Selection and validation of reference genes for normalization of qRT-PCR data to study secondary metabolite related genes in industrial hemp

Michihito Deguchi
Penn State Harrisburg
Shobha Potlakayala
Penn State Harrisburg
Zachary Spuhler
Penn State Harrisburg
Hannah George
Penn State Harrisburg
Vijay Sheri
Penn State Harrisburg
Ruba Agili
Penn State Harrisburg
Aayushi Patel
Penn State Harrisburg
Sairam Rudrabhatla (svr11@psu.edu)
Penn State Harrisburg

Research Article

Keywords: hemp, Cannabis sativa L, gene expression

Posted Date: April 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-394417/v1

Abstract

Industrial hemp (*Cannabis sativa* L.) is a dioecious crop widely known for its production of phytocannabinoids, flavonoids, and terpenes. In the past two years since its legalization, there has been significant interest in researching this important crop for pharmaceutical applications. Although many scientific reports have demonstrated gene expression analysis of hemp through OMICs approaches, accurate validation of omics data cannot be performed because of lack of reliable reference genes for normalization of qRT-PCR data. The differential gene expression patterns of 13 candidate reference genes under osmotic, heavy metal, hormonal, and UV stress were evaluated through four software packages: geNorm, NormFinder, BestKeeper, and RefFinder. The *EF-1a* ranked as the most stable reference gene across all stresses, *TUB* was the most stable under osmotic stress, and *TATA* was the most stable under both heavy metal and hormonal stress. The expression profiles of two cannabinoid pathway genes, *AAE1* and *THCAS*, using the two most stable and single least stable reference genes confirmed that two most stables genes were apt for normalization of gene expression data. This work will contribute to the future studies on the expression analysis of hemp genes regulating the synthesis, transport and accumulation of secondary metabolites.

Introduction

Industrial hemp (*Cannabis sativa* L.) has a rich history with human civilization because it can provide both phytochemicals and lignocellulosic biomass. This crop is believed to have originated in Eurasia and has been distributed all over the world, largely for its use as a fiber crop[1]. After the emergence of other fiber crops, the demand for hemp as a source of fiber has reduced, but in turn, the value has increased as a human food additive. Hemp seed contains essential fatty acids and proteins as well as gamma-linolenic acid which has numerous health benefits[2]. Hemp seeds and oils are also widely used in nutritional supplements and cosmetics. Recently, attention has been focused on its rich repertoire of pharmaceutical compounds[3].

Hemp produces a diverse array of phytocannabinoids, terpenes, and phenolic compounds with prominent nutraceutical potential[4]. Among them, the phytocannabinoids are the most well-known phytochemicals in hemp. The predominant compound, cannabidiol (CBD) and tetrahydrocannabinol (THC) followed by cannabigerol (CBG) and cannabichromene (CBC) are highly promising compounds to improve human health. They act as therapeutic agents for central nervous system diseases such as epilepsy, inflammation, anxiety, and neurodegenerative disorders such as Parkinson's disease, Huntington's disease, Tourette's syndrome, and Alzheimer's disease[3]. Terpenes also present a wide array of pharmacological properties, including anxiolytic, antibacterial, anti-inflammatory, and sedative effects on human diseases[5, 6,7].

The increasing popularity of hemp-based phytochemicals, has spurred the comprehensive analysis of gene expression and protein synthesis utilizing omics approaches[8, 9, 10, 11]. These analysis are necessary

to identify genes involved in the secondary metabolite pathways[12, 13] and have been instrumental in discovering transcription factors, the key proteins that positively or negatively control the synthesis of secondary metabolites[14, 15]. Accurate gene expression studies are essential to confirm the transcriptomic and proteomic analysis data. To date, several different techniques for gene expression analysis such as Northern blotting[16], ribonuclease protection assay (RPA)[17], serial analysis of gene expression (SAGE)[18], and quantitative real-time PCR (qRT-PCR)[19] have been used. Among these methods, qRT-PCR has been used the most frequently for gene expression analysis due to its high sensitivity, specificity, accuracy, reproductivity, and relatively low cost[20]. qRT-PCR also requires a minimal quantity of RNA compared to hybridization-based, RNA quantification methods such as northern blot.

In qRT-PCR, relative gene expression levels can be assessed among different samples in the same parameters. The qRT-PCR analysis is used to detect changes in the expression of genes of interest relative to a reference gene. However, due to the variances in RNA extraction, DNase treatment, cDNA synthesis, and amplification of target genes, the reliability of gene expression results can be affected by sample size, RNA degradation, reverse transcription efficiency, and cDNA quality. To minimize the effect of these factors, reference genes are used as internal controls to provide accurate and reproducible results of gene expression profiles. Recently expression levels have been shown to vary depending on genotypes, tissues, and environmental conditions[21, 22, 23]. Therefore, the use of unstable reference genes will lead to significant biased and misinterpreted gene expression data. Thus, it is critical to identify appropriate and reliable reference genes for each experimental set-up in the respective plant tissue and genotype.

Mangeot-Peter et al (2014) [26] identified suitable reference genes in hemp stem tissue for accurate expression profiling of cell wall synthesizing genes. Subsequently, Guo et al., (2018)[25] studied seven reference genes in various tissues such as root, stem, leaf, and flower. However, to the best of our knowledge, the stability of hemp reference genes was not investigated under different environmental conditions. In this study, we analyzed the stability of 13 reference genes under 11 abiotic stresses/hormone stimuli in hemp leaves to select the best reference genes for qRT-PCR studies. We selected the elite hemp cultivar “Thunderbird” which accumulates CBDA at high concentrations. The following 13 reference genes were used in the current study: *18S ribosomal RNA* (*18S*), *40S ribosomal protein* (*40S*), *E3 ubiquitin-protein ligase* (*UBE2*), *Elongation factor 1-α* (*EF-1α*), *Glyceraldehyde-3-phosphate dehydrogenase* (*GAD*), *TATA-box-binding protein* (*TATA*), *Tubulin α-1* (*TUB*), *chalcone synthase* (*CHAL*), *F-box family* (*F-box*), *Phytochelatin synthase* (*PCS1*), *protein phosphatase 2A subunit* (*PP2A*), sand family (*SAND*), and *TIP41-like family protein* (*TIP41*). The reliability of these reference genes was validated using qRT-PCR of two cannabinoid pathway genes *acyl-activating enzyme 1* (*AAE1*) and *CBDA synthase* (*CBDAS*) normalized to the most stable and least stable genes.

Materials And Methods

Reference to *Cannabis sativa*:

Industrial hemp and medical marijuana plants share *Cannabis sativa* as their common scientific name. Therefore, in this paper, the authors chose to refer to industrial hemp as “hemp”, to distinguish it from medical marijuana.

Plant material, greenhouse conditions, generation of clones, growth, and care

The hemp strain, Thunderbird, was grown following the approved guidelines for industrial hemp provided by the Pennsylvania Department of Agriculture - Bureau of Plant Industry under the regulated permits IH-16-P-2017 and IH-17-P-2017.

During this experiment, the greenhouse was maintained at 25°C with a 14-hour light photoperiod at 25-40 $\mu Em^{-2}s^{-1}$. Hemp clones were obtained by collecting a 3-inch segment containing two axillary buds and coating the 45-degree cut with Clonex Rooting gel (Hydrodynamics International, Inc. Lansing, MI). The explant was then placed in Root Riot plugs (Hydrodynamics International, Inc. Lansing, MI) and maintained under propagation domes for two weeks at which point they were transferred to four-inch pots containing high porosity soil, HP Mycorrhizae from Pro-Mix (Rivière-du-Loop, Québec, Canada). The clones grew to 15cm by week four and flowered by week nine. All cloned plants used in this study were obtained from the same female mother plant by vegetative cutting to prepare genetically identical seedlings of similar size.

During the rooting period, the clones were kept under 24-hour light under propagation domes and 12-hour light during the pre-flowering and flowering periods. Humidity for rooting clones was maintained at 65% and decreased gradually to 45% once the clones started flowering. Lost Coast Plant Therapy (Plant Protector, Inc. Loleta, CA) was applied to the clones biweekly at a dilution of 30mL per 4 liters to control pests.

Plant stress treatments

Four weeks old, cloned plants were subjected to the following 11 stress treatments. Except for the UV light treatment, all the remaining 10 stress treatments were performed in the greenhouse on the same day. For drought and salt stress, hemp plants were subjected to 100mM of mannitol and 100mM of NaCl for eight hours, respectively. For heavy metal stress, hemp plants were subjected to either 200μM of $CuSO_4$, 100μM $CdCl_2$, or 100μM of $Pb(NO_3)^2$ and 200μM of $ZnSO_4$ for eight hours. For the plant hormone treatment, plants were soaked in 100μM abscisic acid (ABA), 100μM of methyl jasmonate (MeJA), 1mM of gibberellic acid (GA_3), or 100μM of salicylic acid (SA) for eight hours. For UV treatment, hemp plants were exposed to UV-C radiation for 10 minutes. After stress treatment, the 3rd and 4th leaves from the top of the plant were sampled and immediately frozen in liquid nitrogen and stored in the -80°C freezer until total RNA was extracted. All the treatments were performed in three biological replicates. For the mock plants, distilled water was used to soak the hemp plants.

Total RNA extraction and cDNA synthesis

Total RNA from 100mg of each plant sample was extracted using the Spectrum^{TM} Plant Total RNA kit (Sigma Aldrich, St. Louis, MO, USA). RNA concentration and absorbance ratios (A260/280 and A260/230) were measured using a NanoVue Plus spectrophotometer (General Electric Healthcare Limited, UK) to measure the quantity and quality of the total RNA. After treatment with DNase I (TaKaRa Bio, Dalian, China) to remove genomic DNA contamination, 2μg of total RNA was used to synthesize cDNA using the high-capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol.

Candidate reference genes selection and primer design

We identified 13 candidate reference genes (Table 1) and two target genes using a BLAST search from NCBI (https://www.ncbi.nlm.nih.gov/) and the *Cannabis sativa* genome browser gateway (http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway). The *Cannabis sativa* genome browser gateway is based on the Purple Kush strain of medical marijuana. Primers were designed based on the sequences of 13 genes using Primer3 Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) with the criteria as follows: amplicon size 80–200bp, primer size 18–24bp, Tm 60°C, GC content 45–60% (Table 1). All primer sequences are listed in Table 1.

qRT-PCR amplification

PCR was performed using cDNA as the template to confirm the specificity of the primers to the target genes. Using a 2% (w/v) agarose gel, all PCR products were analyzed using electrophoresis to confirm a single band of the expected size for each of the primer pairs. To test the PCR amplification efficiency, the regression coefficient (R^2) for each gene was calculated using a standard curve generated from a fivefold dilution series of cDNA (1, 1/10, 1/100, 1/1000, and 1/10000). qRT-PCR was performed with 5μL of SYBR Select Master Mix (Applied Biosystems, Waltham, MA, USA) in a 10μL total reaction mixture containing 400nM of the gene-specific primers and 1μL of cDNA. PCR reaction was performed using a Bio-Rad CFX96 system (Bio-Rad, Hercules, CA, USA) under the following reaction conditions: Initial denaturation at 95°C for 10 minutes, 40 cycles of 95°C for 10 seconds, and 60°C for 1 minute. Two technical replicates were used for each biological replicate and average Ct was used for data analysis. As a negative control, water and total RNA were used instead of cDNA to confirm that there was no amplification from contaminated DNA or hemp genomic DNA, respectively.

Gene expression stability and statistical analysis

Boxplots of quantitative cycle (Cq) values for the 13 candidate reference genes were depicted in all leaf samples with every treatment using the boxplot R package to show the variation of each gene expression. The expression of 13 reference genes was analyzed under 11 different stresses using four algorithms, geNorm, NormFinder, BestKeeper, and RefFinder to rank the stability of the candidate reference genes. The pairwise variation (Vn/Vn+1) between two sequential normalization factors was calculated with geNorm to determine the optimal number of candidate reference genes for accurate normalization.

Validation of identified reference genes

Two cannabinoid pathway genes, *CBDAS* and *AAE1*, were used as target genes to validate the reliability of the selected reference genes using the most stable candidate reference genes as well as the least stable reference genes. Primer design, qRT-PCR analysis, and gene expression stability analysis for these genes were performed as described above. Relative gene expression levels of *CBDAS* and *AAE1* were measured and calculated using the $2^{-\Delta\Delta Ct}$ method [26]. Statistical analysis was performed using a 1-way ANOVA with Tukey's multiple comparison test (a=0.05).

Results

PCR specificity and amplification efficiency of the candidate reference genes

Thirteen housekeeping genes (*18S*, *40S*, *CHAL*, *UBE2*, *EF-1α*, *F-box*, *GAD*, *PCS1*, *PP2A*, *SAND*, *TATA*, *TIP41*, and *TUB*) were identified from NCBI and the *Cannabis sativa* genome browser gateway based on a homology search with *Arabidopsis* genes (Table 1). Specific primers were designed and used to confirm their specificities based on the amplification efficiency and specificity (Table 1). All gene primers amplified single bands with predicted sizes resolved in agarose gel electrophoresis (Figure 1, Figure S1). For the qRT-PCR amplification, the PCR efficiency (%) ranged from 91.22 to 113.87, and the regression coefficient (R^2) varied from 0.9893 to 0.9994 (Table 1).

Ct values of candidate reference genes

Transcript abundances of 13 candidate reference genes were assessed by qRT-PCR for each gene, tested in triplicates across all 11 treatments and control, which was in a total of 36 biological samples (Figure 2). A majority of the candidate reference genes Ct values ranged from 20 to 30. The *PCS1* gene showed the lowest expression level with the highest Ct values between 27.4 and 32.1. The *EF-1α* gene showed the highest expression level with the lowest Ct values ranging from 18.9 to 25.3. The *CHAL* gene displayed the highest difference among all 36 samples tested, with a minimum Ct value of 22.3 and a maximum Ct value of 30.9. These Ct value analyses indicated that the transcription levels of candidate reference genes are unstable under different stress conditions.

Analysis of gene expression stability by geNorm

The geNorm was used for evaluating the expression stability of the 13 candidate reference genes (Table 2). Data analysis was calculated based on individual 11 different treatments and three different groups of treatments such as osmotic stress (OS: mannitol, NaCl), heavy metal stress (HM: $CdCl_2$, $CuSO_4$, $PbNO_3$, $ZnSO_4$), and hormonal stimuli (PH: ABA, GA_3, MeJA, SA). The total ranking was also shown by combining all 11 treatments together. This algorithm evaluated the gene expression stability (M) based on the average pairwise variation of all tested genes[27]. In this analysis, the lower the M value, the more stable the gene expression. A reference gene that has an M value less than 1.5 is used for qRT-PCR. *PP2A* and *TIP41* were the most stable reference genes with the lowest M value (0.46) whereas *CHAL* had an M

value of 1.07 and was ranked as the least stable gene. Individually, *EF-1a* and *SAND* were the most stably expressed genes under osmotic stresses while *F-box* and *TATA* were the least stably expressed genes with an M value of 0.22. The *TUB* and *TATA* genes showed the lowest M values of 0.16 among all of the heavy metal stressed clones. Exposure to hormonal stimuli resulted in *PP2A* and *F-box* to be the most stable with an M value of 0.27 and *CHAL* to be the least stable with an M value of 0.75.

F-box was ranked as the second least stably expressed gene under both osmotic and heavy metal stresses, but it was ranked as the first and second most stable gene under UV and plant hormone treatments, respectively. The *TATA* gene was least stably expressed under osmotic stresses but was among the top two and three under heavy metal stress and plant hormone stimulus, respectively. *CHAL* was the least stably expressed in response to UV light application.

geNorm can determine the minimal number of reference genes that should be used to obtain an accurate normalization. The optimal number of reference genes was determined based on the pairwise variation (Vn) between two normalization factors (NFn) composed of an increasing number of reference genes[27]. The threshold value (Vn/Vn+1 = 0.15) indicates if the number of reference genes less than or equal to the value of n is sufficient to use as a reference gene. As shown in Figure 3, the pairwise variation value V2/V3 of all experimental samples was less than 0.15, demonstrating that two reference genes should be sufficient for normalization under all conditions tested.

Analysis of gene expression stability by NormFinder

NormFinder is a quantity-model-based software and uses complex statistical models to compute the variation between the expression of genes across different biological groups[28]. The lowest expression stability value represents the most stable reference genes. Our results obtained from NormFinder analysis are summarized in Table 3. The *EF-1a* and *TUB* genes were the most stably expressed in all samples and were ranked as fourth and third by geNorm, respectively. The *F-box*, *TATA*, and *CHAL* were ranked as the three least stable genes both by NormFinder as well as geNorm. The *TUB*, *PCS1*, and *TATA* genes were the most stably expressed under osmotic stress, heavy metal stress, and plant hormone stimuli, respectively. In comparison to geNorm, *TUB*, *PCS1*, and *TATA* were ranked as third, sixth, and third positions in each category, respectively. The least stably expressed reference genes under osmotic stress (*CHAL*, *F-box*, *TATA*), heavy metal stress (*CHAL*, *F-box*, *TATA*), and plant hormone stimuli (*40S*, *GAD*, *CHAL*) had similar rankings when compared to geNorm rankings. The *GAD* and *F-box* genes were found to be the most stable reference genes under UV stress while *PCS1* and *CHAL* were the least stable which was a similar trend observed in the geNorm analysis.

Analysis of gene expression stability by BestKeeper

The BestKeeper is an excel-based algorithm and uses standard deviation (SD) and coefficient of variation (CV) data of the average Ct values for specific treatments[29](Table 4). Lower CV ± SD values represent higher stability. When using the BestKeeper algorithm, genes with an SD value > 1 are considered as an undesirable reference gene[22]. When all samples were taken into consideration, *TATA* (SD=0.74), *40S*

(SD=0.79), *PCS1* (SD=0.84), *PP2A* (SD=0.90), and *TUB* (SD=0.99) were determined to be reliable reference genes. *TATA* showed the lowest SD among all 13 reference genes in all samples and the SD values were greater than 1 in osmotic stress (1.29) and mannitol (1.82). The *40S* gene was ranked as the second most stable candidate in all samples tested, but the SD value of *40S* under NaCl and $PbNO_3$ stresses were 1.09 and 1.36, respectively. *PCS1* was ranked at the third position in all samples tested and SD values were below 1 in any individual treatment and also the three treatment groups. The *CHAL* gene displayed the highest SD value with 1.95 in all samples indicating that this gene is unsuitable for gene expression normalization. The *CHAL* gene exhibited an SD value less than 1 only under GA_3 (0.71), ABA (0.27), $CdCl_2$ (0.58), and $CuSO_4$ (0.98) treatments.

Analysis of gene expression stability by RefFinder.

RefFinder is a web-based tool for comprehensive analysis that integrates geNorm, NormFinder, Delta Ct, and BestKeeper approaches[30]. The reference genes were ranked from the most stable (M value is the lowest) to least stable expression (M value is the highest) using Best Keeper (Table 5). Among them, the most stable candidate was the *EF-1a* gene followed by the *TUB* gene in all samples. The *EF-1a* and *TUB* genes were also ranked at third and first places under osmotic stress conditions, respectively. The *TATA* gene was most stably expressed under heavy metal and plant hormone treatments while this gene was the least stable under osmotic stress. The *CHAL* gene was ranked as the least stable gene in all samples tested. The *GAD* and *CHAL* genes were the most and least stably expressed genes respectively under UV application, which was the same findings as the NormFinder software.

Validation of selected reference genes

To validate the selected reference genes, gene expression levels of *AAE1* and *CBDAS* were measured (Figure 4). Each of the two most stable reference genes, *EF-1a* and *TUB*, a combination of these two stable reference genes (*EF-1a*+*TUB*), and the least stable reference gene (*CHAL*) were used as internal controls. *AAE1* expression was significantly reduced under drought (Mannitol) and salinity (NaCl) stresses. *EF-1a*, *TUB*, a combination of *EF-1a* and *TUB* were used for normalization of qRT-PCR analysis. There was no significant difference in the *AAE1* expression between the mock treatment and osmotically stressed samples (Mannitol and NaCl) when *CHAL* was used as an internal control. The expression of *CBDAS* was also reduced under osmotic stresses when expression data was normalized with *EF-1a*, *TUB*, and a combination of *EF-1a* and *TUB* except when the *CBDAS* expression under NaCl stress was normalized with the *TUB* gene. When *CHAL* was used as a reference gene, *CBDAS* gene expression was reduced under mannitol treatment but there was no difference between mock and NaCl treatments.

Discussion

Hemp is a fiber-type *Cannabis sativa* and has gained importance as a medicinal crop, due to its potential for the production of secondary metabolites such as cannabinoids, terpenes, and phenolic compounds[31]. According to Schluttenhofer and Yuan (2017)[4], hemp was cultivated for commercial or research purposes

in at least 47 countries in 2017 and the global hemp market doubled from the year 2016 to 2020. Recent research has indicated comprehensive gene expression analysis is aimed at elucidating the metabolic pathways for cannabinoids and terpene synthesis to improve hemp traits[32, 33, 34]. To validate this data, qRT-PCR analysis is suitable, however, appropriate hemp reference genes for accurate gene expression analysis has not been well established.

In this report, we evaluated 13 hemp reference genes under 11 different stress conditions. Previous studies in other plant species revealed that different tissues and different environmental conditions would require unique reference genes to accurately interpret the expression of specific genes[35, 36]. Eleven different conditions including osmotic stresses, heavy metal stresses, plant hormone stimulus, and UV light application were reported to affect the cannabinoid synthesis[37, 38, 39, 40]. Results obtained in this study from four programs: geNorm, NormFinder, BestKeeper, and RefFinder were not consistent, particularly BestKeeper which was much more distinct from the other software methods. This finding was expected because the BestKeeper algorithm evaluates data quite differently when compared to the three other programs[41].

To rank the most suitable reference genes across all treatments, there was no unanimity when compared among four different algorithms. In most cases, one candidate gene was ranked as the most stable gene by two or three programs, which demonstrated that it could potentially be a good reference gene under various treatments. Based on the combined rankings of the four programs used in our study, the overall results indicated that the most stable genes varied while the least stable genes were almost the same. Across all plants tested, both NormFinder and RefFinder determined *EF-1a* as the most stable gene in all samples tested. In previous reports, *EF-1a* was demonstrated to be the most stable gene under different stresses in a variety of crops such as tobacco[42], maize[43], soybean[44], potato[45, 46]. Interestingly, this gene was not the most stable in any of the three groups (OS, HM, PH). The *TUB* gene appeared to be best the candidate under osmotic stresses because this gene was ranked as the most stable by both NormFinder and RefFinder which is consistent with the results obtained in Parsley under abiotic stresses[47]. Under heavy metal stress, *TATA* was ranked as the most stable gene by BestKeeper and RefFinder and the second most stable gene by geNorm. This gene was also ranked as the best reference gene in hormone stimuli by NormFinder and RefFinder. Interestingly, *TATA* was the least stable gene under osmotic stresses by geNorm, NormFinder, and RefFinder. *TUB* was the most stably expressed gene under osmotic stresses whereas *TATA* was ranked as the best stable gene under both heavy metal stress and hormone stimuli. Unlike most stable genes, *CHAL* was found to be the least stable gene in most of the rankings with all samples and the three treatment groups (OS, HM, PH) when analyzed by all four programs. According to Wang et al., (2015)[48], candidate genes showing a high-level of variation of Ct values should be avoided as internal controls. Our results showed that variation of the Ct value in *CHAL* was highest among all 13 reference genes (Figure 2), which is consistent with the fact that *CHAL* was ranked as the least stable by all four programs used in this study.

In previous *Cannabis* qRT-PCR studies, the *F-box* gene has been used as an internal control for qRT-PCR26, 38, 49). Mangeot-Peter et al., (2016)[24] performed the reference gene analysis in hemp stems and concluded that the *F-box* gene was ranked as one of the most stable genes among 12 reference genes tested under normal conditions. In this study, however, the *F-box* gene was the second least stable gene by RefFinder and the third least stable gene determined by both the geNorm and NormFinder programs when all samples were analyzed. Based on our group rankings (OS, HM, PH), *F-box* was ranked the second least stable genes by geNorm, BestKeeper, and NormFinder under both osmotic and heavy metal stresses. On the contrary, *F-box* was stably expressed under normal conditions in hemp leaves (data not shown) and relatively stable under hormone stimuli as evident by its second position as ranked by both geNorm and RefFinder. These results indicate that *F-box* may not be a suitable reference gene for hemp qRT-PCR analysis under osmotic and heavy metal stresses. However, it could be acceptable as a reference gene under normal and plant hormone treatments. Overall, our study suggests that the *F-box* gene may not be the best reference gene for *C. sativa*, particularly in plant stress-related studies.

Many studies have proved that the use of more than one reference gene enables the possibility of avoiding variations and achieving more accurate normalization of qPCR data[27]. To assess the optimal number of reference genes for the normalization of qRT-PCR data, we used the geNorm program to perform a stepwise calculation of the pairwise variation (Vn/Vn+1) between sequential normalization factors. In this analysis, a Vn/Vn+1 < 0.15 indicates that introducing an additional reference gene for normalization is not necessary. Under all treatments, V2/V3 values were less than 0.15, which indicated that two reference genes were enough for normalization of the real-time PCR data under any treatments in this study.

To validate the reliability of the selected reference genes, we measured the relative expression of two cannabinoids pathway genes using *EF-1a* and *TUB* as the most stable reference genes as well as *CHAL* as the least stable reference gene (Figure 4). Since CBDA content is known to be decreased by the influence of osmotic stress[37], we measured the expression of *AAE1* and *CBDAS* genes that are involved in the rate-determining enzymatic reactions leading to CBDA synthesis under drought and salinity stresses[50, 51]. The expression of these two genes was significantly reduced under drought and salinity stresses when qRT-PCR data was normalized by *EF-1a*, *TUB*, and the combination of *EF-1a* and *TUB*. Notably, the expression level of both of these genes were normalized by *CHAL* under salinity stress and did not show a significant difference when compared with mock plants. These results suggest that *EF-1a* and *TUB* genes individually or in combination are suitable reference genes for hemp under osmotic stresses. Our validation study demonstrated the effectiveness of the ranking of reference genes by the programs used in particular geNorm, NormFinder, and RefFinder.

To the best of our knowledge, this study is the first report that performed a systematic analysis of hemp reference genes under different abiotic stresses and hormonal stimuli. The knowledge obtained in this study could contribute to enhancing future hemp research related to the elucidation of mechanisms involved in the synthesis, transport, and accumulation of abundant secondary metabolites in hemp.

Declarations

Acknowledgements

We would like to thank Dr. Yuka Imamura, PhD of Penn State College of Medicine Genome Sciences and Bioinformatics Facility for data analysis services. The authors would like to thank Ms. Jessica Wolf for proofreading the manuscript. This project was supported by the PA Options for Wellness.

Author contributions

M.D. and S.R. conceived and designed the experiments; M.D. performed the experiments. M.D, S.P. and Z.S. analyzed the data; H.G. prepared plant materials; M.D. wrote the paper; S.R., S.P., V.S., H.G., R. A., A. P. and Z.S. revised the manuscript.

Competing Interests Statement

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.R.

References

1. Frassinetti, S. *et al*. Nutraceutical potential of hemp (*Cannabis sativa*) seeds and sprouts. *Food Chemistry*. 262, 56–66. (2018).
2. Bonini, S. A. *et al. Cannabis sativa*: A comprehensive ethnopharmacological review of a medicinal plant with a long history. *Journal of Ethnopharmacology*. 227, 300-315. (2018).
3. Izzo, A. A., Borrelli, F. Capasso, R., Di Marzo, V. & Mechoulam, R. Non-psychotropic plant cannabinoids: new therapeutic opportunities from an ancient herb. *Trends Pharmacological Science*. 30, 515–527. (2009)
4. Schluttenhofer, C. & Yuan, L. Challenges towards Revitalizing Hemp: A Multifaceted Crop. *Trends in Plant Science*. 22, 917-929. (2017).
5. McPartland, J. M. & Russo, E. B*. Cannabis* and *Cannabis Journal of Cannabis Therapeutics*. 1, 103-132. (2001).
6. Chadwick, B., Miller, M. & Hurd, Y. *Cannabis* Use during Adolescent Development: Susceptibility to Psychiatric Illness. 4, (2013).
7. Vasas, A. & Hohmann, J. Euphorbia Diterpenes: Isolation, Structure, Biological Activity, and Synthesis (2008–2012*). Chemical Reviews* 114, 8579-8612, (2014).
8. Van Bakel, H. et al. The draft genome and transcriptome of *Cannabis Genome Biology*. 12, R102 (2011)

9. Braich, S., Baillie, R. C., Jewell, L. S., Spangenberg, G. C. & Cogan, N. O. I. Generation of a Comprehensive Transcriptome Atlas and Transcriptome Dynamics in Medicinal *Cannabis*. *Scientific Reports* 9, 16583, (2019).
10. McGarvey, P. et al. De novo assembly and annotation of transcriptomes from two cultivars of *Cannabis sativa* with different cannabinoid profiles. *Gene* 762, 145026, (2020).
11. Vincent, D., Binos, S., Rochfort, S. & Spangenberg, G. Top-Down Proteomics of Medicinal *Cannabis*. *Proteomes* 7, (2019).
12. Gülck, T. & Møller, B. L. Phytocannabinoids: Origins and Biosynthesis. *Trends Plant Sci* 25, 985-1004, (2020).
13. Booth, J. K. & Bohlmann, J. Terpenes in *Cannabis sativa* - From plant genome to humans. *Plant science : an international journal of experimental plant biology* 284, 67-72, (2019).
14. Marks, M. D. et al. Identification of candidate genes affecting Delta9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. *J Exp Bot* 60, 3715-3726, (2009).
15. Bassolino, L. et al. In Silico Identification of MYB and bHLH Families Reveals Candidate Transcription Factors for Secondary Metabolic Pathways in *Cannabis sativa* Plants (Basel) 9, 1540, (2020).
16. Kevil, C. G. et al. An Improved, Rapid Northern Protocol. Biochemical and Biophysical Research Communications 238, 277-279, (1997).
17. Stalder, A. K., Pagenstecher, A., Kincaid, C. L. & Campbell, I. L. Analysis of Gene Expression by Multiprobe RNase Protection Assay. *Methods in molecular medicine* 22, 53-66, (1999).
18. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. Serial analysis of gene expression. *Science (New York, N.Y.)* 270, 484-487, (1995).
19. Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nature Protocols* 1, 1559-1582, (2006).
20. Taylor, S. C. et al. The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends in Biotechnology* 37, 761-774, (2019).
21. Zhang, L. et al. Selection of reference genes for qRT-PCR and expression analysis of high-altitude-related genes in grassland caterpillars (Lepidoptera: Erebidae: Gynaephora) along an altitude gradient. *Ecol Evol* 7, 9054-9065, (2017).
22. Yu, Y. et al. Selection of Reference Genes for qPCR Analyses of Gene Expression in Ramie Leaves and Roots across Eleven Abiotic/Biotic Treatments. *Scientific Reports* 9, 20004, (2019).
23. Ambroise, V. et al. Selection of Appropriate Reference Genes for Gene Expression Analysis under Abiotic Stresses in Salix viminalis. *Int J Mol Sci* 20, 4210, (2019).
24. Mangeot-Peter, L., Legay, S., Hausman, J.-F., Esposito, S. & Guerriero, G. Identification of Reference Genes for RT-qPCR Data Normalization in *Cannabis sativa* Stem Tissues. 17, 1556 (2016).
25. Guo, R. et al. Evaluation of reference genes for RT-qPCR analysis in wild and cultivated *Bioscience, Biotechnology, and Biochemistry* 82, 1902-1910, (2018).

26. Deguchi, M. et al. Establishment and optimization of a hemp (*Cannabis sativa*) agroinfiltration system for gene expression and silencing studies. *Scientific Reports* 10, 3504, (2020a).
27. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3, research0034.0031, (2002).
28. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research* 64, 5245-5250, (2004).
29. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnology Letters* 26, 509-515, (2004).
30. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Molecular Biology 80, 75-84, (2012).
31. Andre, C. M., Hausman, J. F. & Guerriero, G. *Cannabis sativa*: The plant of the thousand and one molecules. *Frontiers in Plant Science*. 7, 1–19 (2016).
32. Prentout, D. et al. A high-throughput segregation analysis identifies the sex chromosomes of *Cannabis sativa*. 721324, J bioRxiv (2019).
33. Gao, C. et al. Genome-Wide Expression Profiles of Hemp (*Cannabis sativa*) in Response to Drought Stress. *International Journal of Genomics* 2018, 3057272, (2018).
34. Adal, A. M., Doshi, K., Holbrook, L. & Mahmoud, S. S. Comparative RNA-Seq analysis reveals genes associated with masculinization in female *Cannabis sativa*. *Planta* 253, 17, (2021).
35. Poli, M., Salvi, S., Li, M. & Varotto, C. Selection of reference genes suitable for normalization of qPCR data under abiotic stresses in bioenergy crop Arundo donax L. *Scientific Reports* 7, 10719, (2017).
36. Xu, L. et al. Validation of Reference Genes for Quantitative Real-Time PCR during Bicolor Tepal Development in Asiatic Hybrid Lilies (*Lilium*). *Frontiers in plant science*, 8, 669 (2017).
37. Yep, B., Gale, N. V. & Zheng, Y. Aquaponic and Hydroponic Solutions Modulate NaCl-Induced Stress in Drug-Type *Cannabis sativa* 11, (2020).
38. Husain, R. et al. Enhanced tolerance of industrial hemp (*Cannabis sativa*) plants on abandoned mine land soil leads to overexpression of cannabinoids. *PLoS One* 14, e0221570, 0221570 (2019).
39. Burgel, L., Hartung, J., Schibano, D. & Graeff-Hönninger, S. Impact of Different Phytohormones on Morphology, Yield and Cannabinoid Content of *Cannabis sativa* Plants (Basel) 9, 725, (2020).
40. Bilodeau, E., S., Wu, B.-S., Rufyikiri, A.-S., MacPherson, S. & Lefsrud, M. An Update on Plant Photobiology and Implications for *Cannabis* 10, (2019).
41. Guo, J., Ling, H., Wu, Q., Xu, L. & Que, Y. The choice of reference genes for assessing gene expression in sugarcane under salinity and drought stresses. *Scientific Reports* 4, 7042, (2014).
42. 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. *Molecular genetics and genomics: MGG* 283, 233-241, (2010).

43. Lin, Y. et al. Validation of Potential Reference Genes for qPCR in Maize across Abiotic Stresses, Hormone Treatments, and Tissue Types. *PLOS ONE* 9, e95445, (2014).
44. Saraiva, K. D., Fernandes de Melo, D., Morais, V. D., Vasconcelos, I. M. & Costa, J. H. Selection of suitable soybean EF1α genes as internal controls for real-time PCR analyses of tissues during plant development and under stress conditions. *Plant cell reports* 33, 1453-1465, (2014).
45. Nicot, N., Hausman, J.-F., Hoffmann, L. & Evers, D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot* 56, 2907-2914, J Journal of Experimental Botany (2005).
46. Tang, X., Zhang, N., Si, H., & Calderón-Urrea, A. Selection and validation of reference genes for RT-qPCR analysis in potato under abiotic stress.*Plant methods*, 13, 85 (2017).
47. Li, M.-Y., Song, X., Wang, F. & Xiong, A.-S. Suitable Reference Genes for Accurate Gene Expression Analysis in Parsley (Petroselinum crispum) for Abiotic Stresses and Hormone Stimuli. *Front Plant Sci* 7, 1481-1481, (2016).
48. Wang, H.-L. et al. Evaluation of Appropriate Reference Genes for Reverse Transcription-Quantitative PCR Studies in Different Tissues of a Desert Poplar via Comparision of Different Algorithms. *Int J Mol Sci* 16, 20468-20491, (2015).
49. Guerriero, G. et al. Identification of fasciclin-like arabinogalactan proteins in textile hemp (*Cannabis sativa*): in silico analyses and gene expression patterns in different tissues. *BMC Genomics* 18, 741-741, (2017).
50. Deguchi, M. et al. Metabolic Engineering Strategies of Industrial Hemp (*Cannabis sativa*): A Brief Review of the Advances and Challenges. 11, (2020b).
51. Stout, J. M., Boubakir, Z., Ambrose, S. J., Purves, R. W. & Page, J. E. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* 71, 353-365, (2012).

Tables

Due to technical limitations the Tables are available as a download in the Supplementary Files.

Figures

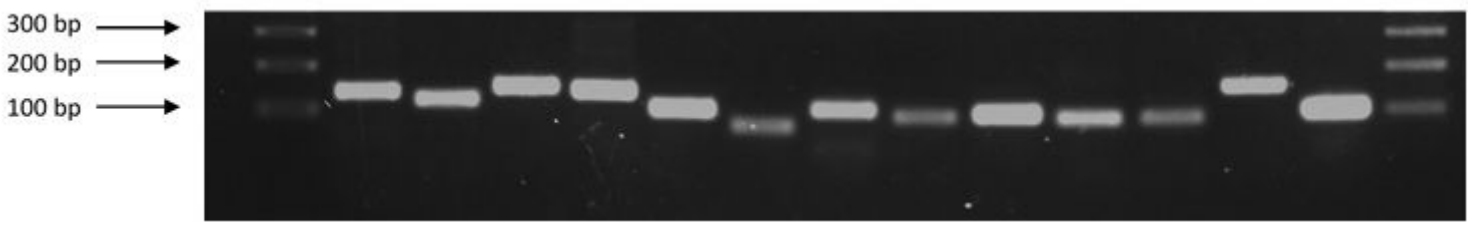


Figure 1

Amplification results for 13 candidate genes using cDNA synthesized from hemp leaf sample to confirm primer specificity and amplicon size.

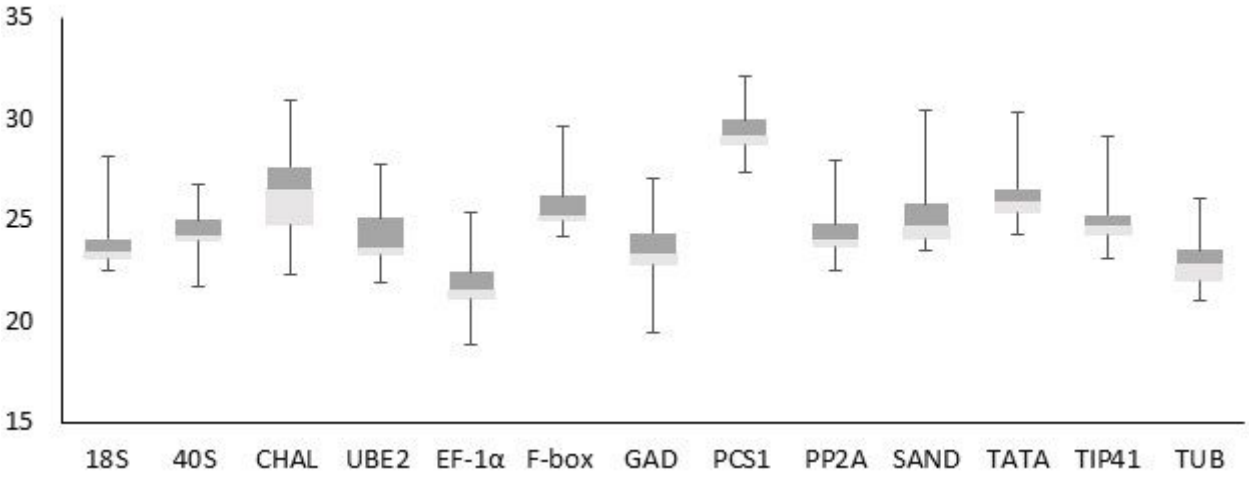
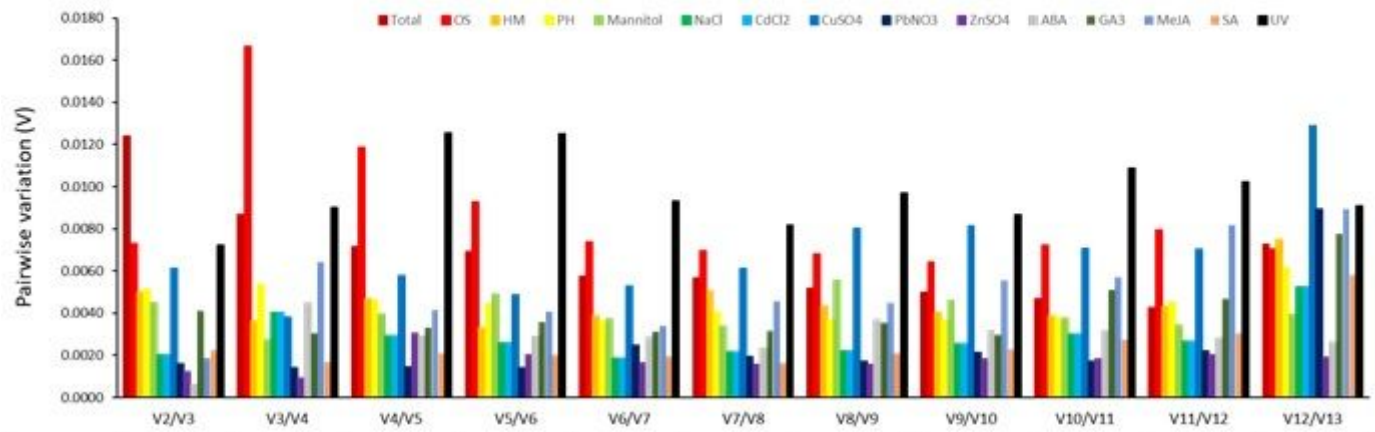


Figure 2

Expression level variability of each candidate reference gene to examine all leaf samples (n=36). Boxes indicate the 25th and 75th percentiles, whisker caps represent the minimum and maximum values, lines across the box represent the median Ct-values.



	V2/V3	V3/V4	V4/V5	V5/V6	V6/V7	V7/V8	V8/V9	V9/V10	V10/V11	V11/V12	V12/V13
Total	0.0124	0.0087	0.0072	0.0069	0.0058	0.0057	0.0052	0.0050	0.0047	0.0043	0.0073
OS	0.0073	0.0167	0.0119	0.0093	0.0074	0.0070	0.0068	0.0064	0.0072	0.0079	0.0070
HM	0.0050	0.0036	0.0047	0.0033	0.0039	0.0050	0.0044	0.0040	0.0039	0.0043	0.0075
PH	0.0051	0.0053	0.0046	0.0045	0.0036	0.0041	0.0037	0.0037	0.0038	0.0045	0.0062
Mannitol	0.0045	0.0027	0.0039	0.0049	0.0037	0.0034	0.0055	0.0046	0.0038	0.0034	0.0039
NaCl	0.0020	0.0040	0.0029	0.0026	0.0019	0.0022	0.0022	0.0025	0.0030	0.0027	0.0053
CdCl2	0.0020	0.0040	0.0029	0.0026	0.0019	0.0022	0.0022	0.0025	0.0030	0.0027	0.0053
CuSO4	0.0061	0.0038	0.0058	0.0049	0.0053	0.0061	0.0080	0.0081	0.0071	0.0070	0.0129
PbNO3	0.0016	0.0014	0.0014	0.0014	0.0025	0.0019	0.0017	0.0021	0.0017	0.0022	0.0089
ZnSO4	0.0012	0.0009	0.0030	0.0020	0.0016	0.0016	0.0016	0.0018	0.0018	0.0020	0.0019
ABA	0.0006	0.0045	0.0029	0.0029	0.0028	0.0023	0.0037	0.0032	0.0032	0.0028	0.0026
GA3	0.0041	0.0031	0.0033	0.0036	0.0031	0.0032	0.0035	0.0030	0.0051	0.0047	0.0078
MeJA	0.0019	0.0064	0.0041	0.0041	0.0034	0.0046	0.0045	0.0056	0.0057	0.0082	0.0089
SA	0.0023	0.0017	0.0021	0.0020	0.0019	0.0016	0.0021	0.0023	0.0027	0.0030	0.0058
UV	0.0072	0.0090	0.0125	0.0125	0.0093	0.0082	0.0097	0.0087	0.0109	0.0102	0.0091

Figure 3

Determination of best reference gene number calculated by geNorm pairwise variation (Vn/Vn+1) under different stress treatments in hemp leaf.

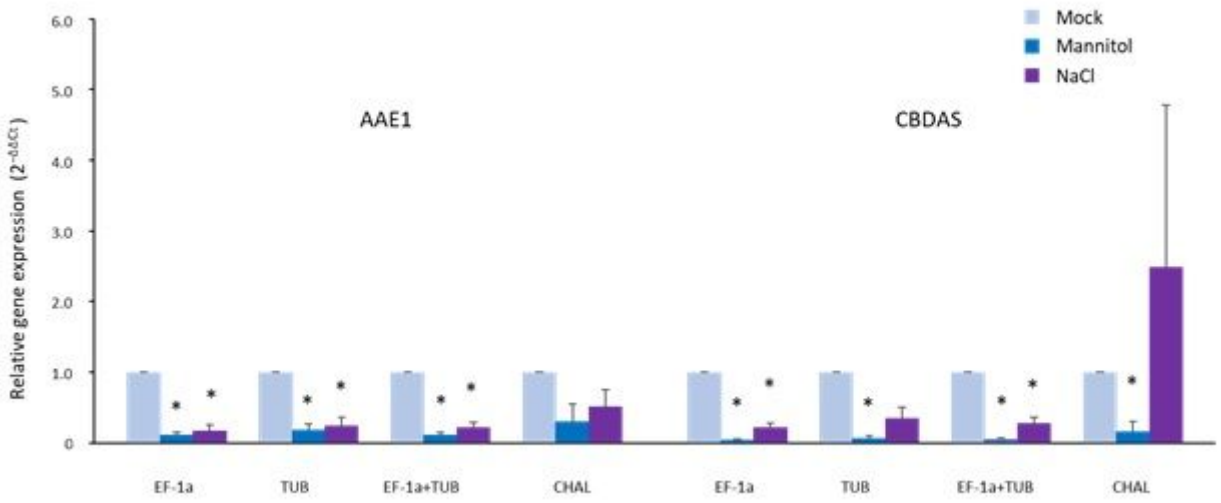


Figure 4

Relative expression of target genes in hemp leaf under osmotic stresses using most and least stably expressed reference genes for normalization. Error bars for qRT-PCR show the standard error of three replicates for EF-1a, TUB, and CHAL, and six replicates for mixture of EF-1a and TUB. Asterisk indicates a significant difference in the comparison with mock treatment by the statistical analysis ($P < 0.05$) in paired t-tests.

Supplementary Files

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

- Supplementaryinformation1.pdf
- Tables.pdf