Identification and validation of appropriate reference genes for gene expression analysis in *Schima superba*

Zhongyi Yang (1317250450@qq.com)
Chinese Academy of Forestry https://orcid.org/0000-0001-6516-918X
Rui Zhang
Research Institute of Subtropical Forestry Chinese Academy of Forestry
Zhichun Zhou
Research Institute of Subtropical Forestry Chinese Academy of Forestry

Research

Keywords: Reference gene, Real-time quantitative PCR, Schima superba, Tissues, Stability evaluation

Posted Date: April 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-361995/v1

Abstract

Background: Real-time quantitative PCR (RT-qPCR) is a reliable and high-throughput technique for gene expression studies, but its accuracy depends on the expression stability of reference genes. *Schima superba* is a strong resistance and fast-growing timber specie. However, so far, reliable reference gene identifications have not been reported in *S. superba*. In this study, we screened and verified the stably expressed reference genes in different tissues of *S. superba*.

Results: Nineteen candidate reference genes were selected and evaluated for their expression stability in different tissues. Three software programs (geNorm, NormFinder, and BestKeeper) were used to evaluate the reference gene transcript stabilities, and comprehensive stability ranking was generated by the geometric mean method. Our results identified that *SsuACT* was the most stable reference gene, *SsuACT* + *SsuRIB* was the best reference genes combination for different tissues. Finally, the stable and less stable reference genes were verified using the *SsuSND1* expression in different tissues.

Conclusions: This is the first report to verify the appropriate reference genes for normalizing gene expression in *S. superba*, which will facilitate future elucidation of gene regulations in this species, and useful references for relative species.

Background

At present, plant gene expression analysis methods include northern blot, in situ hybridization, RT-PCR, and Real-time quantitative PCR (RT-qPCR), and so on. RT-qPCR has been widely used in molecular biology research, and expression analysis is realized by real-time detection of fluorescence signal changes in the whole PCR reaction process due to its high sensitivity, accuracy, specificity, throughput capability, and cost-effectiveness [1–5]. However, the accuracy of relative quantification in RT-qPCR is always affected by many variable factors. For example, the RNA quality, integrity, reverse transcription efficiency, and amplification efficiency [2, 6]. To ensure accurate results and eliminate errors, it is necessary to use one or more stable reference genes to normalize the expression data of target genes [7].

Reference genes, also known as housekeeping genes, refer to a class of genes that can be stably expressed in different tissues and organs. In plant researches, the commonly used reference genes are mainly the genes that constituted the cytoskeleton or participated in the cells basic biochemical metabolic activities, including *actin* (*ACT*), *β-tubulin* (*TUB*), *ribosomal RNA* (*18S rRNA*, *26S rRNA*), *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*), *ubiquitin* (*UBQ*) and *elongation factors 1 alpha* (*EF-1α*) [8–10]. However, studies have shown that the expression levels of these gene is specific, and not stable under different species and treatments. The reference gene suitable for any condition does not exist [11–13]. Therefore, in the qualitative and quantitative research of genes, it is necessary to select the appropriate reference genes according to the specific experimental conditions [14].

Schima superba is an evergreen broad-leaved tree in Theaceae and it valued commercially for its timber and fire protecting [15–17]. Theaceae contains about 700 species, which possessed important economic values, such as the tea plant *Camellia sinensis* with their special drinking value, the traditional oil tree *C. oleifera* that produces high-quality edible seed oil and the ornamental plant *C. azalea* with their attractive flowers. The genus *Schima* has approximately 20 species and is mainly distributed in southern China and the adjacent parts of East Asia, with 13 species (six endemic) present in China [18]. Some reference genes of Theaceae had been reported. For example, *β-actin* could be used as a reference gene for tissues and *GAPDH* for mature leaves and callus in *C. sinensis* [19]. *TUA-3*, *ACT7α* and *CESA* were relatively stable in the different tissues of six oil-tea camellia species [20]. *TUA* and *GAPDH* were optimal reference genes in different organs and *TUB* and *UBQ* in petals in *C. azalea* [21]. However, there were few reports on gene expression analysis and reference genes expression stability in *S. superba*. Only Yang [22] used the *C. oleifera*'s *GAPDH* to verify the differential expression genes between self and cross pollination of *S. superba*, but the expression abundance and

stability have not been reported. With the completion of genome sequencing and construction of high-density genetic map [23], molecular design breeding and molecular assisted breeding of *S. superba* have been carried out gradually. So it is necessary to select appropriate reference genes for gene expression analysis.

In this study, 19 candidate reference genes, namely, *ColGAPDH*、*SsuACT*、*SsuGAPDH*、*SsuHis* (*histone*)、*SsuTUA1*(*alpha tubulin*)、*SsuTUA2*、*SsuUBC1*(*ubiquitin-conjugating enzyme*)、*SsuUBC2*、*SsuUBC17*、*SsuUBCJ2*、*SsuMDH*(*malate dehydrogenase*)、*SsuCal7*(*calmodulin-7*)、*SsuCas*(*caspase*)、*SsueIF5*(*eukaryotic translation initiation factor 5A4*)、*SsuMet2*(*metallothionin 2a*)、*SsuGTP*(*GTP-binding protein*)、*SsuRIB*(*60S ribosomal protein*)、*SsuTUB*(*tubulin beta-3*)、*SsuUDP*(*UDP-galactose transporte*), were assessed by RT-qPCR in *S. superba* different tissues (root, xylem, phloem, leaf, bud and fruit). To obtain the most suitable reference genes, three different statistical tools (geNorm, NormFinder and BestKeeper) were selected to evaluate the expression stability. In addition, *SECONDARY WALL-ASSOCIATED NAC DOMAIN 1* (*SND1*) gene, belonging to the NAC gene family, is involved in the initiation of secondary wall thickening in plant fibroblasts and is the primary switch in the transcriptional regulatory network of secondary wall thickening [24]. So to validate the selected best-ranked reference genes, the *SsuSND1* expression levels in different tissues were investigated using the most and less stable reference genes or the combination.

Results

Candidate reference genes and PCR amplification

Eighteen candidate reference genes were selected from the transcriptome of *S. superba* and *ColGAPDH* was cited from Yang [22]. The presence of a single PCR product expected size (Figure S1) and single peak in the melting curve (Figure S2) confirmed the specific amplification. The amplification efficiency (E) of all PCR reactions ranged from 93.47 % for *SsuTUA1* to 109.03 % for *SsuUBCJ2* (Table 1), suggested that these genes were suitable for further gene expression analysis. Meanwhile, the standard curves showed good linear relationships, with correlation coefficients (R^2) above 0.99 (Table 1).

Ct values of candidate reference genes

To assess the expression stability of 19 candidate reference genes in different tissues, the transcript abundances were presented as their Ct values. The Ct values varied in different tissues, which varied from 16.752 (*SsuMet2*) to 33.379 (*SsuCas*), while the mean Ct values varied from 18.032 (*SsuMet2*) to 25.556 (*SsuMDH*) (Table 2). The Ct range > 4 of *SsuCas, SsuUBC17, SsuUBC2* and *SsuUDP* in different tissues, suggested these genes varied greatly and were unstable in different tissues.

Analysis of reference gene stability using three bioinformatic programs

In order to reduce the analysis error, the candidate gene stability ranking in different tissues was determined separately using geNorm, NormFinder and BestKeeper. So as to screen out the reference genes suitable for the experimental treatment, and provide beneficial reference for subsequent research.

The geNorm program is used to rank the genes expression stability by calculating the average expression stability values (M) based on the $2^{-\Delta Ct}$ value [25]. The smaller M value of reference gene, the more stably expressed. Meanwhile, if M value > 1.5, it is not suitable as a reference gene [25]. The M values of tested genes evaluated by geNorm were shown in Fig.1. *SsuTUA1* and *SsuRIB* were ranked as the two most stable genes in different tissues, while *SsuCas* and *SsuUDP* were the two least stable genes.

The pairwise variation value (V_n/V_{n+1}) calculated by the geNorm determines the optimal number of reference genes. When V_n/V_{n+1}< 0.15, the optimal number of reference genes is n, otherwise, the number is n+1 [25]. In this study, except for V_{18}/V_{19}, the other value of V_n/V_{n+1} < 0.15 (Fig.2), indicating that two reference genes would be sufficient for gene normalization and an increase do not improve sensitivity.

NormFinder ranks the expression stability of reference genes by calculating the average pairwise variation in one relative to other candidate genes and the smaller stability value, the more suitable as reference gene [26]. For different tissues, the most stable gene was *SsuACT*, followed by *SsuRIB*, while the least stable gene was *SsuCas*, which were not altogether with the genes selected by the geNorm (Fig.3).

The expression stability is represent by the standard deviation (SD), coefficient of variance (CV) and correlation coefficient (r) of Ct values in the BestKeeper program, and the most stable reference genes are identified as those with the lowest SD and CV and the most r [27]. In this study, *SsuACT* and *SsuUBCJ2* were identified as the most stable genes for different tissues, while *SsuUBC17*, *SsuTUB*, *SsuUBC2*, *SsuUDP*, *SsuCas* are unstable because of the SD > 1(Table 3).

The rankings of the nineteen tested genes were not perfectly consistent by geNorm, NormFinder and BestKeeper (Table 4). To provide a comprehensive evaluation of candidate reference genes, further analysis was thus carried out using geometric mean, which integrates geNorm, Normfinder and BestKeeper. The comprehensive ranking recommended by the geometric mean method was shown in Table 4, and *SsuACT* was the most stable genes for different tissues.

The best combination of reference genes was determined based on the optimal number calculated by geNorm and the ranking list obtained using geometric mean method. So *SsuACT* + *SsuRIB* was found to be the best combination of reference genes for different tissues.

Validation of the identified reference genes

In order to examine the reliability of the candidate reference genes for normalization, *SsuSND1* expression profiles in different tissues were normalized using the two most stable candidate reference genes (*SsuACT* and *SsuRIB)*, combination of stable genes (*SsuACT* + *SsuRIB*), the least stable reference gene (*SsuCas*) well as *ColGAPDH* (Fig.4). When *SsuACT*, *SsuRIB*, *SsuACT* + *SsuRIB* and *ColGAPDH* were used for normalization, the expression patterns of *SsuSND1* were similar, and the relative expression of xylem, leaf and fruit were higher than others, and *SsuSND1* was hardly expressed in bud and root, but the expression was most appropriate by *SsuACT* and *SsuACT* + *SsuRIB*. However it was analyzed by *SsuCas*, the expression pattern was not compatible and the expression level were too high in fruit and too low in bud. It was suggested that the selected reference genes were reliable.

Discussion

RT-qPCR has become a common technique for molecular biology research [28]. In the analysis process, reference genes are often used to reduce or correct the errors in the quantitative process of target genes. Therefore, the selection of appropriate reference gene is the key to realize the research of target gene expression under different experimental conditions or tissues [29]. *S. superba* has economic value for its timber, and the wood is used for furniture and construction. According to the phylogenetic analysis of Theaceae, Theeae and Gordonieae are closer related to each other, and the genera *Schima* belongs to the Gordonieae [30]. Moreover we have constructed a high-density genetic map and obtained 168 QTLs for 14 phenotypes [23], but still not focused on its molecular function and gene expression. Therefore, to carry out the follow-up experiment smoothly, a stable and suitable reference gene would be selected and evaluated for the normalization of gene expression analysis by RT-qPCR in our research.

To avoid the limitations of using only single software analysis, three bioinformatic programs (geNorm, NormFinder and BestKeeper) were used to evaluate the expression stability of candidate reference genes in our analysis. The basis for evaluating gene stability in GeNorm is to use the $2^{-\Delta Ct}$ value of each gene to calculate the M value [25]. Meanwhile, GeNorm can determine the optimal number of reference genes required for quantitative analyses. In this study, the gene expression analysis needs two reference genes to achieve the best performance. The algorithm of NormFinder is similar to GeNorm using the $2^{-\Delta Ct}$ value as the relative expression to calculate the stability of gene expression [26]. BestKeeper focuses on the standard coefficient variation (SD) and variation correlation coefficient (CV) to screen the stability of internal reference genes [27]. The ranking from different programs showed some substantial discrepancies (Table 4). For instance, *SsuTUA1* and *SsuRIB* were the best reference genes identified by geNorm (Fig. 1), while it turned out that *SsuACT* was evaluated as the best by NormFinder (Fig. 3) and BestKeeper (Table 3). Differences in rankings among these programs have also been reported in other studies [12, 31, 32], which is likely the result of the different algorithms that they employ [33]. Therefore, to provide a comprehensive evaluation of candidate reference genes, the geometric mean was used to generate comprehensive stability ranking, and then the best combinations were determined based on the optimal number of reference genes calculated by geNorm. So *SsuACT*, *SsuACT + SsuRIB* were the most stable reference gene and combination for different tissues (xylem, phloem, leaf, bud, fruit and root). But in study processes, we found that Ct values of 19 candidate reference genes in seeds were significantly greater than other tissues, so the transcription levels fluctuated greatly in the seven tissues (root, phloem, xylem, leaf, bud, fruit and seed). And the rankings of three programs were contradictory, so we excluded seeds and analyzed only the remaining six tissue. It is worth mentioning that, *SsuRIB* was the best reference gene predicted among the seven tissues by NormFinder, which indicated that *SsuRIB* could also be used to normalize the gene expression of seeds, because NormFinder is more suitable for the situation when the genes transcription level fluctuates greatly [26, 27].

Actin is a widely used as reference gene in plants. In this study, *SsuACT* is also the most stable gene. *SsuRIB* and *SsuCal7* are the novel genes screened from genome and transcriptome of *S. superba*, and they have not been reported as reference genes in other species. And *ColGAPDH* is not suitable as a reference gene of *S. superba* because of low expression and poor stability. We compared the reference genes in *S. superba, C. sinensis*, *C. oleifera* and *C. azalea* in Theaceae. The optimal reference genes in different tissues are *SsuACT*, *SsuRIB* and *SsuTUA1* of *S. superba*, *β-actin* of *C. sinensis* [19], *TUA-3*, *ACT7α* and *CESA* of *C. oleifera* [20] and *TUA* and *GAPDH* of *C. azalea* [21]. So *ACT* could be used as the reference gene of *S. superba*, *C. sinensis and C. oleifera*, and *TUA* could be used of *S. superba*, *C. oleifera and C. azalea.* Then we think *ACT* and *TUA* have wide applicability as the reference genes in Theaceae.

To validate the identified reference genes suitability, *SsuSND1* expression patterns were investigated in different tissues using different reference genes. The expression pattern by normalized of *SsuCas* was not compatible with *SsuACT + SsuRIB*. The data once again demonstrated that reference genes play a key role in normalizing the data from RT-qPCR, and the use of inappropriate reference genes may lead to inaccurate results. Moreover, NAC plays a key role in the formation and development of apical meristem [34], lateral root [35] and secondary wall [36]. SND1 plays a similar role in *S. superba* and the relative expression of secondary xylem was higher than others.

Conclusions

This study is the first systematic report about selection and verification of reliable stable reference genes for different tissues in *S. superba*, showing that *SsuACT* was the most stable reference gene, *SsuACT + SsuRIB* was the best combination for different tissues of *S. superba*. This study could provide a basis at the gene expression level of *S. superba* and Theaceae.

Materials And Methods

Plant materials

The plant materials were collected from the germplasm bank of *S. superba* clones (119°06'E, 28°03'N) in Zhejiang Longquan Academy of Forestry Sciences. The clones are the 25-year-old mother trees grafted with scions from Jianou, Fujian (118°31'E, 27°8'N) in 2008. The gene bank covers an area of 6.7 hm^2 and is located at the altitude of 200–300 m. And the relative humidity of the area is 79%, and the average annual rainfall is 1 664.8–1706.2 mm.

In August 2020, different tissues were collected, including secondary xylem, secondary phloem, mature leaf, bud, annual fruit and root of tissue culture seedlings subcultured for 60 days. Each sample was set up in biological triplicate. The samples were frozen with liquid nitrogen and stored in -80℃.

Selection of candidate reference genes

Eighteen candidate reference genes（GeneBank accessions: MW770873~770890）which had a stable expression in the different tissues and age of *S. superba* were selected according to transcriptome data from our laboratory (Unpublished, Novogene, Beijing, China) (Table S1). Eleven of them are often used as housekeeping genes in model plant species. And *ColGAPDH* (*C. oleifera*)（GeneBank accession: KC337052）gene was used as control [22], and the sequence consistency is only 38.68% with *SsuGAPDH*. According to their CDS sequences blasted from Genomic data of *S. superba* (Unpublished, Novogene, Beijing, China) (Table S2), the primers were designed on the Web using the Primer 3.0 (http://www.primer3plus.com/primer3web/primer3web_input.htm) and synthesized by Sangon Biotech Co., Ltd. (Shanghai, China), as shown in Table S3.

RT-qPCR analysis

The total RNA was extracted using the RNAprep Pure Plant Plus Kit (Polysaccharides & Polyphenolics-rich) (Code No. DP441, TIANGEN, China) and were stored at -80℃. RNA was assessed using 1% agarose gel electrophoresis and was quantified with a Nanodrop ND-2000 ultra micro nucleic acid protein analyzer (Thermo, USA). The RNA samples with A260/A280 ratios between 1.8 and 2.1 (Table S4) were used to synthesize the first strand of cDNA with the PrimeScript™ RT master mix (Perfect Real Time) (Code No. RR036A, Takara, Japan) in a 20 μL reaction mixture, once they were adjusted to the same concentration of RNA to 1 μg. The cDNA was diluted 1:9 with nuclease-free water prior to the RT-qPCR analysis.

RT-qPCR was restricted to the guidelines as follows in system (Applied Biosystems Q7, USA): the reaction mixture (20 μl) contained 10 μl 2X TB GreenPremix Ex TaqII (Tli RNaseH Plus) (Code No. RR820A, TaKaRa, Japan), 2 μl diluted cDNA, 0.8 μl each primer (10 mM), 0.4 μl ROX Reference Dye (50X), and 6 μl water. The reactions were incubated under the following cycling conditions: 30 s at 95℃, 40 cycles of 95℃ for 5 s, and Tm 60℃ for 30 s, with a single melt cycle from 65 to 94℃ in 5 s intervals.

The primer specificity was verified by the presence of a single peak in the melt curve analysis during the RT-qPCR process. Three independent biological replicates and three technical repetitions were performed for each of the quantitative PCR experiments. The threshold cycle (Ct) was measured automatically and correlation coefficients (R^2) together with slope were calculated from the standard curve based on a tenfold series dilution of the cDNA templates. The corresponding RT-qPCR efficiencies (E) for each gene were determined from the given slope.

Validation of identified reference genes

The *SND1* （GeneBank accession: MW796194） was selected as target gene to validate reliability of identified reference genes from transcriptome data (Unpublished, Novogene, Beijing, China) (Table S1). The gene expression profiles at

different tissues were normalized using the most and least stable reference gene and *ColGAPDH*. Sample collections and experiments were performed as described above.

Statistical data analysis

The average Ct value was calculated from three biological replicates and three technical replicates, relative gene expression levels were calculated using the $2^{-\triangle\triangle Ct}$ method [37].

GeNorm, NormFinder and BestKeeper algorithms were used to evaluate the stability of 19 candidate reference genes according to the Ct values. And using geometric mean to provide a comprehensive stability evaluation of candidate reference genes.

Declarations

Authors' contributions

ZY, RZ and ZZ designed the research; ZY conducted the experiment, analyzed the data and wrote the manuscript. All authors read and approved the manuscript.

Author details

[1]Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, China. [2]Nanjing Forestry University, Nanjing 210037, China. [3]Zhejiang Provincial Key Laboratory of Tree Breeding, Hangzhou 311400, China

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its Additional files].

Consent for publication

All authors have consented for publication.

Ethics approval and consent to participate

Not applicable.

Funding

This study was supported by the Special Fund for Breeding New Varieties of Forest Species in Zhejiang Province (grant number 2016C020563), the Science and Technology Innovation Project in Jiangxi Province (grant number [2019]19), and the Sixth Stages of Planting and Seedling Science and Technology of Fujian Province (grant number 201906).

Abbreviations

RT-qPCR: Real time quantitative PCR; *S. superba*: *Schima superba*; Ct: Cycle threshold; *SsuSND1*: Transcription factor SECONDARY WALL-ASSOCIATED NAC DOMAIN 1; CV: Coefficient of variation; SD: Standard deviation; CDS: Coding sequence; cDNA: Complementary DNA

References

1. Wong ML, Medrano JF. Real-time PCR for mRNA quantitation. Biotechniques. 2005;39(1):75–85.
2. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. Nat Protoc. 2006;1(3):1559–82.
3. Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa M, Alves-Ferreira M. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. BMC Plant Biol. 2010;10(1):49.
4. Wen L, Tan B, Guo WW. Estimating transgene copy number in precocious trifoliate orange by TaqMan real-time PCR. Plant Cell Tiss Org. 2012;109(2):363–71.
5. Su XY, Shi YB, Yang XM, Wang GB, Cao FL. Selection and validation of reference genes for quantitative real-time PCR analysis in *Ginkgo biloba*. Plant Physiology Journal. 2019;55(6):875–82.
6. Yoo WG, Kim TI, Li SY, Kwon OS, Cho PY, Kim T-S, Kim KJ, Hong S-J. Reference genes for quantitative analysis on *Clonorchis sinensis* gene expression by real-time PCR. Parasitol Res. 2009;104(2):321–8.
7. Huang T, Long JM, Liu SW, Yang ZW, Zhu QJ, Zhao XL, Peng CC. Selection and validation of reference genes for mRNA expression by quantitative real-time PCR analysis in *Neolamarckia cadamba*. Sci Rep. 2018;8(1):9311.
8. Nicot N, Hausman JF, Hoffmann L, Evers D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot. 2005;56(421):2907–14.
9. V Patankar H, Assaha M, Al-Yahyai DV, Sunkar R, Yaish R. MW. Identification of reference genes for quantitative real-time PCR in Date Palm (*Phoenix dactylifera* L.) subjected to drought and salinity. PLoS ONE. 2006;11(11):e0166216.
10. Qu RJ, Miao YJ, Cui YJ, Cao Y, Zhou YW, Tang XQ, Yang J, Wang FY. Selection of reference genes for the quantitative real-time PCR normalization of gene expression in *Isatis indigotica* fortune. BMC Mol Biol. 2019;20(1):9.
11. Li J, Han J, Hu Y, Yang J. Selection of reference genes for quantitative real-time PCR during flower development in Tree Peony (*Paeonia suffruticosa* Andr.). Front Plant Sci. 2016;7:516.
12. Sudhakar Reddy P, Srinivas Reddy D, Sivasakthi K, Bhatnagar-Mathur P, Vadez V, Sharma KK. 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. 2016;7:529.
13. Yu YT, Zhang G, Chen YK, Bai QQ, Gao CS, Zeng LB, Li ZM, Cheng Y, Chen J, Sun XP, Guo LT, Xu JP, Yan Z. Selection of reference genes for qPCR analyses of gene expression in ramie leaves and roots across eleven abiotic/biotic treatments. Sci Rep. 2019;9(1):20004.
14. Xiao Z, Sun XB, Liu XQ, Li C, He LS, Chen SP, Su JL. Selection of reliable reference genes for gene expression studies on *Rhododendron molle* G. Don. Front Plant Sci. 2016;7:1547.
15. Yang HB, Zhang R, Zhou ZC. Pollen dispersal, mating patterns and pollen contamination in an insect-pollinated seed orchard of *Schima superba* Gardn. Et champ. New Forest. 2017;48:431–44.
16. Yang HB, Zhang R, Song P, Zhou ZC. The floral biology, breeding system and pollination efficiency of *Schima superba* Gardn. Et Champ Forests. 2017;8:404.
17. Zhou ZC, Zhang R, Fan HH, Chu XL. Schima superba. Beijing: Chinese Science Press; 2020. pp. 1–9.
18. Stevens PF, Dressler S, Weitzman AL. Theaceae. In: Flowering plants Dicotyledons. Berlin: Springer; 2004. pp. 463–71.
19. Sun ML, Wang YS, Yang DQ, Wei CL, Gao LP, Xia T, Shan Y, Luo Y. Reference genes for real-time fluorescence quantitative PCR in *Camellia sinensis*. Chin Bull Bot. 2010;45:579–87.

20. Zhou CF, Lin P, Yao XH, Wang KL, Chang J, Han XJ. Selection of reference genes for quantitative real-time PCR in six oil-tea camellia based on RNA-sEq. Mol Biol. 2013;47(6):836–51.
21. Liu XF, Yu B, Huang LL, Sun YB. Screening and Validation of Reference Genes of *Camellia azalea* by Quantitative Real-time PCR. Guangdong Agricultural Sciences. 2020;47(12):203–11.
22. Yang HB. Characteristics of the reproductive biology and the mating system in seed orchard of *Schima superba*. Chinese Academy of Forestry. 2017.
23. Zhang R, Yang HB, Zhou ZC, Shen B, Xiao JJ, Wang BS. A high-density genetic map of *Schima superba* based on its chromosomal characteristics. BMC Plant Biol. 2019;19(1):41.
24. Takata N, Awano T, Nakata MT, Sano Y, Sakamoto S, Mitsuda N, Taniguchi T. Populus NST/SND orthologs are key regulators of secondary cell wall formation in wood fibers, phloem fibers and xylem ray parenchyma cells. Tree Physiol. 2019;39(4):514–25.
25. 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(7):research0034.
26. Andersen CL, Jensen JL, Orntoft 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;64:5245–50.
27. 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(6):509–15.
28. Xu LF, Xu H, Cao YW, Yang PP, Feng YY, Tang YC, Yuan SX, Ming J. Validation of reference genes for quantitative real-time PCR during bicolor tepal development in Asiatic Hybrid Lilies (*Lilium* spp.). Front Plant Sci. 2017;8:669.
29. Han XJ, Lu MZ, Chen YC, Zhan ZY, Cui QQ, Wang YD. Selection of reliable reference genes for gene expression studies using real-Time PCR in Tung Tree during seed development. PLoS ONE. 2012;7(8):e43084.
30. Li MM. Phylogenetic and biogeographic studies of Theaceae with a special reference to *Stewartia*. Zhejiang University, 2011.
31. Delporte M, Legrand G, Hilbert J-L, Gagneul D. Selection and validation of reference genes for quantitative real-time PCR analysis of gene expression in *Cichorium intybus*. Front Plant Sci. 2015;6:651.
32. Mangeot-Peter L, Legay S, Hausman JF, Esposito S, Guerriero G. Identification of reference genes for RT-qPCR data normalization in *Cannabis sativa* stem tissues. Int J Mol Sci. 2016;17(9):1556.
33. Zhao JM, Zhou M, Meng Y. Identification and validation of reference genes for RT-qPCR analysis in Switchgrass under heavy metal stresses. Genes. 2020;11(5):502.
34. Souer E, Van Houwelingen A, Kloos D, Mol J, Koes R. The no apical meristem gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell. 1996;85(2):159–70.
35. Mao CJ, He JM, Liu LN, Deng QM, Yao XY, Liu CM, Qiao YL, Li P, Ming F. OsNAC2 integrates auxin and cytokinin pathways to modulate rice root development. Plant Biotechnol J. 2020;18(2):429–42.
36. Zhong RQ, Demura TK, Ye ZH. SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. Plant Cell. 2006;18(11):3158–70.
37. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods. 2001;25(4):402–8.

Tables

Table 1 Amplification efficiency and correlation coefficient of candidate reference genes			
Gene	Amplification efficiency (%)	Correlation coefficient (R^2)	Ct value
ColGAPDH[21]	106.14	0.9991	21.562±0.865
SsuACT	99.17	0.9981	23.877±1.059
SsuCal7	106.74	0.9998	21.984±0.619
SsuCas	104.35	0.9977	23.388±1.467
SsueIF5	103.76	0.9996	25.185±0.807
SsuGAPDH	97.39	0.9996	20.608±0.903
SsuGTP	100.02	0.9929	23.709±4.902
SsuHis	109.7	0.9991	21.455±1.150
SsuMDH	102.6	0.9994	20.113±0.546
SsuMet2	97.27	0.9995	22.875±0.808
SsuRlB	94.92	0.9931	22.073±1.003
SsuTUA1	93.47	0.9991	25.556±1.004
SsuTUA2	104.27	0.9986	18.032±0.984
SsuTUB	103.17	0.9984	22.314±0.848
SsuUBC1	105.46	0.9907	25.320±0.817
SsuUBC17	107.89	0.9993	21.644±1.555
SsuUBC2	104.03	0.9985	21.556±1.681
SsuUBCJ2	109.03	0.9975	24.512±1.180
SsuUDP	108.71	0.9989	24.877±1.782
Ct (Cycle threshold) means the number of cycles experienced when the fluorescent signal in each reaction tube reaches the set domain value. Ct values are mean ± SD (n = 3).			

Table 2 Average Ct values of candidate reference genes in different tissues of *S. superba*										
Gene	Leaf	Bud	Fruit	Phloem	Root	Xylem	Average	Min	Max	Range
SsuACT	21.125	20.818	22.681	21.606	22.504	20.638	21.562	20.638	22.681	2.043
SsuTUA1	23.116	22.855	24.844	23.801	25.474	23.173	23.877	22.855	25.474	2.62
SsuTUA2	21.717	21.637	22.489	21.927	22.918	21.214	21.984	21.214	22.918	1.704
SsubTUB	23.802	22.222	24.35	22.667	25.599	21.687	23.388	21.687	25.599	3.911
ColGAPDH	24.569	24.188	26.065	26.214	25.079	24.997	25.185	24.188	26.214	2.026
SsuCal7	20.469	19.505	21.703	20.858	21.444	19.672	20.608	19.505	21.703	2.198
SsuCas	22.702	20.807	33.379	21.377	23.713	20.28	23.709	20.28	33.379	13.099
SsueIF5	21.82	19.81	21.7	21.423	23.243	20.736	21.455	19.81	23.243	3.433
SsuGAPDH	19.325	19.818	20.596	20.723	20.409	19.81	20.113	19.325	20.723	1.398
SsuGTP	22.843	21.975	22.92	22.966	24.304	22.239	22.875	21.975	24.304	2.328
SsuHis	21.538	20.623	22.414	22.775	23.434	21.654	22.073	20.623	23.434	2.811
SsuMDH	24.656	24.456	26.864	25.655	26.601	25.104	25.556	24.456	26.864	2.408
SsuMet2	17.014	16.752	17.975	19.275	18.652	18.522	18.032	16.752	19.275	2.523
SsuRIB	21.848	21.358	23.163	22.122	23.539	21.856	22.314	21.358	23.539	2.181
SsuUBC1	24.886	24.18	25.768	26.41	25.82	24.857	25.32	24.18	26.41	2.229
SsuUBC17	22.491	20.03	21.798	21.316	24.137	20.09	21.644	20.03	24.137	4.107
SsuUBC2	22.343	20.277	21.973	20.856	24.277	19.607	21.556	19.607	24.277	4.67
SsuUBCJ2	24.718	23.014	24.883	24.601	26.375	23.482	24.512	23.014	26.375	3.361
SsuUDP	25.303	23.309	25.399	24.038	27.973	23.242	24.877	23.242	27.973	4.73

Table 3 Expression stability of candidate reference genes analyzed by BestKeeper								
Gene	geometric Mean	average Mean	minimum	maximum	SD	CV	r	p-value
SsuACT	21.55	21.56	20.64	22.68	0.7	3.25	0.987	0.001
SsuUBCJ2	24.49	24.51	23.01	26.38	0.84	3.44	0.987	0.001
SsuCal7	20.59	20.61	19.51	21.7	0.73	3.53	0.977	0.001
SsuHis	22.05	22.07	20.62	23.43	0.8	3.63	0.93	0.007
SsuRIB	22.3	22.31	21.36	23.54	0.69	3.1	0.928	0.008
SsuUDP	24.83	24.88	23.24	27.97	1.35	5.42	0.926	0.008
SsuTUA1	23.86	23.88	22.85	25.47	0.85	3.58	0.92	0.009
SsuUBC17	21.6	21.64	20.03	24.14	1.17	5.38	0.919	0.01
SsuTUA2	21.98	21.98	21.21	22.92	0.48	2.18	0.905	0.013
SsuUBC2	21.5	21.56	19.61	24.28	1.31	6.07	0.887	0.018
SsuTUB	23.35	23.39	21.69	25.6	1.2	5.11	0.858	0.029
SsuCas	23.35	23.71	20.28	33.38	3.22	13.6	0.83	0.041
SsuMDH	25.54	25.56	24.46	26.86	0.82	3.2	0.824	0.044
SsueIF5	21.43	21.46	19.81	23.24	0.8	3.72	0.761	0.079
SsuGTP	22.86	22.87	21.98	24.3	0.52	2.28	0.742	0.092
SsuUBC1	25.31	25.32	24.18	26.41	0.68	2.68	0.741	0.092
SsuGAPDH	20.11	20.11	19.32	20.72	0.46	2.3	0.634	0.176
ColGAPDH	25.17	25.19	24.19	26.21	0.64	2.53	0.602	0.206
SsuMet2	18.01	18.03	16.75	19.27	0.78	4.35	0.513	0.296
SD: standard deviation; CV: coefficient of variance; r: correlation coefficient.								

Table 4 Comprehensive evaluation on stability of candidate reference genes					
Gene	geNorm	NormFinder	BestKeeper	geometric Mean	Combined ranking
SsuACT	3	1	1	1.44	1
SsuRIB	1	2	5	2.15	2
SsuTUA1	1	4	6	2.88	3
SsuCal7	4	3	3	3.3	4
SsuMDH	2	5	8	4.31	5
SsuUBCJ2	8	6	2	4.58	6
SsuTUA2	5	7	7	6.26	7
SsuHis	7	9	4	6.32	8
SsuGTP	6	8	10	7.83	9
SsueIF5	9	10	9	9.32	10
SsuUBC1	10	12	11	10.97	11
SsuGAPDH	11	13	12	11.97	12
ColGAPDH	12	14	13	12.97	13
SsuTUB	14	11	16	13.51	14
SsuMet2	13	18	14	14.85	15
SsuUBC17	15	15	15	15	16
SsuUBC2	16	16	17	16.33	17
SsuUDP	17	17	18	17.33	18
SsuCas	18	19	19	18.66	19

Figures

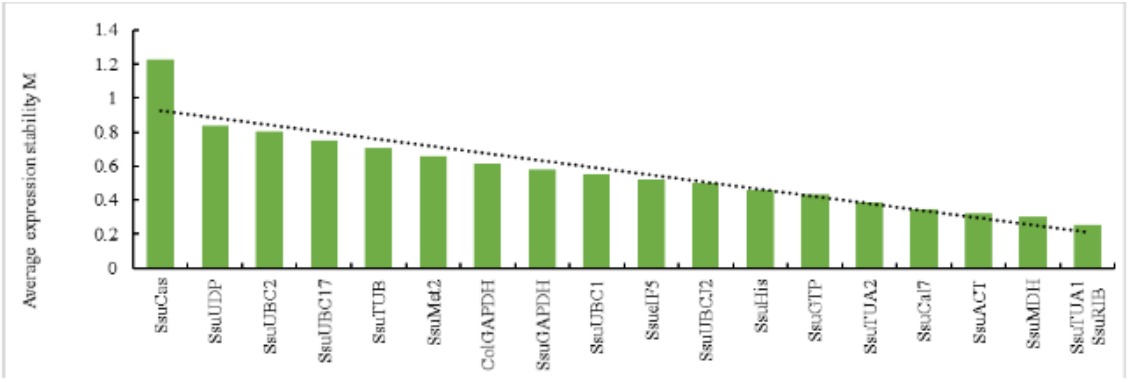

Fig.1 Expression stability values (M) of candidate reference genes calculated

Figure 1

Expression stability values (M) of candidate reference genes calculated by geNorm

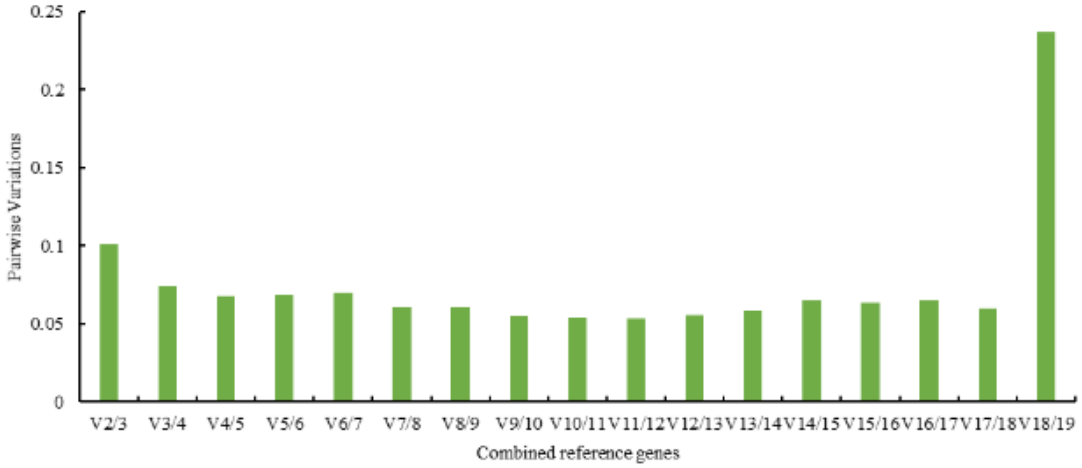

Fig. 2 Pairwise Variation (V) of candidate reference genes calculated

Figure 2

Pairwise Variation (V) of candidate reference genes calculated by geNorm

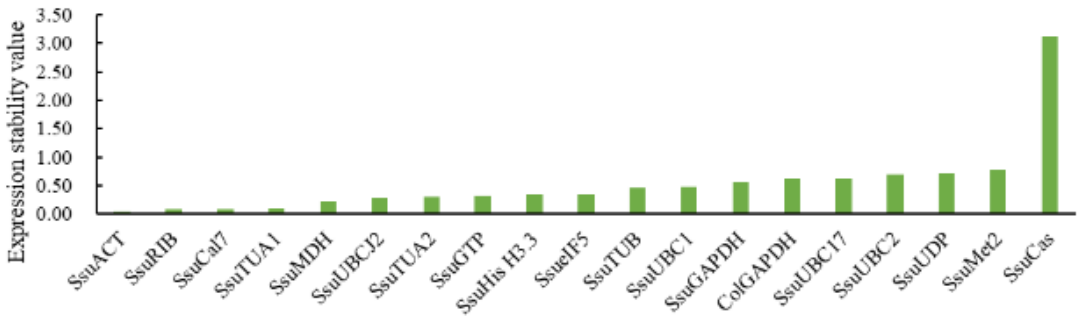

Fig. 3 Expression stability of candidate reference genes analyzed by NormFinder

Figure 3

Expression stability of candidate reference genes analyzed by NormFinder

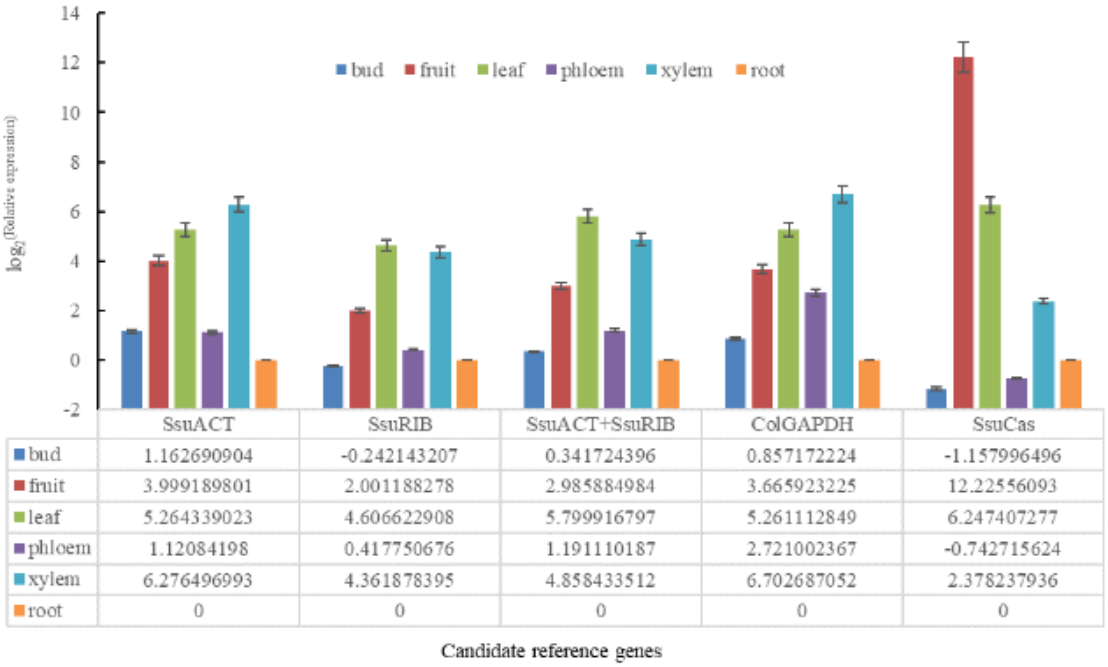

Fig. 4 Relative expression of *SsuSND1* gene using various different reference genes for normalization in different tissues of *S. superba*

Figure 4

Relative expression of SsuSND1 gene using different reference genes for normalization in different tissues of S. superba

Supplementary Files

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

- Additionalfile1FigureS1.docx
- Additionalfile2FigureS2.docx
- Additionalfile3TableS1.xlsx
- Additionalfile4TableS2.docx
- Additionalfile5TableS3.xlsx
- Additionalfile6TableS4.xlsx