
<td>1.99 ± 0.42</td>
<td>2.03 ± 0.40</td>
<td>2.26 ± 0.49</td>
<td>1.99 ± 0.43</td>
<td>2.23 ± 0.54</td>
<td>2.66 ± 0.55</td>
<td>3.64 ± 0.69</td>
</tr>
<tr>
<td>4</td>
<td>YLS8</td>
<td>YLS8</td>
<td>YLS8</td>
<td>ACT</td>
<td>TIP41</td>
<td>CYP2</td>
<td>ACT</td>
<td>CYP2</td>
<td>PTBP</td>
</tr>
<tr>
<td></td>
<td>3.69 ± 0.70</td>
<td>2.32 ± 0.43</td>
<td>2.07 ± 0.39</td>
<td>2.25 ± 0.45</td>
<td>2.30 ± 0.52</td>
<td>2.02 ± 0.46</td>
<td>3.12 ± 0.69</td>
<td>2.73 ± 0.60</td>
<td>3.71 ± 0.80</td>
</tr>
<tr>
<td>5</td>
<td>ACT</td>
<td>ACT</td>
<td>GAPDH</td>
<td>PTBP</td>
<td>CYP2</td>
<td>YLS8</td>
<td>PTBP</td>
<td>YLS8</td>
<td>EF-1α</td>
</tr>
<tr>
<td></td>
<td>3.75 ± 0.71</td>
<td>3.23 ± 0.61</td>
<td>2.23 ± 0.36</td>
<td>2.32 ± 0.52</td>
<td>2.43 ± 0.55</td>
<td>2.02 ± 0.38</td>
<td>3.35 ± 0.77</td>
<td>3.09 ± 0.56</td>
<td>3.96 ± 0.85</td>
</tr>
<tr>
<td>6</td>
<td>PTBP</td>
<td>GAPDH</td>
<td>CYP2</td>
<td>CYP2</td>
<td>ACT</td>
<td>ACT</td>
<td>EF-1α</td>
<td>ACT</td>
<td>ACT</td>
</tr>
<tr>
<td></td>
<td>3.88 ± 0.83</td>
<td>3.67 ± 0.62</td>
<td>2.31 ± 0.51</td>
<td>2.98 ± 0.65</td>
<td>2.85 ± 0.55</td>
<td>2.15 ± 0.43</td>
<td>3.89 ± 0.92</td>
<td>3.15 ± 0.61</td>
<td>4.94 ± 0.97</td>
</tr>
<tr>
<td>7</td>
<td>GAPDH</td>
<td>EF-1α</td>
<td>ACT</td>
<td>YLS8</td>
<td>GAPDH</td>
<td>PP2A2</td>
<td>YLS8</td>
<td>PP2A2</td>
<td>PP2A2</td>
</tr>
<tr>
<td></td>
<td>4.24 ± 0.70</td>
<td>4.28 ± 0.89</td>
<td>2.65 ± 0.49</td>
<td>3.49 ± 0.64</td>
<td>3.70 ± 0.67</td>
<td>6.29 ± 2.05</td>
<td>4.61 ± 0.94</td>
<td>4.00 ± 1.33</td>
<td>5.66 ± 1.88</td>
</tr>
<tr>
<td>8</td>
<td>PP2A2</td>
<td>PP2A2</td>
<td>NCBP2</td>
<td>NCBP2</td>
<td>PP2A2</td>
<td>NCBP2</td>
<td>PP2A</td>
<td>NCBP2</td>
<td>NCBP2</td>
</tr>
<tr>
<td></td>
<td>4.80 ± 1.62</td>
<td>4.35 ± 1.44</td>
<td>5.88 ± 1.35</td>
<td>5.09 ± 1.20</td>
<td>5.52 ± 1.82</td>
<td>9.16 ± 2.26</td>
<td>5.08 ± 0.49</td>
<td>4.80 ± 1.10</td>
<td>7.14 ± 1.70</td>
</tr>
<tr>
<td>9</td>
<td>NCBP2</td>
<td>NCBP2</td>
<td>PP2A2</td>
<td>PP2A</td>
<td>PP2A</td>
<td>PTBP</td>
<td>NCBP2</td>
<td>GAPDH</td>
<td>GAPDH</td>
</tr>
<tr>
<td></td>
<td>6.02 ± 1.43</td>
<td>7.67 ± 1.77</td>
<td>6.20 ± 2.06</td>
<td>5.91 ± 0.59</td>
<td>5.67 ± 0.56</td>
<td>9.42 ± 1.97</td>
<td>5.84 ± 1.52</td>
<td>5.16 ± 0.89</td>
<td>8.04 ± 1.45</td>
</tr>
<tr>
<td>10</td>
<td>PP2A</td>
<td>PP2A</td>
<td>PP2A</td>
<td>PP2A2</td>
<td>NCBP2</td>
<td>PP2A</td>
<td>PP2A2</td>
<td>PP2A</td>
<td>PP2A</td>
</tr>
<tr>
<td></td>
<td>7.97 ± 0.81</td>
<td>8.82 ± 0.90</td>
<td>10.67 ± 1.03</td>
<td>6.70 ± 2.22</td>
<td>6.25 ± 1.47</td>
<td>10.35 ± 1.03</td>
<td>6.22 ± 2.06</td>
<td>10.71 ± 1.07</td>
<td>8.19 ± 0.82</td>
</tr>
</tbody>
</table>

**NormFinder analysis**

Similar to that in BestKeeper analysis, a lower Ct value indicates a higher stability of gene expression. As shown in Table 3, EF-1α and GAPDH displayed the highest expression stability, showing the lowest Ct values under H$_2$O$_2$ and PEG stress. ACT showed the highest expression stability in the NaCl and UV stress subsets. In the CuSO$_4$ and Cold...
subsets, \textit{YLS8} was the most stable reference gene. \textit{TIP41} had the highest stability in the high temperature and wild type subsets. \textit{NCPB2} and \textit{PP2A2} showed the lowest expression stability under the different stress conditions. Among all sample subsets, \textit{YLS8} was the most stable reference gene, followed by \textit{TIP41} and \textit{CYP2}. Interestingly, in the overall ranking, the top three genes from the BestKeeper and NormFinder analyses were consistent.

Table 3

<table>
<thead>
<tr>
<th>Rank</th>
<th>\textit{H}_2\textit{O}_2</th>
<th>\textit{NaCl}</th>
<th>\textit{CuSO}_4</th>
<th>high-temperature</th>
<th>low-temperature</th>
<th>PEG</th>
<th>UV</th>
<th>WT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>\textit{EF-1\alpha}</td>
<td>\textit{ACT}</td>
<td>\textit{YLS8}</td>
<td>\textit{TIP41}</td>
<td>\textit{YLS8}</td>
<td>\textit{GAPDH}</td>
<td>\textit{ACT}</td>
<td>\textit{TIP41}</td>
<td>\textit{YLS8}</td>
</tr>
<tr>
<td></td>
<td>0.090</td>
<td>0.072</td>
<td>0.075</td>
<td>0.083</td>
<td>0.080</td>
<td>0.045</td>
<td>0.039</td>
<td>0.034</td>
<td>0.242</td>
</tr>
<tr>
<td>2</td>
<td>\textit{YLS8}</td>
<td>\textit{GAPDH}</td>
<td>\textit{PTBP}</td>
<td>\textit{ACT}</td>
<td>\textit{PTBP}</td>
<td>\textit{YLS8}</td>
<td>\textit{PTBP}</td>
<td>\textit{YLS8}</td>
<td>\textit{TIP41}</td>
</tr>
<tr>
<td></td>
<td>0.102</td>
<td>0.118</td>
<td>0.075</td>
<td>0.117</td>
<td>0.080</td>
<td>0.069</td>
<td>0.257</td>
<td>0.099</td>
<td>0.348</td>
</tr>
<tr>
<td>3</td>
<td>\textit{GAPDH}</td>
<td>\textit{YLS8}</td>
<td>\textit{GAPDH}</td>
<td>\textit{GAPDH}</td>
<td>\textit{TIP41}</td>
<td>\textit{ACT}</td>
<td>\textit{GAPDH}</td>
<td>\textit{CYP2}</td>
<td>\textit{CYP2}</td>
</tr>
<tr>
<td></td>
<td>0.232</td>
<td>0.167</td>
<td>0.115</td>
<td>0.137</td>
<td>0.122</td>
<td>0.080</td>
<td>0.316</td>
<td>0.278</td>
<td>0.365</td>
</tr>
<tr>
<td>4</td>
<td>\textit{TIP41}</td>
<td>\textit{TIP41}</td>
<td>\textit{TIP41}</td>
<td>\textit{EF-1\alpha}</td>
<td>\textit{EF-1\alpha}</td>
<td>\textit{YLS8}</td>
<td>\textit{PTBP}</td>
<td>\textit{ACT}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.289</td>
<td>0.225</td>
<td>0.115</td>
<td>0.270</td>
<td>0.227</td>
<td>0.170</td>
<td>0.415</td>
<td>0.210</td>
<td>0.420</td>
</tr>
<tr>
<td>5</td>
<td>\textit{ACT}</td>
<td>\textit{PTBP}</td>
<td>\textit{EF-1\alpha}</td>
<td>\textit{YLS8}</td>
<td>\textit{ACT}</td>
<td>\textit{CYP2}</td>
<td>\textit{EF-1\alpha}</td>
<td>\textit{CYP2}</td>
<td>\textit{EF-1\alpha}</td>
</tr>
<tr>
<td></td>
<td>0.342</td>
<td>0.333</td>
<td>0.135</td>
<td>0.289</td>
<td>0.334</td>
<td>0.274</td>
<td>0.427</td>
<td>0.278</td>
<td>0.420</td>
</tr>
<tr>
<td>6</td>
<td>\textit{CYP2}</td>
<td>\textit{CYP2}</td>
<td>\textit{CYP2}</td>
<td>\textit{PTBP}</td>
<td>\textit{CYP2}</td>
<td>\textit{TIP41}</td>
<td>\textit{TIP41}</td>
<td>\textit{EF-1\alpha}</td>
<td>\textit{PTBP}</td>
</tr>
<tr>
<td></td>
<td>0.365</td>
<td>0.358</td>
<td>0.194</td>
<td>0.320</td>
<td>0.360</td>
<td>0.408</td>
<td>0.466</td>
<td>0.443</td>
<td>0.778</td>
</tr>
<tr>
<td>7</td>
<td>\textit{PTBP}</td>
<td>\textit{EF-1\alpha}</td>
<td>\textit{ACT}</td>
<td>\textit{CYP2}</td>
<td>\textit{GAPDH}</td>
<td>\textit{PP2A}</td>
<td>\textit{CYP2}</td>
<td>\textit{GAPDH}</td>
<td>\textit{GAPDH}</td>
</tr>
<tr>
<td></td>
<td>0.382</td>
<td>0.367</td>
<td>0.329</td>
<td>0.524</td>
<td>0.413</td>
<td>1.187</td>
<td>0.480</td>
<td>0.599</td>
<td>0.928</td>
</tr>
<tr>
<td>8</td>
<td>\textit{PP2A}</td>
<td>\textit{PP2A}</td>
<td>\textit{PP2A}</td>
<td>\textit{PP2A}</td>
<td>\textit{PP2A}</td>
<td>\textit{PP2A}</td>
<td>\textit{PP2A}</td>
<td>\textit{NCBP2}</td>
<td>\textit{PP2A}</td>
</tr>
<tr>
<td></td>
<td>0.747</td>
<td>0.543</td>
<td>0.901</td>
<td>0.790</td>
<td>0.599</td>
<td>1.624</td>
<td>0.680</td>
<td>1.042</td>
<td>1.006</td>
</tr>
<tr>
<td>9</td>
<td>\textit{NCBP2}</td>
<td>\textit{PP2A2}</td>
<td>\textit{NCBP2}</td>
<td>\textit{NCBP2}</td>
<td>\textit{NCBP2}</td>
<td>\textit{PTBP}</td>
<td>\textit{NCBP2}</td>
<td>\textit{PP2A}</td>
<td>\textit{NCBP2}</td>
</tr>
<tr>
<td></td>
<td>1.390</td>
<td>1.282</td>
<td>1.561</td>
<td>1.135</td>
<td>1.233</td>
<td>2.121</td>
<td>1.367</td>
<td>1.335</td>
<td>1.436</td>
</tr>
<tr>
<td>10</td>
<td>\textit{PP2A2}</td>
<td>\textit{NCBP2}</td>
<td>\textit{PP2A2}</td>
<td>\textit{PP2A2}</td>
<td>\textit{PP2A2}</td>
<td>\textit{NCBP2}</td>
<td>\textit{PP2A2}</td>
<td>\textit{PP2A2}</td>
<td>\textit{PP2A2}</td>
</tr>
<tr>
<td></td>
<td>1.668</td>
<td>1.420</td>
<td>2.302</td>
<td>2.161</td>
<td>1.679</td>
<td>2.330</td>
<td>1.431</td>
<td>1.614</td>
<td>1.745</td>
</tr>
</tbody>
</table>

\textbf{geNorm analysis}

As shown in Fig. 2, a lower average M value indicates a higher gene expression stability. Moreover, a gene with M value > 1.5 was considered unsuitable as a reference gene. In all treatment subsets, the top three reference genes
all showed excellent stability with low M values. YLS8 was the most stable reference gene with the lowest M values in all treatment subsets, except for the high temperature subset. PTBP also presented highest expression stability under all stress conditions, except for PEG stress. TIP41 showed the least fluctuation of M value in all stress subsets, and maintained the highest gene expression stability in the total and control subsets. On the contrary, PP2A, PP2A2, and NCBP2 had the lowest gene expression stability under all stress conditions, indicating that they are the least suitable as reference genes. This result is consistent with that of BestKeeper and NormFinder analysis.

In addition to comparing and ranking the expression stability of the reference genes, geNorm can also optimize the number of reference genes by using pairwise variation \( V_{n/n+1} \) as a parameter. A threshold value of 0.15 was used to determine whether one more reference gene should be added for normalization. A value of < 0.15 indicates that an additional reference gene has no notable significance for normalization. In this study, we calculated the \( V_{n/n+1} \) value by using geNorm (Fig. 3 and Table S3). In all stress condition except low temperature, two reference genes were sufficient for normalization, with \( V_{2/3} < 0.15 \). In the low-temperature subset, the \( V_{3/4} \) value was below 0.15, which suggested that three reference genes should be applied for normalization.

**Comprehensive analysis of reference gene validation**

RefFinder was used to systematically select the optimal reference genes under different treatment conditions. By comparing the results of the above three algorithms with the \( \Delta Ct \) method, RefFinder can assess and screen out the most stable reference genes [40]. The results are listed in Table S4. In addition, all gene pairs were compared with each other and ranked according to the value of \( \Delta Ct \), from low to high, as shown in Table S5. According to these results, TIP41 showed high stability in most treatment groups, which was consistent with the results of geNorm analysis.

Next, the geometric means of the three algorithms were calculated according to the ranks of the reference genes. TIP41, CYP2, and YLS8, ranked top three in all algorithms. In contrast, PP2A, NCBP2, and PP2A2, which suggested that these genes were not suitable for use as reference genes under different stress conditions (Fig. 4). Table S6 shows the ranking of the reference genes in all four algorithm tools revealing the genes that had stable expression under different stress conditions.

**Discussion**

In this study, among 10 candidate genes in *P. corylifolia*, the genes that show expression stability under various abiotic stress conditions were selected as the most suitable reference genes for use in RT-qPCR analyses. First, we verified the primer specificity of the candidate reference genes and the PCR experimental conditions. A single peak in the dissolution curve confirmed the specificity of the primers (Fig. S1). Moreover, according to the slope of the standard curve, all gene primers, except for PP2A and PP2A2 primers, had high PCR efficiency, with their R values are higher than 0.99 (Table 1). For these genes, there was a good linear relationship between Ct value and diluted cDNA concentration, as well as acceptable PCR conditions [40]. As a qualified reference gene, its expression level should be within a reasonable range under various stress treatments. As illustrated in Fig. 1, the mean Ct values for the 10 candidates ranged from 9.95 to 33.16; the majority of these values were found to be between 18 and 24 in all tested samples, which meant that most candidates were likely to provide accurate normalization [41]. Low Ct values correspond to high expression levels and a narrow Ct distribution range implies lower variability and higher
stability. Thus, NCBP1 and PP2A2 may not be suitable as a reference gene. In contrast to PP2A, PP2A presented a narrow range of very low Ct values.

To systematically evaluate and compare the expression stability of the candidate genes, we used three Excel-based mathematical tools to process the raw data. Although the results of all three methods were reasonable, they were not completely consistent. The main reason is that they depend on different algorithms [19], therefore, another software RefFinder was applied to conduct a comprehensive analyze and rank the results of the three algorithms to increase the credibility of the results. Interestingly, the three algorithms consistently showed the same three genes as the most stable genes in the high temperature, PEG, and control subsets. Although BestKeeper revealed TIP41 as the most stable gene in most subsets (Table 2), NormFinder showed that TIP41 had good stability in only three subsets (Table 3), this difference was mainly due to the different algorithms used to process the data. YLS8 also showed good stability in most stress conditions suggesting its potential as an appropriate reference gene.

In this study, GAPDH exhibited the highest stability under PEG stress, indicating that it can be a reliable reference gene under drought conditions. Although GAPDH has been widely used as a reference gene in different plants such as Linum usitatissimum L. and Coffea arabica [42, 43], several other studies have also shown that GAPDH exhibited instability under different conditions [44, 45], which may be partly due to the involvement of GAPDH in various cellular processes besides glycolysis [46].

Our results also showed that EF-1α, and ACT presented high stability in individual groups. In Arabidopsis, the expression of EF-1α was relatively stable [47]. However, under nutritional deficiencies or abiotic stresses, EF-1α shows poor expression stability [48]. ACT has been successfully applied to study the effects of low temperature and light in Festuca pratensis [49]. Considering that the average relative expression level of genes obtained using multiple internal reference genes is less variable than that obtained using a single internal reference gene [50], RT-qPCR with multiple reference genes may yield more stable and reliable results [22]. The number of reference genes used for normalization can be analyzed based on the paired variation value (V\textsubscript{n}/V\textsubscript{n+1}) calculated using geNorm software. Generally, once the V\textsubscript{n}/V\textsubscript{n+1} value is above the suggested threshold value of 0.15, an additional reference gene must be introduced to calibrate the experimental results. As shown in Table S3, only the low-temperature subset required three reference genes, whereas the other subsets needed only two reference genes for normalization. NCBP2, PP2A, and PP2A2, showed the lowest stability in all stress conditions, according to the ranking of the three algorithms, indicating that they were not suitable reference genes for P. corylifolia gene research. Although PP2A showed a high expression level with low Ct value and was proven to be stable in some studies [51], another report showed its instability under several environments as well as tissue expression specificity [52]. Indeed, the use of a single, untested reference gene to standardize RT-qPCR results is usually not acceptable [24]. Thus, RefFinder analysis was conducted to evaluate the ranking generated by the three algorithms, and the result was similar to that of geNorm analysis. The final ranking is presented in Fig. 4, with TIP41, CYP2, and YLS8 shown as the top three reference genes. Therefore, we recommended using these three genes as reference genes for RT-qPCR normalization in the study of P. corylifolia gene expression profile.

**Conclusion**

Comprehensive analysis of BestKeeper, NormFinder, and geNorm results revealed TIP41, CYP2, and YLS8 as the reference genes with the most stable expression under all stress conditions tested, as well as PP2A, NCBP2, and
PP2A2 as the least stable genes. geNorm analysis also showed that two stable reference genes were sufficient for normalization under most treatment conditions. This is the first study on the screening of reference genes in *P. corylifolia*. Our findings not only lays a foundation for future research but also has a profound significance for guiding the screening of appropriate reference genes in other plant species.

**Abbreviations**

ACT: actin; Ct: cycle threshold; CV: coefficient of variation; CYP2: cyclophilin 2; EF-1α: elongation factor 1-alpha; FPKM: fragments per kilobase of exon model per million mapped fragments; GAPDH: glyceraldehyde 3-phosphate dehydrogenase; NCBP2: nuclear cap-binding protein subunit 2; PP2A: protein phosphatase 2A subunit A3; PP2A2: protein phosphatase 2A-2; PTBP: polypyrimidine tract-binding protein; RT-qPCR: Reverse transcription quantitative polymerase chain reaction; TIP41: TIP41-like family protein; YLS8: thioredoxin-like protein YLS8.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

All authors have consented for publication.

**Availability of data and materials**

The gene sequence information mentioned in this study could be found from the literature based on the gene list in Additional Table S2. The raw data can be accessed from the NCBI Sequence Read Archive (SRA) platform under the accession number PRJNA533929. The materials are available from the corresponding author on reasonable request after the publication of the work.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This work, the materials planting, sample collection was supported by the Practical Training Program for Natural Science Foundation of the Jiangsu Higher Education Institutions of China (16KJB150042 and 19KJD430005). The RT-qPCR experiment was supported by the Jiangsu “Six one” Project of Health top talent program (LGY2018089), and Innovation and entrepreneurship training program for college students in Jiangsu Province (201714255009Y, 202014255013H, and 202014255037T). The data collection, analysis and interpretation was supported by the Qinglan project of excellent teaching team in Jiangsu and teaching and research project of Jiangsu Health
Vocational College (JKA201902, JKB201911, JKA202004, JKB202002, JKB202005, and JKB202008). The funders had no other role during the experiment and preparation of the manuscript.

Authors’ contributions

JW, YDH and YZ designed the experiments, YZ, YNG and JJG performed the experiments, WFN, ZZB and WQW analyzed the data, XQC, YNG and JJG contributed reagents, materials, and analysis tools. YZ, YDH wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Editage (www.editage.cn) for English language editing.

References


**Figures**
Figure 1

The cycle threshold (Ct) values of 10 candidate reference genes in P. corylifolia. Lines across the Ct values box chart represent the median values. The boxes indicate the 25th and 75th percentiles, and the whisker caps indicate the maximum and minimum values.
Figure 2

The mean expression stability (M) of 10 candidate reference genes sorted using geNorm. The stability of gene expression increased from left to right, the most stable gene owned the lowest m value, which was distributed on the right.
Figure 3

The pairwise variation values of 10 candidates calculated using geNorm. Different treatment conditions are identified by different colors. The cut-off value of the optimized reference gene numbers for RT-qPCR normalization is 0.15.
Figure 4

Comprehensive analysis rankings of BestKeeper, NormFinder, and geNorm. The geomean ranks of different software are presented by polylines with different colors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx