UDP and NTF2 are the most consistently expressed genes in *Panax ginseng* roots at different growth stages

MEICHEN LIU, QUN WANG, HONGMEI XIE, SHICHAO LIU, SIMING WANG, HUI ZHANG and YU ZHAO

Center of Chinese Medicine and Bio-Engineering Research, Changchun University of Chinese Medicine, Changchun, Jilin 130117, P.R. China

Received January 4, 2016; Accepted December 19, 2016

DOI: 10.3892/mmr.2017.6494

Abstract. Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis relies on normalization against a consistently expressed reference gene. However, it has been reported that reference gene expression levels often vary markedly between samples as they are usually selected based solely on convention. The advent of RNA sequencing technology offers the opportunity to select reference genes with the least variability in steady-state transcript levels. To identify the most consistently stable genes, which are a prerequisite for obtaining reliable gene expression data, the present study analyzed transcriptomes from six *Panax ginseng* transcriptome data sets, representing six growth stages, and selected 21 candidate reference genes for screening using RT-qPCR. Of the 21 candidate genes, 13 had not been reported previously. The geNorm, NormFinder and BestKeeper programs were used to analyze the stability of the 21 candidate reference genes. The results showed that UDP-N-acetylgalactosamine transporter and nuclear transport factor 2 were likely to be the optimal combination of reference genes for use in investigations of ginseng. The novel reference genes were validated by correlating the gene expression profiles of four pathogenesis-related protein genes generated from RT-qPCR, with their expression levels calculated from the RNA sequencing data. The expression levels were well correlated, which demonstrated their value in performing RT-qPCR analyses in ginseng.

Correspondence to: Professor Yu Zhao, Center of Chinese Medicine and Bio-Engineering Research, Changchun University of Chinese Medicine, 1035 Boshuo Road, Changchun, Jilin 130117, P.R. China
E-mail: cnzhaoyu1972@126.com

Abbreviations: UDP, UDP-N-acetylgalactosamine transporter; NTF2, nuclear transport factor 2; RT-qPCR, reverse transcription-quantitative polymerase chain reaction; RNA-Seq, RNA sequencing; RPKM, reads per kilobase of transcript per million; UBE2, ubiquitin conjugating enzyme isoform 2; GAGA, GAGA-binding transcriptional activator; PDI, protein disulfide isomerase; MPP, mitochondrial-processing peptidase; G6P, glucose-6-phosphate; PPS, probable prefoldin subunit 5; ARF1, auxin response factor 1; 3-IPMDH, putative 3-isopropylmalate dehydrogenase; ECH1, δ(3,5)-δ(2,4)-dienoyl-CoA isomerase, mitochondrial; EIF-4E1, eukaryotic translation initiation factor 4E-1; ACT1, actin 1; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; UBQ, ubiquitin; CYP, cyclophilin; PP2A, PP2Ac-3-phosphatase 2A isoform 3

Key words: Panax ginseng, RNA sequencing, reverse transcription-quantitative polymerase chain reaction, reference gene

Introduction

Ginseng (*Panax ginseng* CA Meyer) is a medicinal herb, which has been used in Asia for >1,000 years (1). Ginseng root, the most commonly used region of the plant, contains bioactive constituents with complex and multiple pharmacological effects (2). Previous reports have demonstrated that ginseng grown for longer durations shows improved efficacy and a greater concentration of bioactive components, including ginsenosides (3-6). Various studies have focused on the genetics underlying these findings, particularly on marker gene identification or authentication, and on key enzymes involved in the ginsenoside biosynthetic pathway (7,8). However, the molecular mechanisms remain to be fully elucidated.

Gene expression analysis is an effective and widely used approach to identify marker genes and elucidate biological mechanisms. Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis is the preferred method for gene expression analysis owing to its rapidity, sensitivity and specificity (9). Measurements of the expression levels of genes of interest are normalized against a consistently expressed reference gene to improve the accuracy of the RT-qPCR results. However, reference genes are usually selected based solely on convention, and reference gene levels have been found to vary substantially among samples (10-13). In previous studies analyzing gene expression in ginseng, reference genes, predominantly actin 1 (ACT1) and 18S rRNA, have been selected based on previous studies of various plant species (14-18). However, it has been shown that the expression levels of these two genes are not consistent in different ginseng organs (19). Thus, the selection of suitable reference genes is an important prerequisite for gene expression analysis in ginseng.

RNA sequencing (RNA-Seq) is an ideal method to identify the most consistently expressed genes for use as reference genes (20), as large-scale gene expression data can be generated

at the same time and gene expression values can be converted to reads per kilobase of transcript per million (RPKM) (21) for direct comparisons between gene data sets. Publically available RNA-Seq data have been used previously to identify superior reference genes (21-24). In the present study, RNA-Seq data was obtained from our previous ginseng RNA-Seq sequencing project, which included a panel of six ginseng transcriptome databases, and were used to identify reference genes with lower variations across multiple developmental stages in ginseng root. Statistical methods were implemented in geNorm (25), NormFinder (26) and BestKeeper (27), and the effectiveness of the candidate genes for RT-qPCR normalization were then compared with traditional reference genes.

Materials and methods

Ginseng samples. *P. ginseng* CA Meyer plants were used in the present study. The *P. ginseng* samples, which had been grown for 3, 5, 7, 10, 15 and 20 years, were originally collected from Fu-song County (longitude, 127.28; latitude, 42.33) of Jilin, China. A single sample was harvested for each growth period. The primary roots were collected, immediately frozen in liquid nitrogen and stored at -80˚C until used for library construction.

Selection of candidate reference genes from ginseng RNA-Seq data. The RNA-Seq data were generated on an Illumina sequencing platform (HiSeq 2,000; Illumina, Inc., San Diego, CA, USA), as described previously (28). Briefly, the samples from the six growing stages were processed according to the manufacturer's protocol and used for transcriptome analysis, including cDNA library construction, sequencing, assembly and gene expression analyses.

Gene expression levels were expressed as RPKM using the following formula: RPKM=109C/NL, where C is the number of mappable reads uniquely aligning to a unigene, N is the total number of mappable reads that uniquely align to all unigenes, and L is the length of a unigene in base pairs. Candidate reference genes were selected by calculating the coefficient of variation (CV) and the maximum fold change (MFC) across multiple samples within each data set, where CV represents the standard deviation (SD) divided by the mean RPKM, and MFC represents the maximum RPKM divided by the minimum RPKM value.

RT-qPCR analysis. Total RNA was isolated using TRIzol reagent (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA) according to the manufacturer's protocol. The RT-qPCR analyses were performed using the One Step SYBR PrimeScript PLUS RT-qPCR kit (Takara Biotechnology, Co., Ltd. Dalian, China, TaKaRa code: DRR096A). The PCR amplification was performed in a 25 μl mixture containing 2.0 μl cDNA, 0.5 μl each primer, 12.5 μl SYBR Premix Ex Taq, 0.5 μl ROX reference dye II and 9 μl distilled water. Data were collected using an ABI Prism 7500 real-time PCR system (Applied Biosystems; Thermo Fisher Scientific, Inc.). The thermal cycling conditions comprised an initial denaturation step at 95˚C for 30 sec and 40 cycles at 95˚C for 5 sec and 65˚C for 34 sec, followed by a dissociation stage at 95˚C for 15 sec, 60˚C for 1 min and 95˚C for 15 sec. All samples were amplified in triplicate and the mean was used for RT-qPCR analysis. Relative gene expression was calculated using the $2^{-\Delta\Delta Cq}$ method (29). The primer sequences were designed using Primer 6.0 software (www.premierbiosoft.com/primerdesign/index.html) (30).

Stability analysis of candidate reference genes. The mRNA expression profiling data sets were prepared and generated from the RNA-Seq data. To compare the stability of the candidate reference genes, the following three Visual Basic for Applications were used for Microsoft Excel: GeNorm (https://genorm.cmgg.be/), NormFinder (http://moma.dk/normfinder-software) and BestKeeper (http://www.gene-quantification.de/bestkeeper.html).

Expression analysis of pathogenesis-related (PR) proteins. To determine whether RT-qPCR normalization with different reference genes altered the expression profiles, PR proteins were used to validate candidate reference genes. The genes and their primers are listed in Table I. The ΔCq values for each sample were calculated using either a traditional reference gene (ACT1) or a novel reference gene (UDP-N-acetylgalactosamine transporter; UDP) or the combination of UDP and nuclear transport factor 2 (NTF2), as identified by geNorm. All these analyses were performed in compliance with Minimum Information for Publication of Quantitative Real-Time PCR Experiments guidelines (31).

Results

Construction of RNA-Seq databases. The RNA-Seq data used in the present study were obtained from our ginseng project, which covered six growing stages between 3 and 20 years. The data included >39,000,000 high-quality sequencing reads for each sample. Following reads clustering, >80,000 unigenes were obtained in each data set, which comprises the gene sequence, gene expression level, annotation and other information for each unigene.

CV and MFC values were used to estimate the stability of the RPKM values in order to select candidate reference genes expressed at moderate or high levels in all six data sets, based on the following three criteria: %CV <25, MFC <5, mean RPKM >100. As a result, 21 candidate reference genes were identified; these comprised eight traditional reference genes: ACT1, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 18s rRNA, ubiquitin (UBQ), tubulin, β-tubulin, cyclophilin (CYP), and PP2Ac-3-phosphatase 2A isoform 3 (PP2A), and 13 non-traditional reference genes: Ubiquitin conjugating enzyme isoform 2 (UBE2), GAGA-binding transcriptional activator (GAGA) protein disulfide isomerase (PDI), mitochondrial-processing peptidase (MPP), glucose-6-phosphate (G6P); UDP, probable prefoldin subunit 5 (PPS); auxin response factor 1 (ARF1), putative 3-isopropylmalate dehydrogenase (3-IPMDH), δ(3,5)-δ(2,4)-dienoyl-CoA isomerase, mitochondrial (ECH1), eukaryotic translation initiation factor 4E-1 (EIF-4E1), SKP1 and NTF2. A summary of the sequence information for these genes is provided in Table II.

Expression profiles of the candidate reference genes. The expression profiles of the 21 candidate reference genes in the

Table I. Primer sequences and amplicon sizes of PR proteins.

Symbol	Gene name	Primer sequence (5'→3')	Amplicon size (bp)
PR1	Pathogenesis-related protein 1	TGTTTCCTTCCTCCCTCG CCCCTTCGCTGATTGGT	145
PR2	Pathogenesis-related protein 2	GCTCCATCCTCAGTCCCA GGTTCCAACTCCACCATCTC	132
PR5	Pathogenesis-related protein 5	CCATTTTCCTTTTCATTTCCA CGTTAATGGTCCAGGTTTGG	147
PR10	Pathogenesis-related protein 10	TTGAAGCACTGGATTGATGAG CCACCATTGGATGATGCC	134

RNA-Seq data sets across the six growth stages were analyzed. The 21 genes were ranked from lowest to highest CV values based on the RPKMs (Fig. 1), which allowed direct comparisons within and between samples with no bias for short genes. The results showed that the non-traditional reference genes, UDP, NTF2 and UBE2, were the most stably expressed genes, whereas traditional reference genes, including ACT1, GAPDH and 18s rRNA, were less consistently expressed in the six growth stages.

Subsequently, the expression profiles of the 21 candidate reference genes were determined using RT-qPCR analysis. The Cq values for individual genes reflect the actual mRNA levels in the samples and can be compared directly. The Cq distribution is shown as a box-plot in Fig. 2. The average Cq values for the 21 genes ranged between 15 and 29 cycles, with the majority falling between 20 and 26 cycles. Consistent with the results from the analysis of the RNA-Seq data, the non-traditional UDP, NTF2, and UBE2 reference genes had more consistent Cq values, compared with the traditional reference genes.

Statistical analysis of RT-qPCR data using geNorm, NormFinder and BestKeeper. The consistency of the expression levels of each reference gene was analyzed using the geNorm, NormFinder and BestKeeper software packages.

geNorm was designed to analyze the expression stability of candidate reference genes based on the assumption that the ratio of the expression levels of two ideal reference genes is constant in all samples. The average expression stability (M value), for each reference gene is calculated using the average of pairwise variations, according to which the expression stability of all reference genes is ranked. The least stable gene, which has the highest M value, is then excluded and the M value is recalculated in a stepwise manner until the two most stably expressed genes are identified (32). The geNorm analyses of all six samples revealed that the UDP and NTF2 combination had the lowest M value (0.18), whereas GAGA had the highest M value (0.58; Fig. 3). geNorm also calculates the pairwise variation (Vn/Vn+1) between two sequential normalization factors, NFn and NFn+1, to determine the optimum number of reference genes. As a general rule, the stepwise inclusion of reference genes is performed until Vn/Vn+1 falls below a theoretical threshold of 0.15, when the benefit of adding another gene (n+1) is limited (25,33). In the present study, the pairwise variation V2/3 was below the default cut-off value of 0.15, which indicated that the inclusion of a third reference gene was not necessary. Thus, UDP/NTF2 may be the most suitable combination of reference genes for gene expression analyses in ginseng at different growth stages (Fig. 4).

The present study also used NormFinder to rank the expression stability of the 21 candidate reference genes. NormFinder uses an analysis of variance-based model to estimate intra- and inter-group variations, and combines these estimates to provide a direct measure of the variations in expression for each gene (26). Genes with lower average expression stability values are more stably expressed. NormFinder analyses of all six samples revealed that UDP was the most stable gene, followed by ECH1, which surpassed that of UBE2, whereas GAGA was the least stable (Table III).

The BestKeeper program analyzes the stability of a candidate reference gene based on the CV and standard deviation (SD) of Cq values using the average Cq value of each duplicate reaction (27). Reference genes, which exhibit the lowest CV±SD, are determined as the most stable genes. The BestKeeper analyses revealed that UBE2 and UDP showed the highest expression stability in all six samples (Table IV), whereas GAGA and tubulin showed the least stable expression. Although the preferred reference genes differed marginally for each program, UDP consistently ranked high in expression stability.

Consensus list of candidate reference genes. To provide a consensus result from the outputs of the three statistical programs, an arithmetic mean ranking value was calculated for each gene to obtain the final gene stability ranking order (Table V). The results revealed that UDP, NTF2 and UBE2 were the most stable reference genes, whereas tubulin, CYP and GAGA were the least stable.

Validation of the usefulness of the selected reference genes. Validation of the sets of candidate reference genes involved normalizing the RT-qPCR expression levels of the genes encoding four PR proteins (PR1, PR2, PR5 and PR10) in the six growing stages. To survive under different environmental stresses, ginseng has developed mechanisms to perceive external signals, which trigger adaptive responses and appropriate physiological alterations, with the induction of PR proteins being one such response (34). PR proteins have been

Table II. *Panax ginseng* candidate reference genes, primers and amplicon sizes.

Symbol	Gene name	Primer sequence (5'→3')	Amplicon size (bp)
UBE2	Ubiquitin-conjugating enzyme isoform 2	AGTGCTGGACCTGTTGGTGAAG CTGGTGGGAAATGAATGGATAC	112
GAGA	GAGA-binding transcriptional activator BBR/BPC	AATGAGTAGCGGGGTTGATGAC CCTCCATTTCCCCATTTGTAGC	132
PDI	Protein disulfide isomerase	GCAGACAAAGATAGCCCATTCC AAGGCAACAAAGCAGATGGCAG	173
MPP	Mitochondrial-processing peptidase	CGACCTAAGGAACCACAATCAG CTTCCTTCACATTATGCCAGCC	121
G6P	Glucose-6-phosphate	TGAAGGGGAAGTCTGTTAGTGG TTCCATCCAAGTGCCCACATCT	121
UDP	UDP-N-acetylgalactosamine dual transporter	CGGCAAGCAGAGATAAGACACT CGGCAAGCAGAGATAAGACACT	95
PPS	Probable prefoldin subunit 5	AGCAGTAAAGGAACAAACCGAT ACATAAAGCGACGCCGTAAGAG	159
ARF1	Auxin response factor 1	GAGCGTGGAGAAAAAGGTATTG GCTTCAACTGATAAATGCGACC	142
3-IPMDH	Putative 3-isopropylmalate dehydrogenase	TCCCGCTATCTTCGTGTCTTCT GGATAGGTTGGGAAATGAAGGT	105
ECH1	δ(3,5)-δ(2,4)-dienoyl-CoA isomerase, mitochondrial	AATCTCTTCCTCAATCGCCCAT ATTAGGGTTTTGGTCAAGGGAG	130
EIF-4E1	Eukaryotic translation initiation factor 4E-1	TATTCCACATCCACTTGAGCAC GAAGAGAAAGTGTAGATGGGGC	111
SKP1	Skp1	CGCTAACACCAGTATTCCCCTT GATGTTGAGGTAGTTTGCTGCC	214
NTF2	Nuclear transport factor 2	AGAACATCGTTGCCAAACTCAC CTGACAAAGACGAGCATACCAC	112
ACT1	Actin 1	TGGCATCACACTTTCTACAACG TTTGTGTCATCTTCTCCCTGTT	109
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic	GAGAAGGAATACACACCTGACC CAGTAGTCATAAGCCCCTCAAC	106
18s rRNA	18S rRNA	TTCACACCAAGTATCGCATTTC CCAAGGAAATCAAACTGAACTG	145
UBQ	Ubiquitin, putative	AACCAACTGATACCATTGACCG CTTTTGCTGTTTTGTCATCTCC	120
Tublin	Tubulin α-1 chain	CTCTGTTGTTGGAACGCTTGTC CTGTGTGCTCAAGAAGGGAATG	144
β-Tublin	β-tubulin	TGTTGTGAGGAAAGAAGCCGAG GGAGAAGGGAAGACAGAGAAAG	165
CYP	Cyclophilin	CAGGCAAAGAAAAAGTCAAGTG AAAGAGACCCATTACAATACGC	108
PP2A	PP2Ac-3-phosphatase 2A isoform 3	GCTCCAAACTACTGTTACCGCT ATAATCAGGTGTCTTGCGGGTG	141

classified into 17 families on the basis of structural differences, serological associations and biological activity (35).

In accordance with the results obtained from the RNA-Seq data sets, the alterations in gene expression levels of the PR proteins showed similar patterns when the UDP/NTF2 combination (from geNorm) and the most consistent reference gene, UDP, were used for normalization. However, significantly different gene expression levels were observed for the PR proteins when the traditional reference gene, ACT1, was used for normalization. Spearman's correlation analysis also demonstrated a high degree of correlation between RPKM and relative quantification when the UDP/NTF2 combination or UDP were used as reference genes, and a low degree of correlation when ACT1 was used (Fig. 5).

These results showed that the choice of reference gene had a considerable effect on the normalization results, and that

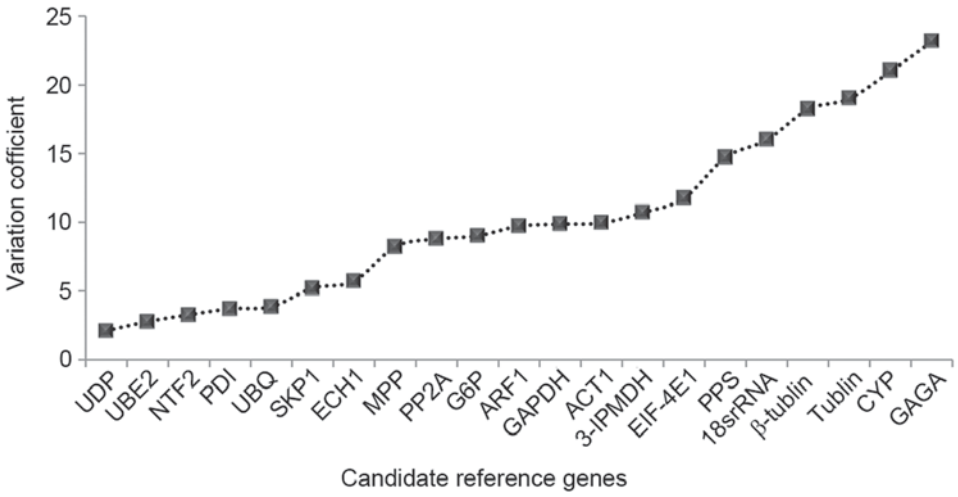

Figure 1. Variation coefficient of reads per kilobase of transcript per million. The horizontal axis lists the candidate reference genes; the vertical axis is the variation coefficient. The lower the variable coefficient, the higher the stability.

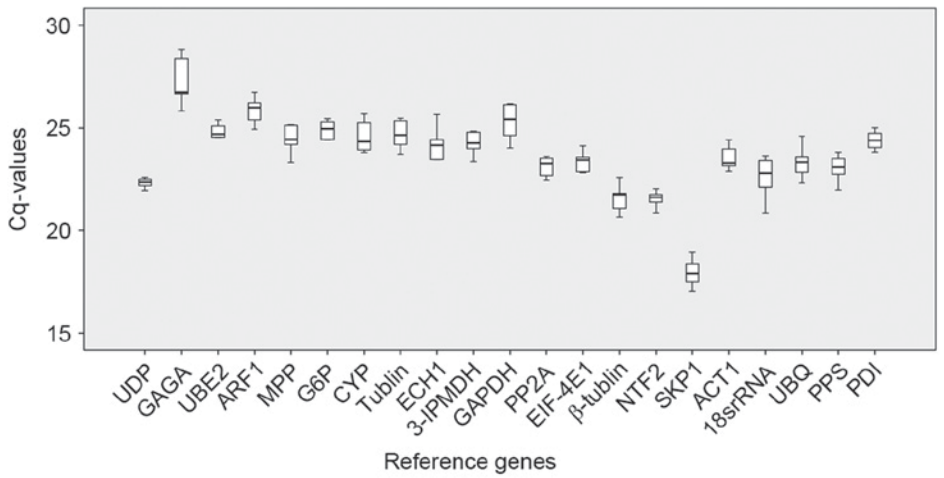

Figure 2. Box and whisker plots of Cq values for the 21 candidate reference genes in six growth stages. Expression data is shown as the Cq value for each reference gene in all ginseng samples. The data are expressed as whisker box plots; the box represents the 25th-75th percentiles. The median is indicated by the bar across the box, and the whiskers on each box represent the minimum and maximum values.

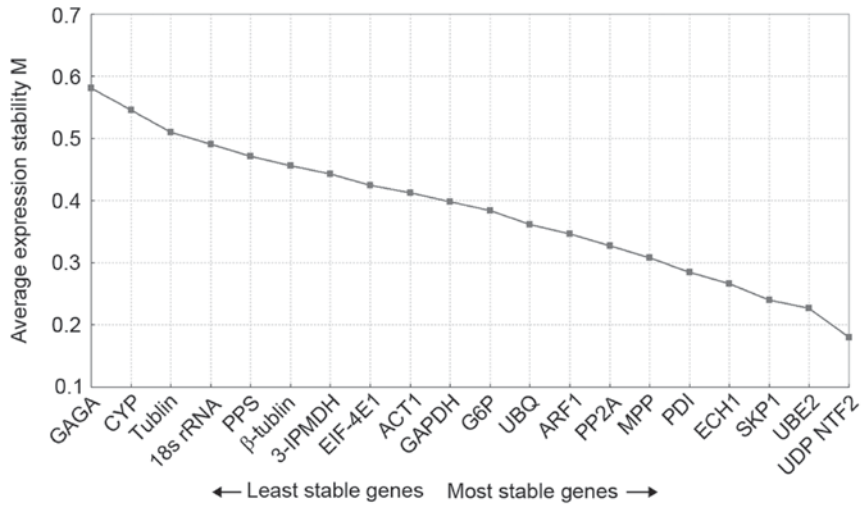

Figure 3. Ranking of candidate reference genes by geNorm. The expression stability of the candidate reference genes was evaluated using geNorm. The M value was determined by assessing the mean pairwise variations of all genes; the least stable gene, with the highest M value, was excluded, and the M value was recalculated until the most stable pair was selected. M, stability.

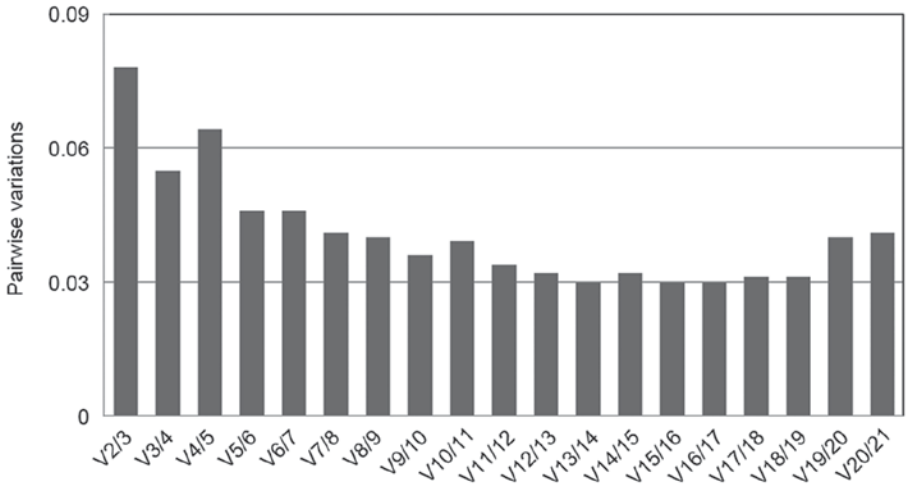

Figure 4. Determination of the optimal number of reference genes. The geNorm program calculated an NF and used the variable, V, to determine pairwise variation (Vn/Vn+1) between two sequential NFs (NFn and NFn+1). Additional genes were included when V exceeded the cut off value, which is typically set at 0.15, but is not always achievable. The number of reference genes is considered optimal when the lowest possible V value is achieved, at which point it is unnecessary to include additional genes in the normalization strategy. NF, normalization factor; V, variation.

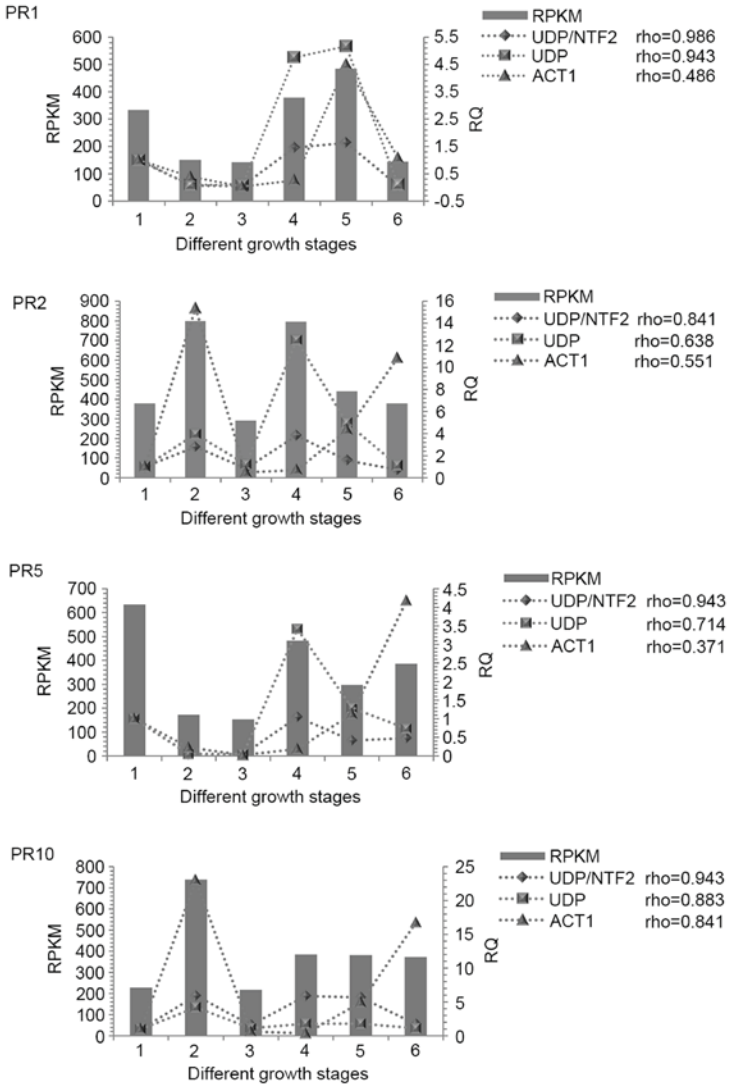

Figure 5. Expression profiles of PR proteins. The RQ of PR proteins were determined using reverse transcription-quantitative polymerase chain reaction analysis, calculated using the $2^{-\Delta\Delta Cq}$ method with either ACT1 or UDP and NTF2 as the reference gene, relative to the 3-year-old ginseng sample. For UDP and NTF2, the geometric mean was calculated and used for normalization. RQ, relative quantification; RPKM, reads per kilobase of transcript per million mapped reads obtained for each ginseng sample by RNA-seq; PR, pathogenesis-related; ACT1, actin 1; UDP, UDP-N-acetylgalactosamine transporter; NTF2, nuclear transport factor 2.

Table III. Ranking of candidate reference genes using NormFinder.

Rank	Tissue (Stability value)
1	UDP (0.101)
2	ECH1 (0.105)
3	UBE2 (0.139)
4	NTF2 (0.143)
5	ARF1 (0.145)
6	MPP (0.157)
7	SKP1 (0.188)
8	PP2A (0.223)
9	EIF-4E1 (0.229)
10	PDI (0.244)
11	G6P (0.253)
12	GAPDH (0.272)
13	UBQ (0.276)
14	ACT1 (0.280)
15	PPS (0.309)
16	3-IPMDH (0.334)
17	β-tublin (0.348)
18	Tublin (0.391)
19	18srRNA (0.420)
20	CYP (0.546)
21	GAGA (0.591)

Table IV. Ranking of candidate reference genes using BestKeeper.

Rank	Tissues (CV%±SD)
1	UBE2 (0.87±0.19)
2	UDP (0.93±0.20)
3	NTF2 (1.08±0.26)
4	PDI (1.11±0.28)
5	PP2A (1.61±0.39)
6	SKP1 (1.65±0.38)
7	ECH1 (1.72±0.42)
8	MPP (1.85±0.48)
9	G6P (1.92±0.47)
10	UBQ (1.95±0.46)
11	GAPDH (2.06±0.44)
12	β-tublin (2.24±0.52)
13	ACT1 (2.28±0.49)
14	3-IPMDH (2.44±0.59)
15	PPS (2.49±0.61)
16	ARF1 (2.53±0.62)
17	EIF-4EI (2.96±0.75)
18	CYP (2.97±0.68)
19	18s rRNA (3.00±0.54)
20	GAGA (3.46±0.94)
21	Tublin (3.91±0.88)

CV, coefficient of variation.

using inappropriate reference genes may introduce bias to the analysis and cause misleading results.

Discussion

Using inaccurate reference genes for normalization can lead to conflicting results in gene expression investigations based on RT-qPCR analysis, particularly when transcription rate variations between sample groups are small (36). Increasing evidence indicates that traditional reference genes do not show stable expression under all conditions (27,37). Therefore, it is important to validate the expression stability of a reference gene under specific experimental conditions prior to use in RT-qPCR normalization.

The present study performed systematic analysis of the stability of mRNA expression levels of 21 candidate reference genes, including eight traditional reference genes from six ginseng transcriptome data sets. RNA-Seq data and three independent methods, geNorm, NormFinder and BestKeeper, were used to identify suitable reference genes for differential gene expression analyses during ginseng growth years. Among the 21 candidate reference genes analyzed, UDP and NTF2 were determined to be the optimal combination of reference genes for analyzing expression.

Microarray and large-scale sequencing technologies have been used to identify stably expressed reference genes (38). RNA-Seq technology is considered a method technology for the following reasons: i) RNA-Seq reads are digital rather than analog; ii) there is low background signal; and iii) there is virtually no upper limit for detection results in a substantially larger dynamic range (20,21,39-42). A higher degree of technical reproducibility with RNA-Seq, compared with microarrays has been reported, and RNA-Seq expression data correlate well with RT-qPCR data, regardless of the sequencing platform used (39,41).

As a single software package may introduce bias, three statistical approaches, geNorm, NormFinder and BestKeeper, were used in the present study to determine the stability of the expression of the 21 candidate reference genes. GeNorm and BestKeeper identified UDP, NTF2 and UBE2 as the genes with the least variation, and NormFinder identified UDP, ECH1 and UBE2 as the genes with the highest expression stabilities. Inconsistencies between the three methods can be expected, as they use different statistical algorithms (43). To summarize the results, a comprehensive ranking order of each reference gene was calculated, and it was found that the reference gene with the highest stability was the combination of UDP and NTF2.

UDP, a novel nucleotide sugar transporter with dual substrate specificity, is important in the development of plants, and may also be involved in glucuronidation and chondroitin sulfate biosynthesis (44). NTF2 is indispensable in plants, as it facilitates protein transport into the nucleus. It may be a component of a multicomponent system of cytosolic factors, which assemble at the pore complex during nuclear import (45,46). These two reference genes exhibited similar expression patterns in the six growth stages of ginseng, possibly due to them being involved in basic cell metabolism and cellular functions. In addition to their

Table V. Comprehensive ranking order.

Ranking order	Genorm	Normfinder	Bestkeeper	Comprehensive ranking (mean)
1	UDP	UDP	UBE2	UDP (1.67)
2	NTF2	ECH1	UDP	NTF2 (3.67)
3	UBE2	UBE2	NTF2	UBE2 (3.83)
4	SKP1	NTF2	PDI	PDI (4.67)
5	ECH1	ARF1	PP2A	ECH1 (5.33)
6	PDI	MPP	SKP1	SKP1 (5.57)
7	MPP	SKP1	ECH1	MPP (6.23)
8	PP2A	PP2A	MPP	PP2A (6.58)
9	ARF1	EIF-4E1	G6P	ARF1 (9.17)
10	UBQ	PDI	UBQ	UBQ (10.31)
11	G6P	G6P	GAPDH	G6P (10.99)
12	GAPDH	GAPDH	β-tublin	GAPDH (11.65)
13	ACT1	UBQ	ACT1	ACT1 (12.98)
14	EIF-4E1	ACT1	3-IPMDH	EIF-4E1 (13.31)
15	3-IPMDH	PPS	PPS	3-IPMDH (14.98)
16	β-tublin	3-IPMDH	ARF1	β-tublin (15.00)
17	PPS	β-tublin	EIF-4E1	PPS (15.67)
18	18s rRNA	Tublin	CYP	18s rRNA (18.67)
19	Tublin	18sr RNA	18sr RNA	Tublin (19.33)
20	CYP	CYP	GAGA	CYP (19.34)
21	GAGA	GAGA	Tublin	GAGA (20.67)

high expression stability, the superiority of UDP and NTF2 over the traditional reference genes was based on their lower expression levels. The use of reference genes with low expression levels similar to the target genes has been recommended in order for the comparisons to fall on the same linear scale (47). The data obtained in the present study supported the unsuitability of the traditional reference genes, including ACT1, for normalization, which was in accordance with other studies (48,49). The reference genes selected in the present study may be superior reference genes for the normalization of a wide range of genes, particularly weakly expressed genes. This result is significant as the majority of transcripts in tissues are expressed at low levels (50).

The results of the present study revealed that the expression levels normalized by a single top-ranked reference gene were less accurate, compared with expression levels normalized using two reference genes. Therefore, for investigations of ginseng development and growth, it is recommended that two reference genes are used for reliable quantification.

In conclusion, the present study used RNA-Seq data to identify 21 candidate reference genes in ginseng root grown for different durations, and identified UDP and NTF2 as the most suitable reference genes using geNorm, NormFinder and BestKeeper. These genes were validated using RT-qPCR analysis for use as reference genes in ginseng investigations. The results showed that the use of unsuitable reference genes for normalization may result in biased expression levels. These findings are useful for further gene expression analyses of ginseng growth, particular associated with marker identification, environmental stress and the characterization of gene function. In addition, the results of the present study provide useful guidelines for reference gene selection in investigations of other species.

Acknowledgements

This study was supported by grants from the National Natural Foundation of China (grant nos. 81373937 and 81503212), the Scientific and Technological Development Program of Jilin, China (grant nos. 20140622003JC and 20150520139JH) and the Strategic Adjustment of the Economic Structure of Jilin Province to Guide the Capital Projects (grant no. 2014N155).

References

1. Xiang YZ, Shang HC, Gao XM and Zhang BL: A comparison of the ancient use of ginseng in traditional chinese medicine with modern pharmacological experiments and clinical trials. Phytother Res 22: 851-858, 2008.
2. James AD: The Green Pharmacy Herbal Handbook: Your comprehensive reference to the best herbs for healing. Rodale: Emmaus, USA, pp115-116, 2000.
3. Dan M, Xie G, Gao X, Long X, Su M, Zhao A, Zhao T, Zhou M, Qiu Y and Jia W: A rapid ultra-performance liquid chromatography-electrospray Ionisation mass spectrometric method for the analysis of saponins in the adventitious roots of Panax notoginseng. Phytochem Anal 20: 68-76, 2009.
4. Shan SM, Luo JG, Huang F and Kong LY: Chemical characteristics combined with bioactivity for comprehensive evaluation of Panax ginseng C.A. Meyer in different ages and seasons based on HPLC-DAD and chemometric methods. J Pharm Biomed Anal 89: 76-82, 2014.
5. Wan JY, Fan Y, Yu QT, Ge YZ, Yan CP, Alolga RN, Li P, Ma ZH and Qi LW: Integrated evaluation of malonyl ginsenosides, amino acids and polysaccharides in fresn and processed ginseng. J Pharm Biomed Anal 107: 89-97, 2015.

6. He JM, Zhang YZ, Luo JP, Zhang WJ and Mu Q: Variation of ginsenoside in ginseng of different ages. Nat Prod Commun 11: 739-740, 2016.
7. Sathiyaraj G, Srinivasan S, Subramanium S, Kim YJ, Kim YJ, Kwon WS and Yang DC: Polygalacturonase inhibiting protein: Isolation, developmental regulation and pathogen related expression in Panax ginseng C.A. Meyer. Mol Biol Rep 37: 3445-3454, 2010.
8. Han JY, Kim HJ, Kwon YS and Choi YE: The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II duringginsenoside biosynthesis in Panax ginseng. Plant Cell Physiol 52: 2062-2073, 2011.
9. Bustin SA: Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. J Mol Endocrinol 29: 23-39, 2002.
10. Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A and Heinen E: Housekeeping genes as internal standards: Use and limits. J Biotechnol 75: 291-295, 1999.
11. Bustin SA: Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25: 169-193, 2000.
12. Schmittgen TD and Zakrajsek BA: Effect of experimental treatment on housekeeping gene expression: Validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods 46: 69-81, 2000.
13. Lee PD, Sladek R, Greenwood CM and Hudson TJ: Control genes and variability: Absence of ubiquitous reference transcripts in diverse mammalian expression studies. Genome Res 12: 292-297, 2002.
14. Huang Z, Lin J, Cheng Z, Xu M, Guo M, Huang X, Yang Z and Zheng J: Production of oleanane-type sapogenin in transgenic rice via expression of β-amyrin synthase gene from Panax japonicas C. A. Mey. BMC Biotechnol 15: 45, 2015.
15. Lim W, Shim MK, Kim S and Lee Y: Red ginseng represses hypoxia-induced cyclooxygenase-2 through sirtuin1 activation. Phytomedicine 22: 597-604, 2015.
16. Oh GS, Yoon J, Lee GG, Oh WK and Kim SW: 20 (S)-protopanaxatriol inhibits liver X receptor α-mediated expression of lipogenic genes in hepatocytes. J Pharmacol Sci 128: 71-77, 2015.
17. Qi J, Sun P, Liao D, Sun T, Zhu J and Li X: Transcriptomic analysis of american ginseng seeds during the dormancy release process by RNA-Seq. PLoS One 10: e0118558, 2015.
18. Zhu L, Li J, Xing N, Han D, Kuang H and Ge P: American ginseng regulates gene expression to protect against premature ovarian failure in rats. Biomed Res Int 2015: 767124, 2015.
19. Liu J, Wang Q, Sun M, Zhu L, Yang M and Zhao Y: Selection of reference genes for quantitative real-time PCR normalization in Panax ginseng at different stages of growth and in different organs. PLoS One 9: e112177, 2014.
20. Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621-628, 2008.
21. Wang Z, Gerstein M and Snyder M: RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet 10: 57-63, 2009.
22. Macrae T, Sargeant T, Lemieux S, Hébert J, Deneault E and Sauvageau G: RNA-Seq reveals spliceosome and proteasome genes as most consistent transcripts in human cancer cells. PLoS One 8: e72884, 2013.
23. Cankorur-Cetinkaya A, Dereli E, Eraslan S, Karabekmez E, Dikicioglu D and Kirdar B: A novel strategy for selection and validation of reference genes in dynamic multidimensional experimental design in yeast. PLoS One 7: e38351, 2012.
24. de Jonge HJ, Fehrmann RS, de Bont ES, Hofstra RM, Gerbens F, Kamps WA, de Vries EG, van der Zee AG, te Meerman GJ and ter Elst A: Evidence based selection of housekeeping genes. PLoS One 2: e898, 2007.
25. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A and Speleman F: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: research0034, 2002.
26. Andersen CL, Jensen JL and Ørntoft TF: Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64: 5245-5250, 2004.
27. Pfaffl MW, Tichopad A, Prgomet C and Neuvians TP: Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol Lett 26: 509-515, 2004.
28. Yao B, Zhao Y, Zhang H, Zhang M, Liu M, Liu H and Li J: Sequencing and de novo analysis of the Chinese Sika deer antler-tip transcriptome during the ossification stage using Illumina RNA-Seq technology. Biotechnol Lett 34: 813-822, 2012.
29. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402-408, 2001.
30. Rozen S and Skaletsky H: Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386, 2000.
31. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, *et al*: The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55: 611-622, 2009.
32. Pombo-Suarez M, Calaza M, Gomez-Reino JJ and Gonzalez A: Reference genes for normalization of gene expression studies in human osteoarthritic articular cartilage. BMC Mol Biol 9: 17, 2008.
33. Warzybok A and Migocka M: Reliable reference genes for normalization of gene expression in cucumber grown under different nitrogen nutrition. PLoS One 8: e72887, 2013.
34. Gibbs GM, Roelants K and O'Bryan MK: The CAP superfamily: Cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins-roles in reproduction, cancer and immune defense. Endocr Rev 29: 865-897, 2008.
35. Sels J, Mathys J, De Coninck BM, Cammue BP and De Bolle MF: Plant pathogenesis-related (PR) proteins: A focus on PR peptides. Plant Physiol Biochem 46: 941-950, 2008.
36. Etschmann B, Wilcken B, Stoevesand K, von der Schulenburg A and Sterner-Kock A: Selection of reference genes for quantitative real-time PCR analysis in canine mammary tumors using the GeNorm algorithm. Vet Pathol 43: 934-942, 2006.
37. Huis R, Hawkins S and Neutelings G: Selection of reference genes for quantitative gene expression normalization in flax (Linum usitatissimum L.). BMC Plant Biol 10: 71, 2010.
38. Hamalainen HK, Tubman JC, Vikman S, Kyrölä T, Ylikoski E, Warrington JA and Lahesmaa R: Identification and validation of endogenous reference genes for expression profiling of T helper cell differentiation by quantitative real-time RT-PCR. Anal Biochem 299: 63-70, 2001.
39. Oshlack A, Robinson MD and Young MD: From RNA-seq reads to differential expression results. Genome Biol 11: 220, 2010.
40. Mane SP, Evans C, Cooper KL, Crasta OR, Folkerts O, Hutchison SK, Harkins TT, Thierry-Mieg D, Thierry-Mieg J and Jensen RV: Transcriptome sequencing of the microarray quality control (MAQC) RNA reference samples using next generation sequencing. BMC Genomics 10: 264, 2009.
41. Wilhelm BT and Landry JR: RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48: 249-257, 2009.
42. Shendure J: The beginning of the end for microarrays? Nat Methods 5: 585-587, 2008.
43. Jian B, Liu B, Bi Y, Hou W, Wu C and Han T: Validation of internal control for gene expression study in soybean by quantitative real-time PCR. BMC Mol Biol 9: 59, 2008.
44. Muraoka M, Kawakita M and Ishida N: Molecular characterization of human UDP-glucuronic acid/UDP-N-acetylgalactosamine transporter, a novel nucleotide sugar transporter with dual substrate specificity. FEBS Lett 495: 87-93, 2001.
45. Bayliss R, Ribbeck K, Akin D, Kent HM, Feldherr CM, Görlich D and Stewart M: Interaction between NTF2 and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP. J Mol Biol 293: 579-593, 1999.
46. Goepfert S, Vidoudez C, Rezzonico E, Hiltunen JK and Poirier Y: Molecular identification and characterization of the Arabidopsis delta(3,5), delta(2,4)-dienoyl-coenzyme A isomerase, a peroxisomal enzyme participating in the beta-oxidation cycle of unsaturated fatty acids. Plant Physiol 138: 1947-1956, 2005.
47. Czechowski T, Stitt M, Altmann T, Udvardi MK and Scheible WR: Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol 139: 5-17, 2005.
48. Goidin D, Mamessier A, Staquet MJ, Schmitt D and Berthier-Vergnes O: Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. Anal Biochem 295: 17-21, 2001.
49. Selvey S, Thompson EW, Matthaei K, Lea RA, Irving MG and Griffiths LR: Beta-actin: An unsuitable internal control for RT-PCR. Mol Cell Probes 15: 307-311, 2001.
50. Warrington JA, Nair A, Mahadevappa M and Tsyganskaya M: Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol Genomics 2: 143-147, 2000.