This is an open access article distributed under the terms of the Creative Common Common BY-NC-ND Licence

Transcriptome-based validation of proper reference genes for reverse trascription quantitative PCR analysis of *Sinocalycanthus chinensis*

C. ZHANG, Q.-N. JIANG, Y.-G. WANG, J.-X. FU, B. DONG, L.-H. ZHOU*, and H.-B. ZHAO*

Zhejiang Provincial Key Laboratory of Germplasm Innovation and Utilization for Garden Plants, Key Laboratory of National Forestry and Grassland Administration on Germplasm Innovation and Utilization for Southern Garden Plants, School of Landscape Architecture, Zhejiang Agriculture and Forestry University, Hangzhou 311300, Zhejiang, P.R. China

Abstract

Reverse transcription quantitative PCR is a widely used method to detect gene expressions. To obtain accurate expression results, the selection of proper reference genes is important and necessary. However, related works concerning reference gene selection have not been carried on many plant species, especially endangered ones. The aim of the present study was to select dependable reference genes for expression normalization of an endangered plant species with medicinal and ornamental value: Sinocalycanthus chinensis (Calycanthaceae). Nine reference genes with stable expressions were chosen for further analysis according to transcriptomic sequencing data from S. chinensis. The expression stability of these candidate genes in inner and outer petals at different floral developmental stages and in many different tissues was then analyzed with the geNorm and NormFinder software. The reference genes were evaluated by normalizing the expression of the anthocyanidin synthase gene in the outer and inner petals at different floral developmental stages to further verify the expression stability of these genes. Elongation factor 1-alpha (ScEF) and 50S ribosomal protein L27 were found to be the two most stable genes in the overall ranking of all the samples and different tissues. Furthermore, ScEF and protein phosphatase 2A were stably expressed in all petal samples. Moreover, among the nine candidate reference genes, phosphoglycerate kinase performed poorly in all sample sets. Our results will help to obtain reliable expression data in molecular studies of S. chinensis.

Additional key words: anthocyanidin synthase, floral developmental stages, geNorm, NormFinder, petals.

Introduction

Sinocalycanthus chinensis is an endangered plant species in China that has high medicinal and ornamental value (Cheng and Chang 1964). S. chinensis bears large creamy white flowers with a whorl of yellow inner petals, sometimes with pink to purplish around the edges of the outer petals.

Plant petal coloration is caused by the accumulation of pigments such as flavonoids, carotenoids, or betacyanins (Tanaka *et al.* 2008). Biosynthesis of related pigments is likely responsible for the difference in petal coloration between the outer petals and inner petals. However, at present, the factors that regulate the petal coloration of

S. chinensis are still unclear. Therefore, investigating the expression of genes involved in the biosynthesis of related pigments will help elucidate the underlying mechanisms of petal coloration in *S. chinensis*.

Reverse transcription quantitative PCR (RT-qPCR) is the most reliable technology to detect gene expression because of its high accuracy and sensitivity (Gachon et al. 2004, Bustin et al. 2005). With respect to RT-qPCR, a suitable reference gene is crucial for gene expression analysis. Actin (ACT), beta-tubulin (TUB), 18S ribosomal RNA (18S), and other traditional genes are often employed as reference genes. However, the expression performance of reference genes is quite different in different plant species or under different experimental conditions. For

Submitted 5 September 2019, last revision 17 January 2020, accepted 28 January 2020.

Abbreviations: ACT - actin; ANS - anthocyanidin synthase; Cp - crossing point; EF - elongation factor 1-alpha; EIF - eukaryotic initiation factor; M value - gene expression stability measure; PGK - phosphoglycerate kinase; PP2A - protein phosphatase 2a; PV - pairwise variation; RPL - 50S ribosomal protein L27; RT-qPCR - reverse transcription quantitative PCR; TUB - beta-tubulin; UBC - ubiquitin-conjugating enzyme.

Acknowledgements. This study was supported by the Key Research and Development Program of Zhejiang Province (2019C02036), the National Natural Science Foundation of China (Grant No.31401902 and 31101571), and Zhejiang Provincial Major Program of New Cultivar Breeding (Grant No. 2016C02056-12). The first two authors contributed equally to this work.

^{*} Corresponding authors; e-mails: zhoulihua@zafu.edu.cn, zhaohb@zafu.edu.cn

instance, *18S* is considered a dependable reference gene in developing *Rhododendron molle* flowers (Xiao *et al.* 2016), but this gene is not recommended as a suitable reference gene in *Osmanthus fragrans* (Zhang *et al.* 2015).

To obtain accurate expression results, the selection of proper reference genes under specific research conditions is essential for the normalization of RT-qPCR data. Therefore, in this study, based on the expression data of four transcriptomes of S. chinensis, nine reference genes whose expression changed little were chosen: S. sinensis actin 1 (ScACT1), ScACT2, elongation factor 1-alpha (ScEF), eukaryotic initiation factor (ScEIF), phosphoglycerate kinase (ScPGK), protein phosphatase 2a (ScPP2A), 50S ribosomal protein L27 (ScRPL), ScTUB, and ubiquitin-conjugating enzyme (ScUBC). The expression performance of these nine genes was then analyzed in different developmental stages of petals and different tissues. The geNorm (Vandesompele et al. 2002) and NormFinder (Andersen et al. 2004) were employed to evaluate expression stability in different experimental sample sets of S. chinensis. After stability assessment, the reference genes with good or poor performance were further used to normalize the expression of an S. chinensis anthocyanin synthase gene, ScANS, in the outer and inner petals at different floral developmental stages.

Materials and methods

Plants: Sinocalycanthus chinensis W.C. Cheng & S.Y. Chang (Calycanthaceae) plants that were grown in the Wintersweet Garden of Zhejiang Agriculture and Forestry University were used for sample collection. The flower opening of S. chinensis could be divided into six developmental stages: stage 1 (S1; tight buds, 5 - 7 mm in diameter), S2 (tight buds, 12 - 15 mm in diameter), S3 (soft bud stage), S4 (initial flowering stage), S5 (half flowering stage), and S6 (full flowering stage). The outer and inner petals of flowers at six developmental stages were collected and constituted the samples of all petal sample sets. Additionally, ten different tissue types were collected: pedicels, sepals, outer petals, inner petals, both the androecia and gynoecia from flowers at S6, young leaves, mature leaves, young stems, and mature stems. In total, 22 experimental samples comprised 12 petal samples and ten different tissue samples.

Extraction of RNA and first-strand cDNA synthesis: Total RNA was extracted from all samples using a *MiniBEST* universal RNA extraction kit (TaKaRa, Dalian, China). The concentration and quality of RNA samples were measured using a $NanoDrop\ ND-1000$ spectrophotometer ($NanoDrop\ Technologies$, Rockland, DE, USA). Samples of RNA with absorbance ratios A_{260}/A_{280} of 1.8 - 2.2 and $A_{260}/A_{230} > 1.8$ were used for cDNA synthesis. One microgram of total RNA was used to synthesize the first-strand cDNA with the reverse transcriptase $M-MLV\ (TaKaRa)$.

Selection of candidate reference genes: Four cDNA

libraries were constructed from the outer and inner petals of *S. chinensis* at S2 and S4 stages and sequenced *via* the *Illumina HiSeq 2000* platform. In total, 125 014 unigenes were assembled and identified. Additionally, the FPKM (fragments per kilobase per million reads) method was used to calculate the unigene expression. The differences in gene expressions among the samples were determined using the FPKM value. Nine reference genes (*ScACT1*, *ScACT2*, *ScEF*, *ScEIF*, *ScPGK*, *ScPP2A*, *ScRPL*, *ScTUB*, and *ScUBC*) were chosen as candidate reference genes in *S. chinensis* because their expression was stable (1/1.5 < fold change <1.5) between any two samples (Table 1 Suppl.).

Amplification efficiency test of primers: The primers for the nine candidate reference genes with amplicon lengths of 87 - 180 bp are shown in Table 2 Suppl. The amplification efficiency of each primer pair was tested as follows: the slope of the standard curve obtained by serial 5-fold dilutions of the cDNA samples was used to calculate the amplification efficiency as:

 $E [\%] = [-1 + 10 (-1/slope)] \times 100.$

RT-qPCR was performed using a *LightCycler 480 II* (*Roche*, town?, Switzerland) with *SYBR Premix Ex Taq* (*TaKaRa*). The reaction mixture (20 mm³ total volume) consisted of 10 mm³ of *SYBR Premix Ex Taq*, 0.8 mm³ of each primer (10 μM), 2 mm³ of cDNA (~50 ng), and 6.4 mm³ of ddH₂O. The amplification was performed under the following conditions: an initial denaturation step at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. Afterward, a melting curve was obtained. Crossing point (Cp) values of the triplicate reactions were then obtained.

Expression stability analysis: The expressions of the nine genes were determined according to Cp values, and the expression stability was analyzed by *geNorm* (v. 3.5) and *NormFinder* (v. 0.953) algorithms. Before their input into this two software programs, the raw Cp values were transformed into relative quantities (Q values) using the following formula: $Q = 2^{-\Delta Cp} (\Delta Cp = Cp \text{ value of each sample - the minimum Cp value).$

The *geNorm* software determines gene expression stability based on the gene expression stability measure (M value). Additionally, pairwise variation (PV) values were used to choose the optimal number of reference genes (Vandesompele *et al.* 2002). The *NormFinder* software ranks the reference genes by the expression stability of the candidate genes in a certain sample set (Andersen *et al.* 2004).

Expression normalization of ScANS: As a target gene, ScANS was employed to indicate the usefulness of the selected candidate reference genes. The expression of ScANS in the outer and inner petals at different developmental stages was quantified with the three most stable reference genes (ScEF, ScPP2A, ScRPL), the least stable reference gene (ScPGK) and a combination of reference genes (ScEF+ScPP2A) based on the

overall ranking calculated by *geNorm* and *NormFinder*. Primers (5'-GTGTTCGGGCTGTCGGTAAG-3' and 5'-ACTGGCATTGTTGGCGAGCT-3') were used to amplify *ScANS*.

Results

Based on the melting curve, only a single peak was observed in the template, while there was no amplicon in the negative control for each gene pair (Fig. 1 Suppl.). In addition, the PCR efficiencies varied from 95.4 % for *ScPP2A* to 101.0 % for *ScTUB*, and the R² values ranged from 0.9975 for *ScEIF* to 0.9997 for *ScEF* (Table 2 Suppl.), which indicates that the primers could be used for further gene expression stability with RT-qPCR.

A box plot (Fig. 1) was constructed that gives an overview of the raw Cp values of the nine candidate reference genes in the tested samples. The raw Cp values ranged from 12.51 to 25.85 (Fig. 1). The *ScPP2A* showed the highest average Cp value (21.25), followed by *ScRPL*, and *ScEF* exhibited the lowest mean Cp value (15.23). Furthermore, the Cp value variation was quite different for each gene (Fig. 1). For example, the Cp value variation of *ScPP2A* was below 6 cycles with stable gene expression, while that of *ScPGK* was above 9 cycles with obvious expression variation.

The *geNorm* ranked the candidate genes according to their M values because the genes showed the most stable expression with the lowest M values (Table 3 Suppl.). The M value of 1.5 is usually recommended as a threshold to identify stable reference genes (Vandesompele *et al.* 2002). All M values of the nine reference genes were lower than 1.5 in the four sample sets (Table 3 Suppl.), revealing that all nine genes met the basic reference gene requirements.

The ScEF and ScRPL were the most stable genes with the lowest M value (0.534; Table 3 Suppl.) for all 22 samples, followed by ScPP2A, with the M value of 0.619. Similarly, ScEF and ScRPL were also the most stable genes in the different tissue samples, followed by ScPP2A (Table 3 Suppl.). In the petal sample set, ScRPL and ScPP2A were the most stable reference genes, followed by ScEF. Generally, across all sample sets, ScEF, ScRPL, and ScPP2A were the three most stable reference genes. In addition, ScPGK, with the highest M value in all sample sets, was considered the least stable gene (Table 3 Suppl.).

A pairwise variation (V_n/V_{n+1}) of 0.15 is recommended as a threshold to determine the optimal number of reference genes. The V2/3 values of all petals were less than 0.15 (Fig. 2), indicating that the top two reference genes were sufficient for expression normalization. For all samples and different tissues, the V2/3 and V3/4 values were higher than 0.15, but the V4/5 values were lower than 0.15, suggesting that a fifth reference gene was not required to normalize the gene expression and that the top four reference genes were needed for accurate normalization.

The *ScEF* was the most stable gene ranked by *NormFinder* for all samples, including all petal samples and different tissue samples (Table 4 Suppl.). The ranking order of gene stability calculated by *geNorm* and

NormFinder was the same in the sample set of different tissues and was similar in all petal samples. Regarding the least stable gene, ScPGK was ranked last out of the nine candidate reference genes in all sample sets, which is consistent with the results determined by geNorm.

The overall ranking with the aggregate results calculated by *geNorm* and *NormFinder* (Table 5 Suppl.) showed that *ScEF* and *ScRPL* were the most stable genes in the overall ranking of all the samples and different tissues. In addition, *ScEF* and *ScRPL* were the most stably expressed genes based on the overall ranking calculated by *geNorm* and *NormFinder* in all petal samples. On the other hand, *ScPGK* was the least stable gene out of all reference genes in all the sample sets.

The three most stable reference genes (ScEF, ScPP2A, and ScRPL), the least stable reference gene (ScPGK) and a combination of reference genes (ScEF+ScPP2A) were used as internal controls for normalization of ScANS according to the overall ranking (Table 5 Suppl.) calculated by geNorm and NormFinder. The expression

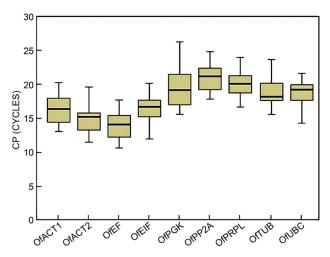


Fig. 1. Expression profiles of nine candidate reference genes from 22 samples. The expression data are displayed as Cp (crossing point) values for each reference gene in all samples. The line across the box is the median, n=22. The *boxes* indicate the 25/75 percentiles. The *whisker caps* indicate the minimum and maximum values.

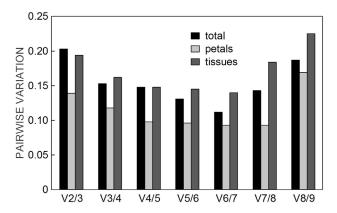


Fig. 2. Pairwise variation (PV, $V_{n/n+1}$) analysis of nine candidate genes in three sample sets by *geNorm*.

patterns of *ScANS* in the outer petals and inner petals at different developmental stages were similar when *ScEF*, *ScPP2A*, *ScRPL*, and a combination of reference genes (*ScEF+ScPP2A*) were used for normalization (Fig. 3). However, the expression of *ScANS* with *ScPGK* as an internal control was greatly different, revealing that *ScPGK* performed poorly as a reference gene in all sample sets.

Discussion

Transcriptome sequencing is an excellent technology that can help obtain additional gene information and reveal the complexity of the transcriptomes of many species (Ma et al. 2016, Ning et al. 2016, Zhang et al. 2016). Nine reference genes whose expression changed little in *S. chinensis* transcriptomes were chosen as candidate reference genes in the present study. The expression of these candidate reference genes was tested in outer petals, inner petals, and all petals at different floral developmental stages and in different tissues by RT-qPCR and evaluated using statistical algorithms.

Among the different statistical algorithms, geNorm (Vandesompele et al. 2002) was the first choice to be used for the evaluation of the stability of candidate reference genes, followed by NormFinder (Andersen et al. 2004). In our study, the gene stability rankings generated by geNorm and NormFinder were the same in the sample set of different tissues; however, they were inconsistent in terms of the total samples and all petals (Table 3 Suppl. and Table 4 Suppl.), which can be explained by the fact that the two algorithms are based on different principles (Vandesompele et al. 2002, Andersen et al. 2004). However, the last ranking order by the two methods was identical in all sample sets. Similarly, based on the rank, the most unstable gene was also the same in many other plant species when different statistical algorithms were used (Fu et al. 2013, Zhang et al. 2015, Xiao et al. 2016). In addition, the two most stable genes generated by the two methods were identical in most sample sets.

With geNorm software, a cutoff PV value of 0.15 is recommended (Vandesompele et al. 2002). The top four reference genes were required for expression normalization in all samples and different tissues (Fig. 2). In fact, expression normalization with additional reference genes could generate better results, and a PV value of 0.15 is not a necessary criterion (Vandesompele et al. 2002). In previous studies, although geNorm recommended that more than two genes were needed for optimal results, the top two reference genes were sufficient to obtain accurate expression of interesting genes in Salix matsudana (Chen et al. 2017) and Sorghum bicolor (Sudhakar Reddy et al. 2016). Similarly, in our study of S. chinensis, the top two reference genes were sufficient for accurate normalization in all samples and different tissues. In other words, ScEF and ScRPL were proper reference genes used for accurate normalization in all samples and different tissues.

ScEF was an optimal reference gene for all samples, all petals, and different tissues based on the overall rankings calculated by geNorm and NormFinder (Table 5 Suppl.). ScEF is a homolog of the EF gene. The EF is the most stably expressed gene in different tissues of Chinese cabbage (Qi et al. 2010) and tobacco (Schmidt and Delaney 2010). In addition, EF is considered the most stable gene not only in flower development but also in the leaf development of 'Mitchell' petunia (Mallona et al. 2010) and during the different fruit developmental stages of Litsea cubeba (Lin et al. 2013).

The *PP2A* was another optimal reference gene for all petals. In cineraria, the expression of *PP2A* was the least variable in developing flowers (Jin *et al.* 2013). In addition, *PP2A* was the third most suitable reference gene for different tissues in *S. chinensis*, and *PP2A* was the most stable gene in all the sample sets of *Sorghum bicolor*, including in different tissues (Sudhakar Reddy *et al.* 2016).

In our study, *ScRPL* was recommended as one of the optimal reference genes for all samples, inner petals and different tissues. Ribosomal proteins are integral to ribosome biogenesis and function in protein synthesis (Szakonyi and Byrne 2011). Thus, homologs of the *RPL*

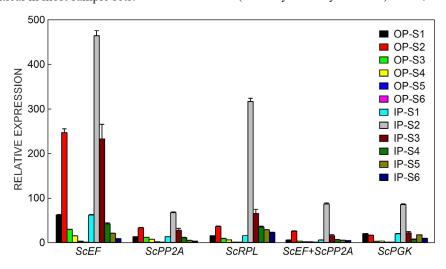


Fig. 3. Relative quantification of ScANS expression in outer and inner petals at different developmental stages using different reference genes (ScEF, ScPP2A, ScRPL, and ScPGK) and a combination of reference genes (ScEF+ScPP2A). Means \pm SDs, n = 3.

gene are often found to serve as suitable reference genes in other plant species, such as wheat (Liu et al. 2014), tomato (Løvdal and Lillo 2009), *Theobroma cacao* (Pinheiro et al. 2011), and *Rhododendron molle* (Xiao et al. 2016). In developing flowers and different tissues of *R. molle*, the *RPL* gene was recommended for the normalization of RT-qPCR data (Xiao et al. 2016).

In conclusion, our research reports on the transcriptome-based validation of proper reference genes for RT-qPCR expression normalization in petals at different floral developmental stages and in different tissues of *S. chinensis*. *geNorm* and *NormFinder* were employed to determine the most suitable reference genes. Overall, *ScEF* and *ScRPL* were the two most stable genes in the overall ranking of total samples and different tissues, and *ScEF* and *ScPP2A* were the most stably expressed genes in all petal samples. *ScPGK* performed poorly in all sample sets.

References

- Andersen, C., Jensen, J., Ørntoft, T.: Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. - Cancer Res. 64: 5245-5250, 2004.
- Bustin, S.A., Benes, V., Nolan, T., Pfaffl, M.W.: Quantitative real-time RT-PCR a perspective. J. mol. Endocrinol. **34**: 597-601, 2005.
- Chen, Y., Cui, P., Ren, X., Zhang, C., Jin, C., Zhang, Z., Shih, C.K.: Fabrication of MoSe2 nanoribbons via an unusual morphological phase transition. - Nat. Commun. 8: 15135, 2017.
- Cheng, W.C., Chang, S.Y.: Genus novum Calycanthacearum chinae orientalis. Acta phytotaxon. sin. 9: 137-138, 1964.
- Fu, J., Wang, Y., Huang, H., Zhang, C., Dai, S.: Reference gene selection for RT-qPCR analysis of *Chrysanthemum lavandulifolium* during its flowering stages. - Mol. Breed. 31: 205-215, 2013.
- Gachon, C., Mingam, A., Charrier, B.: Real-time PCR: what relevance to plant studies? - J. exp. Bot. 55: 1445-1454, 2004.
- Jin, X., Fu, J., Dai, S., Sun, Y., Hong, Y.: Reference gene selection for qPCR analysis in cineraria developing flowers. Sci. Hort. **153**: 64-70, 2013.
- Lin, L., Han, X., Chen, Y., Wu, Q., Wang, Y.: Identification of appropriate reference genes for normalizing transcript expression by quantitative real-time PCR in *Litsea cubeba*. Mol. Genet. Genom. **288**: 727-737, 2013.
- Liu, Y.K., Li, H.P., Huang, T., Cheng, W., Gao, C.S., Zuo, D.Y., Zhao, Z.X., Liao, Y.C.: Wheat-specific gene, *ribosomal protein L21*, used as the endogenous reference gene for qualitative and real-time quantitative polymerase chain reaction detection of transgenes. J. Agr.Food Chem. **62**: 10405-10413, 2014.

- Løvdal, T., Lillo, C.: Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. - Anal. Biochem. 387: 238-242, 2009.
- Ma, R., Xu, S., Zhao, Y., Xia, B., Wang, R.: Selection and validation of appropriate reference genes for quantitative realtime per analysis of gene expression in *Lycoris aurea*. - Front. Plant Sci. 7: 536, 2016.
- Mallona, I., Lischewski, S., Weiss, J., Hause, B., Egeacortines,
 M.: Validation of reference genes for quantitative real-time
 PCR during leaf and flower development in *Petunia hybrida*.
 BMC Plant Biol. 10: 4, 2010.
- Ning, H., Zhang, C., Fu, J., Fan, Y.: Comparative transcriptome analysis of differentially expressed genes between the curly and normal leaves of *Cymbidium goeringii* var. *longibracteatum*. -Genes Genom. 38: 985-998, 2016.
- Pinheiro, T.T., Litholdo Jr., C.G., Sereno, M.L., Leal Jr., G.A., Albuquerque, P.S.B., Figueira, A.: Establishing references for gene expression analyses by RT-qPCR in *Theobroma cacao* tissues. Genet. mol. Res. **10**: 3291-3305, 2011.
- Qi, J., Yu, S., Zhang, F., Shen, X., Zhao, X., Yu, Y., Zhang, D.: Reference gene selection for real-time quantitative polymerase chain reaction of mRNA transcript levels in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). Plant mol. Biol. Rep. 28: 597-604, 2010.
- Schmidt, G.W., Delaney, S.K.: Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. Mol. Genet. Genom. **283**: 233-241, 2010.
- Sudhakar Reddy, P., Srinivas Reddy, D., Sivasakthi, K., Bhatnagar-Mathur, P., Vadez, V., Sharma, K.K.: Evaluation of Sorghum [Sorghum bicolor (L.)] reference genes in various tissues and under abiotic stress conditions for quantitative real-time PCR data normalization. - Front. Plant Sci. 7: 529, 2016.
- Szakonyi, D., Byrne, M.E.: Ribosomal protein L27a is required for growth and patterning in *Arabidopsis thaliana*. - Plant J. 65: 269-281, 2011.
- Tanaka, Y., Sasaki, N., Ohmiya, A.: Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. - Plant J. 54: 733-749, 2008.
- Vandesompele, J., Preter, K.D., Pattyn, F., Poppe, B., Roy, N.V., Paepe, A.D., Speleman, F.: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. - Genome Biol. 3: research0034, 2002.
- Xiao, Z., Sun, X., Liu, X., Li, C., He, L., Chen, S., Su, J.: Selection of reliable reference genes for gene expression studies on *Rhododendron molle* G. Don. - Front. Plant Sci. 7: 1547, 2016.
- Zhang, C., Fu, J., Wang, Y., Bao, Z., Zhao, H.: Identification of suitable reference genes for gene expression normalization in the quantitative real-time PCR analysis of sweet osmanthus (Osmanthus fragrans Lour.). PloS ONE 10: e0136355, 2015.
- Zhang, C., Wang, Y., Fu, J., Bao, Z., Zhao, H.: Transcriptomic analysis and carotenogenic gene expression related to petal coloration in *Osmanthus fragrans* 'Yanhong Gui'. - Trees 30: 1207-1223, 2016.