E3 Journal of Biotechnology and Pharmaceutical Research Vol.2(1). pp. 009-018, September, 2011
Available online http://www.e3journals.org/JBME
ISSN 2141-7474 © 2011 E3 Journals

Full length research paper

Identification and validation of stable reference genes in *Camellia* Species

Bornali Gohain*, Tirthankar Bandyopadhyay, Sangeeta Borchetia, Raju Bharalee, Sushmita Gupta, Priyadarshini Bhorali , Niraj Agarwala and Sudripta Das

Department of Biotechnology, Tea Research Association, Tocklai, Jorhat-785008, Assam, India

Accepted 12 October 2011

We aimed at finding and validating a stable reference gene in *Camellia sinensis* and *Camellia assamica* from a set of four putative housekeeping genes in various samples exposed to different experimental conditions mainly biotic and abiotic stresses. Variation in gene expression across *Camellia sinensis* leaf tissues exposed to nine different kind of experimental sets was studied. The suitability of *18S rRNA, 26s rRNA*, rubisco bis phosphatase (*RuBP*) and camellia actin (*Act)* as reference genes were validated by geNorm and BestKeeper programs and revealed 18S rRNA and RuBP to be the most stably expressing housekeeping gene. We therefore recommend use of RuBP as a stable reference gene for normalisation of transcripts abundance experiments in tea leaf samples.

Keywords: Gene expression; Housekeeping gene; Reference gene.

INTRODUCTION

Till recently gene expression regulation and turnover of gene transcripts (mRNA), and the number of copies of an mRNA transcript of a gene in a cell or tissue was determined by the rates of its expression and degradation using Northern blotting. Purified RNA after agarose gel electrophoresis is transferred to a solid matrix (such as a nylon membrane), and probed with a specific DNA or RNA probe which are complementary to the gene of interest. This method requires relatively large amounts of RNA and provides only qualitative or semi-quantitative information of mRNA levels. For robust detection and quantification of gene expression from very small amount of RNA increasing the amount of transcript prior to detection is necessary. The polymerase chain reaction (PCR) is the widely used method of amplifying DNA, cDNA (reverse transcribed RNA). With the advent of PCR based techniques and fluorescence based detection method, the PCR amplification could be estimated in real time. The abundance of transcripts at each of the PCR cycle and the data acquired can be further analysed with bioinformatics aids to calculate the relative expression across several tissues, different experimental types etc. This requires selection of appropriate reference genes for normalisation. We estimated the gene expression stability of four candidate house-keeping genes viz. *18S, 26S, RuBP* and *Camellia* tubulin in 10 experimental types across leaf tissues of *Camellia sinensis*. Mainly two softwares popularly used in different crops like *Coffea arabica*, chicory, and peach were used in this experiment viz. geNorm and BestKeeper (Cavallari et al., 2009,). BestKeeper is based on a descriptive statistical method.

The approach of accurate data normalization is the most essential component for measurement and analysis of gene expression data (Andersen etal*.,* 2004). Expression of the reference gene used to normalize qPCR analyses should be unaffected throughout many biological contexts; otherwise, it may lead to erroneous results (Hugget et al., 2005; Bustin and Nolan 2002).

Few genes like β-actin, rRNA, β-tubulin, alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, 14-3-3 and poly-ubiquitin have been used as internal controls for gene expression analyses, they are assumed to possess stable expression. Many reports till date have shown that the expression of putative housekeeping genes is different in different tissue and organ types (Cavallari etal., 2009). These genes therefore cannot be used as internal control/housekeeping genes in real time qPCR assays as

*Corresponding Author Email: bornalig6@gmail.com; bornaliz@yahoo.co.in

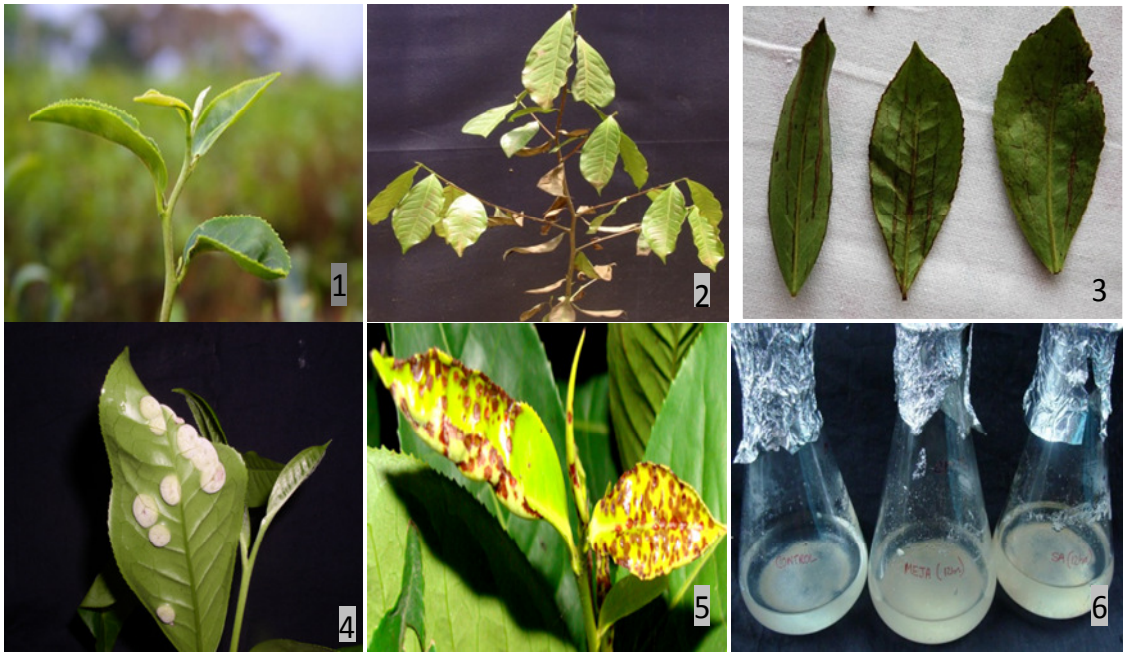

Figure 1. Different leaf samples considered for the study depicting the severity of the stress, both abiotic and biotic are 1. Greenfly infested leaf 2. Drought induced leaf samples 3. Thrips infested leaf samples 4. Blister blight infected tea leaf 5. Helopeltis infested tea leaf 6. MeJA and SA elicited cell suspension cultures

the variation in expression between sample types (tissues and organs) can cause significant experimental errors leading to inaccurate data interpretation. In recent years, it has become clear that no single gene is constitutively expressed in all cell types and under all experimental conditions, implying that the expression stability of the intended control gene has to be verified before each experiment (Andersen etal., 2004).

Tea, *Camellia sinensis*, L (O. Kuntze), belongs to the family Theaceae, and found in the natural environment of South-East Asia. It is a cross-breeding plantation crop (Lupton, 1984). There are mainly three varieties of the plant: Assam*; C. assamica* (Masters Wight), China; *C. sinensis*, L (O. Kuntze) and Cambod; *C. assamica* Sp. *lasiocalyx* (Planchom ex Watt Wight) tea (Hara etal., 1995). Crop productivity or crop yield is mostly an agronomic concept which is controlled by a large number of genes inherited in a quantitative manner. There is a need for information on the nature of molecular patterns underlying behaviour of tea clones during conditions considered to be of economic importance to tea cultivation and obtaining gene expression profiles is an important exercise. Establishing a stable and valid internal control gene is therefore fundamental.

The study undertaken involved a range of experimental conditions encompassing different clones, treatments and time course experiments in the design viz. *Camellia sinensis, C.assamica*, garden series clones, tocklai vegetative clones, tea suspension cultures, greenfly, thrips and *Helopeltis* infested tea leaves, blister blight infected leaf samples and drought induced samples of tea (See figure 1).

The climatic condition which prevails in Darjeeling is in complete contrast to that which prevails in Assam. Thus, considering tea clones grown under these two contrasting climatic conditions will give us an idea of genes which expresses constitutively under both the condition i.e., the variable factor in this experiment was the climate and not the clone. *Exobasidium vexans* (causal organism of blister blight disease in tea) causes great loss to Darjeeling clones and therefore studying its influence on the gene expression profiles of the host is worthwhile. Herbivory causes loss to tea production in Assam. The incidence of tea mosquito bug or *Helopeltis theivora* on tea is profound in Assam tea plantations causing huge losses to tea production in Assam (Muraleedharan etal.,1988). We therefore tried to profile the gene expression changes owing to the interaction between this pest and few important tea clones. Darjeeling tea gardens, on the contrary, find infestation by thrips and greenfly of economic importance as studies have shown that these insects enhance the emission of volatile biochemicals and eventually increasing the worth of Darjeeling tea for aroma and flavour attributes. These groups of insect vis-a-vis tea thus require to be studied and gene expression changes due to herbivory need to be documented and analysed. Apart from these there are certain abiotic stresses prevalent in certain tea growing

belts in the eastern region of the country like drought. Molecular implication of this stress on Assam tea clones have not been documented as yet Therefore it will be interesting to derive the gene expression profiles for such stresses in tea. We also performed qPCR gene expression assays to study the transcriptional cascade in tea suspension cell culture after MeJA and salicyclic acid treatments to better understand the accumulation of several classes of alkaloids, phenolics or coumarins and the defense mechanism they exhibit in response to stress.

Considering all the above molecular needs it is therefore necessary to conduct an experiment which would help determine the best suited housekeeping genes and thus obtain normalised expression changes of target genes in such samples.

We employed two software programmes to calculate the most stable gene for use as housekeeping gene. They were namely geNorm (Vandesompele et al., 2002*)* and BestKeeper (Pfaffl 2003). They were used based on their popularity and usability. The crossing point values for the samples under different experimental setup were imported into excel for analysis in BestKeeper. In geNorm, the crossing point values were first converted into relative quantities using the delta Ct method before analysis. Both the softwares required the input of PCR efficiency for obtaining the normalised estimates for expression in housekeeping genes in the given sample size.

The BestKeeper software first compute a BestKeeper index which is based on the best combination of housekeeping genes from a group of maximum 10 genes. The best suited ones are based on the BestKeeper index computed by the software, which are then compared with the GOI (gene of interest/ target genes) to countercheck their differential expression subject to the experimental conditions and factors applied. For internal standardisation of target gene expression data it is therefore required to determine the stability of standard gene expression or the housekeeping gene expression, based on the stable expression of these gene, the target gene expression is normalised. Best keeper software facilitates descriptive analysis of reference genes/ internal controls.

The geNorm VBA applet for Microsoft Excel determines the most stable reference genes from a set of tested genes in a given cDNA sample panel, and calculates a gene expression normalization factor for each tissue sample based on the geometric mean of a user defined number of reference genes (Vandesompele et al., 2002). GeNorm calculates the gene expression stability measure M for a reference gene as the average pair wise variation V for that gene with all other tested reference genes. Stepwise exclusion of the gene with the highest M value allows ranking of the tested genes according to their expression stability.

GeNorm was also used to estimate the normalisation factor (NFn) using n multiple reference genes, by calculating the geometric mean of the expression levels of the n best reference genes (Cavallari et al., 2009). The optimisation of the number of reference genes starts with the inclusion of the two genes with the lowest M value, and continues by sequentially adding genes with increasing values of M. Thus, geNorm calculates the pair-wise variation Vn/ Vn+1 between two sequential normalisation factors NFn and NFn+1 containing an increasing number of reference genes. A large variation means that the added gene has a significant effect on the normalisation and should preferably be included for calculation of a reliable normalisation factor. Ideally, extra reference genes are included until the variation Vn/Vn+1 drop below a given threshold (Vandesompele et al., 2002). However, the group also mentioned that this threshold should not be taken as a strict cut-off value.

MATERIALS AND METHODS

Experimental condition and tissue types

Within tissue types, variation in expression of housekeeping gene was generally less in leaf tissue samples from the same experimental condition, whereas the expression variation can be substantial while considering different tissue types. Cavallari et al., (2009) mentioned that normalisation to a single gene across different tissue types is unwise, since the variation observed between normal tissues of different types may in part be due to the different metabolic demands of those tissues. Our experiments included leaf tissues subjected to various biotic and abiotic stresses, and the housekeeping genes considered expressed constitutively in leaf tissue types.

The rationale behind considering the 4 genes was based on available data on use of housekeeping genes for qPCR analysis in tea. Sakata et al., (2007) used 26S rRNA gene for carrying out qPCR analysis in tea, they found it to be stably expressed in their analysis. We selected the rRNA genes viz. 18S rRNA and 26S rRNA as they demonstrate somewhat constitutive expression across samples as observed in semi quantitative PCR (figure 2). Kumar etal. 2009 also used 26S rRNA genes as internal control for qPCR assays in *Camellia sinensis*, however, our analysis showed that 18S rRNA and RuBP genes expressed more stably in *Camellia sinensis*. Maroufi et al., (2010) used actin and rRNA genes for validation qPCR data in chicory and found them to be expressed stably across different chicory samples. A myriad of housekeeping genes have been used and successfully evaluated for normalising real time quantitative PCR data in different organisms and few of them have been evaluated for plant species like rice, poplar, coffee, chicory, Soya bean and *Arabidopsis thaliana* (Kim et al., 2003; Brunner et al., 2004; Volkov et al., 2003).

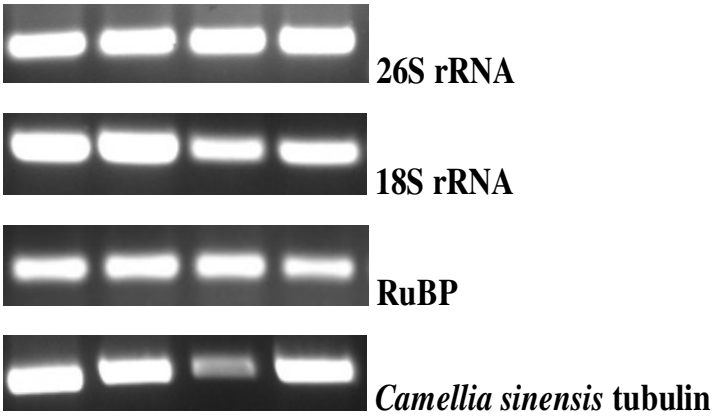

Figure 2 Semi quantitative PCR amplification of the housekeeping gene across cDNA samples isolated from greenfly and thrips infested tea leaf.

However, till date, efficient and robust housekeeping or constitutive gene for carrying qPCR generated gene expression studies have not been defined in *Camellia sinensis* and this report is a first one of its kind in tea

Plant materials and conditions

We performed the quantitative gene expression assays for four candidate reference genes/housekeeping genes namely Camellia sinensis tubulin, rubisco bis phosphatise dioxygenase, 18s rRNA and 26s rRNA. The samples which were considered were collected from a diverse range of conditions and the same kind of tissues i.e. the leaves, since this is the economically important tissue in tea. The following conditions were studied Darjeeling weather condition and Assam weather condition grown tea clones

Blight induced tea leaf samples

Herbivory induced leaf samples (Scirothrips dorsalis infested, greenfly infested and *Helopeltis* theivora infested)

Drought induced tea leaf samples

Methyl jasmonate induced, salicyclic acid induced suspension culture of tea. There is a need to take forward tea research to arrive at improvement programmes for tea breeding, molecular biology and biotechnology. The study comprised of different experimental setups to ultimately derive a set of housekeeping genes which can be used in gene expression profiling experiment in tea under any kind of biotic and abiotic stress.

RNA extraction

RNA was extracted from each of the biological sample to yield six RNA sample replicates from leaves. The amount of leaf used for each extraction was 100 mg to yield 0.4 µg of total RNA (approximately). The leaves were ground with a mortar and pestle using liquid nitrogen. Total RNA was extracted using guanidine-HCl RNA extraction protocol with slight modification.

100 mg fresh leaf tissue was used for the experiment. The leaf tissues were collected and immediately placed in liquid nitrogen. The 100 mg fresh leaf tissues were ground in liquid nitrogen. 1 ml extraction buffer was used to homogenize the ground tissues. The homogenized mixture was then mixed with 1 ml of PCI (25:24:1) and centrifuged at 14500 rpm for 20 minutes. The upper aqueous was transferred to a fresh tube and again mixed with PCI (25:24:1). The mix was then centrifuged at 14500 rpm for 15 minutes. The upper aqueous phase was transferred to a fresh tube. $1/10^{th}$ volume 3M NaOAc was added to the upper aqueous and double volume of absolute ethanol was added. Then the mixture was incubated at -20 °C for an hour. Following incubation the sample was centrifuged at 14500 rpm for 20 minute. A dirty whitish pellet was obtained upon centrifugation. The pellet was then washed with 70% ethanol. The pellet was then air dried and dissolved in DEPC treated water.

Following extraction the isolated total RNA was gel electrophoresed in 1% agarose gel to check for its RNA quality, integrity and concentration. Figure 3 depicts the gel electrophoresis of the isolated total RNA from infested samples and these intact and pure RNA was used for downstream applications in cDNA preparation and real time qPCR assays.

Spectrophotometry

Absorption measurements are carried out in the ultraviolet and visible light ranges. Nucleic acids are measured at 260 nm, proteins directly at 280nm and colorimetrically at 550 to 600 nm. The absorption of RNA is measured in buffered solutions (pH 7-8, e.g. RNA storage buffer, DEPC treated water) at 260 nm. Absorption of 1 OD (E) correspond 40 µg/ml RNA. Impurities in the samples were detected from the ratios of various absorption values. Since proteins absorbs at 280 nm, the purity of a nucleic acid was estimated from the E260/E280 ratio. The E260/ E280 ratio for the samples ranged from 2.00-2.20 after background correction. We considered furthering our experiments in samples which abided by the above mentioned standard.

Reverse transcription

Moloney Murine Leukaemia virus (MMLV) reverse transcriptase with RNase H activity was used to generate

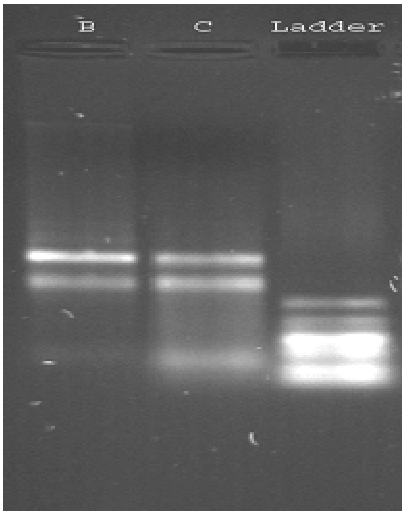

Figure 3 Gel electrophoresis of total RNA in 1% agarose gel the ladder used is high low weight RNA ladder.

cDNA strands following supplier's protocol. Briefly, RT was carried out in a 20 µl mixture including primer (OligodT) and 200 ng to 2 µg total RNA. The reaction was performed at 50 to 60°C using Transcriptor single strand synthesis kit, the incubation was performed for 1 hour and was stopped by hit inactivation of the enzyme, as described by the manufacturers. A brief overview of the protocol is as follows:

1 µg of total RNA or 10 ng of mRNA was taken as a starting material. The RNA was then allowed to mix with oligo dT primers to inactivate the secondary structure of the single stranded RNA by incubating at 65 ºC for 10 minutes. Other components were then added namely first strand synthesis buffer, reverse transcriptase, dNTPs and RNase inhibitor to the secondary structure inactivated RNA-oligo mix. The resulting mix was incubated at 50 ºC for an hour. The activity of the reverse transcriptase was stopped by incubating the reaction mix at 85 ºC for 5 minutes. The first strand of cDNA so synthesised can be stored at -20 ºC till further use.

Housekeeping genes and corresponding primers in the study

Sequences for the house keeping genes were obtained from the public databases viz. the DDBJ and NCBI. The four genes considered in the study were 18S rRNA, 26S rRNA, RuBP and Camellia tubulin. The corresponding GenBank accession number and primer pair sequence for the gene are

18S rRNA (GenBank/AY563528) (5' GGCCGGCTCCGTTACTTTG 3'/ 5' GTTTCAGCCTTGCGACCATACTC 3'),

26S rRNA (GenBank/AY283368) (5'TCAAATTCCGAAGGTCTAAAG3'/5'CGGAAACGGCAAAAGTG3'),

Ribulose-1, 5-bisphosphate carboxylase/oxygenase (GenBank: EF011075.1) (5'AAGCACAATTGGGAAAAGAAG3' /5'AAAGTGAAAATGAAAAGCGACAAT3') and camellia tubulin (GenBank/DQ444294) (5'AGCGTGCGGTTTGCATGA3' /5' GCCCAAAGGTTTGGCATCA3').

Housekeeping genes amplicon and size validation and pcr efficiency

Melting curve analysis, PCR efficiency check and gel electrophoresis of PCR amplicons size is not always indicative that multiple gene is not being amplified, so it is essential that we check and recheck the primers designed and product amplified by these primers. To nullify this factor therefore, the housekeeping genes considered in the study were initially amplified for using the primer specific to these genes designed using Primer 3 (http://frodo.wi.mit.edu/primer3/). (PCR conditions attached as appendix A).The amplicons obtained were of the desired size, which was estimated by calculating the base pairs spanning the region between the forward and reverse primers upon gel electrophoresis. This amplicons were further cloned in pGEMTEasy vector after eluting it from the agarose gel and sequenced in ABI 3130 *XL* genetic analyzer. The sequence thus obtained was then aligned against the public database BLASTx (http://blast.ncbi.nlm.nih.gov/Blastx.) and homology to available sequence in the database was documented. The PCR efficiency for each gene in all samples was obtained and assays with an efficiency of 1.98-2.0 were considered for further analysis.

Quantitative real time pcr

Light cycler 480 SYBR Green I Master was used to carry the expression assays. Unpurified cDNA sample after reverse transcription was used. To improve our results we applied 10 minutes pre incubation at 95 °C. This resulted in lower crossing point (CP) values with a decreased standard deviation. PCR primers at a final concentration of 0.4 uM were used; both forward and reverse in the reaction. The standard procedures were optimized for use with the light cycler 480 II.

Data on the expression levels of studied factors were obtained in the form of crossing points (CP) as described earlier (Rasmussen 2001). The data acquisition was done employing the 'second derivative maximum' method (Rasmussen 2001) as computed by the LightCycler Software (Roche Diagnostics). For further data analysis

Table 1. BestKeeper analysis

	18s	26s	RuBP	*Camellia* tubulin
BestKeeper vs.	HKG 1	HKG 2	HKG 3	HKG 4
coefficient of corr. [r]	0.663	0.23	0.82	0.711
p-value	0.001	0.237165	0.001	0.001

the Excel based application BestKeeper and geNorm was programmed to perform the computing procedure.

Data acquisition and analysis

We analysed the gene expression stability for 4 set of housekeeping gene across various experimental samples. All the experimental condition exhibited different stability patterns for the house keeping gene when each condition was analysed in seclusion, geNorm and BestKeeper are two most popular software used to calculate the stable internal control in real time quantitative PCR.

Michael Pfaffl devised an algorithm in 2003 to analyse the best set of housekeeping gene exhibiting the most stable gene expression across different experimental samples termed the BestKeeper. BestKeeper is an excel based algorithm which require to input the PCR efficiency of each qPCR assay for each sample and the corresponding crossing point value to determine the repeated pair wise correlation analysis. The use of the software, geNorm, is also very popular among biologists to normalise their gene expression data; where Cp values of all samples corresponding to each housekeeping genes are exported to Excel and relative quantities are calculated using the gene-specific PCR amplification efficiency. These relative quantities were then exported to geNorm (version 3.5) to analyse gene expression stability (2004).

RESULTS AND DISCUSSION

Descriptive analysis of gene expression based on BestKeeper

Descriptive statistics of the derived crossing points (assays run in a LightCycler 480 Ver. II), was performed based on BestKeeper program, and were calculated to investigate the variation level of each candidate gene following Pfaffl et al., (2003). The BestKeeper software employ an approach wherein the pair-wise coefficient of correlation is calculated between the housekeeping genes and a best keeper index calculated taking the crossing point values as input data. The expression stability can be calculated as a measure of the coefficient of correlation where the correlation coefficients can range from -1.00 to +1.00. The value of -1.00 represents a perfect negative correlation while a value of +1.00 represents a perfect positive correlation. A value of 0.00 represents a lack of correlation.

Descriptive analysis using BestKeeper in all experimental condition

Following the estimates of variation for the different experimental setup using one of the most popular qPCR analysis software BestKeeper, the results for expression stability is presented in table 1. Here the crossing point values for all the genes in all the samples under different experimental condition were used to compute the pair wise comparison. Table 1 describes the pair-wise correlation analysis between all the experimental conditions taken together and the housekeeping genes. It is an estimate where all the possible combination of the genes and samples are drawn and a relative comparison between them is calculated. The calculated estimate of relatedness describe whether the combination have a positive relationship for the gene in all samples considered or a negative relation between the genes and all experimental conditions considered. Negative correlation is determined by a value -1.00 and a positive correlation is determined by a value +1.00. The data analysed for 4 housekeeping genes in six experimental types and 28 sample (the experimental setup and treatments comprised of tea clones from 2 different species namely *Camellia sinensis, Camellia assamica*) proved RuBP to be the most stable gene with a coefficient of correlation of 0.82.

The results obtained depicted 26S rRNA as the most unstably expressing housekeeping gene. We would suggest the application of expression levels of RuBP for normalising the gene expression variation in GOIs' gene expression. The result also suggests that 26S rRNA gene expression cannot be considered for use in normalising expression data in different target gene across the experimental samples.

The above results were obtained from samples when considered together. To check the worth of the experimental results, we also considered analysing the data from individual experimental setups. There was a variation in stability of housekeeping genes in

Table 2. GeNorm analysis

	18S	26S	Tubulin	rbcl	Experimental condition
M < 1.5	-----	------	------	2.2	Darjeeling clone 2 climates
M < 1.5	------	------	------	1.95	Assam clones 2 climates
M < 1.5	0.971	------	------	------	Blister blight infection time course experiment
M <1.5	0.275	------	------	------	Blister blight infection tolerance study
M < 1.5	------	------	------	1.144	Helopeltis infestation time course experiment
M < 1.5	------	------	------	0.074	Greenfly and thrips infested tea samples
M < 1.5	1.140	------	------	------	Drought induced 2 leaves and a bud
M < 1.5	------	------	0.394	------	Drought induced progeny screening
M < 1.5	2.689	------	------	------	MeJA and SA elicited suspension culture
M < 1.5	------		1.863	------	Assam quality clones

experiments when analysed in isolation, as 5 out of the 10 experimental setups/condition depicted 18S rRNA gene as a stable housekeeping gene. 2 experimental conditions observed RuBP as the most stably expressing gene, 2 experimental conditions reported the expression of 26S rRNA gene to be stable and 1 experimental condition depicted camellia tubulin genes' expression to be stable. Following are the elaborated analysis of the BestKeeper index comparison with the internal control genes across all the experimental samples along with the geNorm analysis.

Analysis using BestKeeper and geNorm in different samples

Following are the experimental conditions considered for study and also the analysis for normalised expression data based on two different softwares namely BestKeeper and geNorm.

Clones grown at 2 different climate:

For this experimental setup we again subdivided our analysis into two conditions based on the genotype of the clones. In first experimental condition Darjeeling clones were taken for study and the second comprised of Assam quality clones, however, the constant factor in both the analysis was the climate in which they were grown, both set of clones were grown in Assam and Darjeeling climate

Assam clones

Lowest expression levels was observed in *Camellia* tubulin in which CP values were around 22-25 and the highest expression was observed for 26 S gene whose CP values ranged from 8-25. The levels of gene expression where the most stable gene expression was measured on the basis of standard deviation and coefficient of correlation for 26S rRNA gene (r=0.959), Ribulose-1, 5-bisphosphate carboxylase/oxygenase (r=0.874) followed by Camellia tubulin gene (r=0.767) using BestKeeper gene expression analysis software. The analysis using geNorm depicted an altogether different story which proved that RuBP and camellia tubulin are the comparatively stable housekeeping gene in Assam clones. Supplementary data sheets depict the corresponding value of expression stability for housekeeping genes across Assam clones grown in 2 different climatic condition and geographical location.

Darjeeling clones

The analysis showed lowest expression level in the housekeeping gene tubulin whose CP value range was from 23-36 cycles. Similarly the highest expression was observed in 26 s whose CP value was obtained around cycles 5-29. These clones exhibited a similar pattern in level of gene expression as in Assam clones. According to the calculated coefficient of correlation values it was observed that Ribulose-1, 5-bisphosphate carboxylase/oxygenase (r=0.988) was the most stable housekeeping gene in the study followed by tubulin and 26 S. The gene expression analysis for stable housekeeping genes in Darjeeling clones grown under both the climatic condition recorded from geNorm analysis showed that RuBP had the most stable expression as depicted by the M value of 2.2 (See table 2). These clones found most stable housekeeping gene to be RuBP.

Blight induced tea leaf samples

We performed two experimental design for blight infected samples; first experiment deals with the time course infection of blister blight and the other was designed to derive with the expression of the most tolerant and susceptible clones upon subjection to blister blight infection.

Time course experiment

When we considered studying the expression stability of 4 housekeeping genes in different blight infected and uninfected leaf sample of tea, we could observe that a gradient in genes expression for the 4 genes was present, where the highest levels of expression was determined for 26s gene which had a range of CP values in cycle number around 8-10. The lowest observed expression levels as per the CP values were determined around cycle number 31-35 for *Camellia* tubulin gene in the blight treated and untreated samples. Upon descriptive analysis of the data in Best keeper software the SD and (r) values for the 4 genes helped us conclude that the most stably expressing housekeeping genes in the blight infected sample was Ribulose-1, 5-bisphosphate carboxylase/oxygenase(r= 0.803) followed by 18S rRNA, 26S rRNA (SD=0.47125) and *Camellia* tubulin, which showed a (r) value of 0.376, 0.248 and 0.134 respectively. The analysis in geNorm showed 18S rRNA gene expression to be the most stable with a M (expression stability measure) value of 0.971 which was a good stability far less than the accepted threshold of 1.5.

Tolerant and susceptible clones for blister blight resistance in tea

The analysis for stable housekeeping gene employing BestKeeper software showed 18S rRNA (r=0.996) gene to be the most stable and valid gene to be considered for experimenting Darjeeling tea samples for resistance to blister blight. The geNorm analysis also exhibited 18S rRNA gene to be the most stable gene for consideration in normalising gene expression assays with a stability of 0.275 which is far above the permissible threshold of 1.5.

Herbivory infested leaf samples

***Helopeltis* infestation**

The level of gene expression was highest for 26 s as depicted by the CP values of the samples for this gene which was around 5-9 cycles. The lowest expression level was recorded for camellia tubulin gene which was around 22-25 cycles. Like the other samples the SD and CV values for the different samples in the experiment showed that the most stably expressing housekeeping gene for this kind of infested tissue was the 18S rRNA with a (r) value of 0.831. However geNorm analysis showed that among the 4 genes considered, RuBP was the most stably expressing gene with a stability measure of 1.144.

Greenfly and thrips infested tea leaf samples

The expression profile for the housekeeping genes in greenfly and thrips infested Darjeeling tea samples depicted the highest expression for RuBP with around 19-22 cycle, and lowest expression of camellia tubulin of around 28-37 cycles. Coefficient of correlation (r=0.916) values depicted 18 S rRNA to be the most stable housekeeping gene in greenfly and thrip infested Darjeeling tea leaf. The most stable gene as depicted by geNorm for greenfly and thrips infested tea leaf samples was RuBP.

Drought induced tea leaf samples

Drought induced two leaves and a bud of tea; the analysis showed that highest expression was recorded for 26 S genes in drought induced two leaves and a bud of tea and the lowest expression was observed in 18 S rRNA genes. The most stable housekeeping gene for the drought induced tea leaf samples was 18S rRNA (r=0.998) using BestKeeper analysis software. The analysis using geNorm established 18S rRNA gene as the most stable gene with a stability of 1.140. This experimental design verified 18S rRNA gene to be the most stable upon analysis by both software.

Drought tolerant and susceptible progenies of tea clones in an artificial drought induction experiment

We found that 18S rRNA gene was the most stable gene with a coefficient of correlation of 0.98 upon analysis in BestKeeper software. The application of geNorm analysis in progenies of drought induced experiment showed *Camellia sinensis* tubulin gene to be the most stable housekeeping gene with an expression stability of 0.394 which is within the permissible stability value range and in fact it is a value under the most stable expression scale regime.

Methyl jasmonate and Salicyclic acid elicited tea suspension culture

The estimate of levels of gene expression revealed the highest expression for 26S gene in the elicited tea suspension cultures, the lowest expression for the same set of samples was recorded in *Camellia* tubulin gene expression; this estimate was based on crossing point values. The most stable gene for studying gene expression change upon elicitation was Ribulose-1, 5-bisphosphate carboxylase/oxygenase which calculated a SD=2.12, followed by *Camellia* tubulin, 18S and 26S which had a SD of 2.42, 4.78 and 5.93 respectively.

Conclusions

Ranking the most stable gene for consideration in *Camellia sinensis*

Stably expressing control gene should be such that its

Table 3. Ranking of stable genes

Sl. No.	Experimental condition	BestKeeper gene expression stability	GeNorm gene expression stability
1	Darjeeling clones in 2 climates	RuBP (r=0.977)	RuBP(M=2.2)
2	Assam clones in 2 climates	26s rRNA (r=0.959)	RuBP and *Camellia tubulin* (M=1.94)
3	Blister blight infected time course experiment	RuBP(r=0.803)	18S rRNA(M=0.971)
4	Experiment on resistance for blister blight in tea	18S rRNA(r=0.996)	18S rRNA(M=0.275)
5	Helopeltis infested tea samples	18S rRNA(r=0.831)	RuBP (M=1.144)
6	Greenfly and thrips infestation	18S rRNA(r=0.916)	RuBP (M=0.074)
7	Drought induced 2 leaves and a bud	18S rRNA(r=0.998)	18S rRNA(M=1.140)
8	Drought induced tea progeny	18S rRNA(r=0.98)	Camellia sinensis tubulin (M=0.394)
9	MeJA and SA elicited tea suspension cultures	26S rRNA(r=0.986)	18S rRNA(M=2.689)
10	Assam quality clones	26S rRNA (r=0.935)	Camellia sinensis tubulin (M=1.863)

expression does not change in different tissues or cells under study for a particular organism. Ideally internal control genes exhibit universal acceptance for valid and constant expression levels across all possible tissue types, cells, experimental treatments and design. Reports till date show that no such housekeeping gene have been found (Schmittgen etal., 2000; Bustin etal., 2000; Suzuki etal., 2000; Thellin etal., 1999; Tricarico etal., 2002; Warrington et al., 2000). In the experiment conducted by us we saw a gradation where the same gene is not expressed stably in all experimental samples. If we consider the descriptive analysis of gene expression stability or the BestKeeper analysis in various tea leaf samples, it was observed that out of the 10 experimental conditions studied, 5 experimental conditions ranked 18S rRNA as the most stably expressing gene. The algorithm ranked the remaining housekeeping genes based on the calculated coefficient of correlation and standard deviation, however, the gene with the most stable expression was different in different experimental conditions. There are reports that no such housekeeping genes exists which expresses stably universally across sample types and conditions. It is therefore clear from these reports and observations from different crops, animal system (Vandesompele etal. 2002) and our crop, i. e, *Camellia sinensis* that a universally accepted valid housekeeping gene does not exist.

Even though such reports are coming in, it is a relief that all experimental setup do not consider necessity of such a gene because most experimental designs are restricted to a few tissue types or a few different developmental stages of the same tissue, e.g., normal and infested, untreated and treated, and so forth, and it is likely that one or more genes are constitutively expressed across such restricted designs. However, the task of identifying these genes is not an easy one. In cases where no stable housekeeping gene is found, it is essential to normalize the expression data using a normalisation factor which have been obtained from multiple genes. Software like geNorm allows the imputation of normalisation factors in different treated samples for multiple genes. We calculated the normalisation factor for the samples in our study using geNorm for 4 housekeeping genes. The justification for such a design is simple; the variation that is observed while considering multiple genes is less than that observed with single gene. This is not an indication of improved normalisation. However, it is expected that if the selection of gene is done with caution an improved normalisation may be obtained. The imputation of normalisation factor from multiple genes requires computation of expression across all experimental conditions for all the genes. The number of genes would be a substitution between practical consideration and also allowing minimization of variation in the normalisation factor. The analysis of the present study is represented in table 3. The table also rank the most stable gene in different experimental conditions.

REFERENCES

Andersen CL, Jensen JL, Orntoft TF (2004). 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(15):5245-5250.

BestKeeper Software [http://www.gene-quantification.de/bestkeeper.html]

Brunner AM, Yakovlev IA, Strauss SH (2004). Validating internal controls for quantitative plant gene expression studies. BMC Plant Biology 4-14

Bustin SA (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol, 29:23-39.

Bustin SA, Nolan T (2002): Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J Biomol Tech 2004, 15(3):155-166.

Bustin SA (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mol.

Endocrinol. 25: 169–193.

Cavallari CFB, Severino FE, Maluf MP, Maia IG (2009). Identification of suitable internal control genes for expression studies in coffea arabica under different experimental conditions. BMC Mol Biol, 10:1

Cho JY, Mizutani M, Shimizu B, Kinoshita T, Ogura M, Tokoro K, Lin M, Sakata K (2007). Chemical Profiling and Gene Expression Profiling during the Manufacturing Process of Taiwan Oolong Tea "Oriental Beauty". Biosci. Biotechnol. Biochem. 71(6): 1476-1486

Gachon C, Mingam A, Charrier B (2004). Real-time PCR: what relevance to plant studies? J Exp Bot 55(402):1445-1454.

geNorm Software [http://medgen.ugent.be/~jvdesomp/genorm/]

Hara Y, Luo SJ, Wickremasinghe RL, Yamanishi T (1995). Special issue on tea. Food Rev. Int. 11(3): 371-545.

Huggett J, Dheda K, Bustin S, Zumla A (2005). Real-time RT-PCR normalisation; strategies and considerations. Genes Immun 6(4):279-284.

Kim BR, Nam HY, Kim SU, Kim SI, Chang YJ (2003). Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice. Biotechnol. Lett. 25: 1869–1872.

Lupton FGH (1984). Biological control-The Plant Breeders objective. Ann. Appl. Biol. 104: 1-16.

Maroufi A, Van Bockstaele E, De Loose M (2010). Validation of reference genes for gene expression analysis in chicory (Cichorium intybus) using quantitative real-time PCR. BMC Mol Biol. 11: 15.

Muraleedharan N, Selvasundaran R, Radhakrishnan B (1988). Natural enemies of certain tea pests occurring in southern India. Insect Sci. Appl. 5:647-54

Pfaffl M (2003). Livestock Transcriptomics: Quantitative mRNA Analytics in Molecular Endocrinology and Physiology, Habilitation submitted to the Faculty Center of Life and Food Sciences, Technische Universität München - Weihenstephan

Pfaffl MW (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res, 29:e45.

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. Biotechnol Lett, 26(6):509-515.

Rani A, Singh K, Sood P, Kumar S, Ahuja PS (2000). p-Coumarate:CoA ligase as a key gene in the yield of catechins in tea [Camellia sinensis (L.) O. Kuntze]. Funct. Integr. Genomics. 9(2): 271-275

Rasmussen R (2001). Quantification on the LightCycler. In: Meuer, S, Wittwer, C. and Nakagawara, K, eds. Rapid Cycle Real-time PCR, Methods and Applications. Springer Press, Heidelberg.

Schmittgen TD, Zakrajsek BA (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods, 46: 69-81

Singh K, Raizada J, Bhardwaj P, Ghawana S, Rani A, Singh H, Kaul K, Kumar S (2004). 26S rRNA-based internal control gene primer pair for reverse transcription-polymerase chain reaction-based quantitative expression studies in diverse plant species. Anal. Biochem. 335:330–333.

Suzuki T, Higgins PJ, Crawford DR (2000). Control selection for RNA quantitation. Biotechniques, 29: 332-7,

Thellin O, Zorzi W, Lakaye B (1999). Housekeeping genes as internal standards: use and limits. J Biotechnol, 75: 291-5,

Tricarico C, Pinzani P, Bianchi Sl (2002). Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. Anal Biochem, 309: 293-300,

Vandesompele J, De Preter K, Pattyn F (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: 34.

Volkov RA, Panchuk II, Schoffl F (2003). Heat-stress-dependency and developmental modulation of gene expression: the potential of house-keeping genes as internal standards in mRNA expression profiling using real-time RT-PCR. J. Exp. Botany 54: 2343-2349

Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M (2000). Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol Genomics, 2: 143-7,