Identification of best housekeeping gene for normalization in qRT-PCR data under different development stages and stress conditions in Vigna mungo

Sudhir Kumar Jha
ICAR-Indian Institute of Pulses Research
Pritee Sagar
ICAR-Indian Institute of Pulses Research
Prateek Kumar
ICAR-Indian Institute of Pulses Research
Shivangi Gupta
ICAR-Indian Institute of Pulses Research
Ankita Anil Chakote
ICAR-Indian Institute of Pulses Research
Vaibhav Kumar
ICAR-Indian Institute of Pulses Research
Kalpana Tewari
ICAR-Indian Institute of Pulses Research
Antara Das
ICAR-Indian Institute of Pulses Research
Meenal Rathore
ICAR-Indian Institute of Pulses Research
Girish Prasad Dixit
ICAR-Indian Institute of Pulses Research
Kuldeep Kumar
kuldeepdbc@gmail.com

ICAR-Indian Institute of Pulses Research

Article

Keywords: Housekeeping gene, Normalization, qRT-PCR, urdbean

Posted Date: April 7th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5819031/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 26th, 2025. See the published version at https://doi.org/10.1038/s41598-025-10295-5.

Abstract

The qRT-PCR has emerged as the most sought-after local gene expression profiling technique. The absolute quantification of the mRNA is highly affected by the handling artifacts. In this regard, the internal control genes serve a vital role in negating the handling errors. So far, the most stable housekeeping genes to be used for the normalization of qRT-PCR data are lacking in *V. mungo*, commonly known as urdbean or blackgram. Therefore, in the current study we intend to explore the stability of the UFO, TUB2, ACT2, RHA, ISTL6, CyP, DNAJ, RBCS3B, RPS34, USP21, USB1, EfTu, UBQ10, and UBQE9 genes of *V. mungo* in 17 different developmental stages and 4 different abiotic stresses. The well-acknowledged algorithms such as geNorm, NormFinder, BestKeeper, the comparative ΔCt method and comprehensive RefFinder were used to assess the expression stability. According to the comprehensive ranking, RPS34 and RHA were identified as the most appropriate genes for qRT-PCR data normalization throughout all developmental stages, whereas ACT2 and RPS34 were optimal under abiotic stress conditions. Further, the suitability of these three candidate housekeeping genes for normalization has been demonstrated in an independent experiment. Based on the findings obtained, the present research advocates the use of a combination of HKGs (RPS34 and RHA for developing tissues and ACT2 and RPS34 for stress-treated tissues) for the normalization of qRT-PCR data. The proposed reference genes will guarantee the reliability and consistency of qRT-PCR results.

Key Message

The gene stability ranking indicated that RPS34 and RHA were best for qRT-PCR data normalization in various developmental stages, whereas ACT2 and RPS34 under abiotic stress.

Introduction

With the advent of sequencing technologies, abundant information and genomics resources are being developed on a daily basis in several legumes, which were otherwise considered as “orphan crops.” To date, genome sequences for several legumes, including *Cajanus cajan*, *Cicer arietinum*, *Vigna mungo*, and *Cyamopsis tetragonoloba* have been published [1–4]. Parallelly, several of these legume crops have been supplemented with multiple transcriptomics experiments to mine genes related to any development stage and their biotic-abiotic response and have been reviewed [5–7]. Gene expression profiling, either ‘global’ or 'local', is becoming vital in any advanced molecular biology and molecular breeding experiments. Even the sole transcriptomics experiments require validation of key genes to make the results more reliable and statistically more robust.

The expression of any gene is spatiotemporal and dynamic, as it is influenced by several internal as well as external factors. Quantification of mRNA is essential for understanding the gene expression levels in the biological samples. Various techniques, including quantitative reverse transcription polymerase chain reaction (qRT-PCR), northern blotting, RNA-seq, micro-array, *in-situ* hybridization, the Nanostring nCounter system, and the Serial Analysis of Gene Expression (SAGE), have been developed for this purpose, each with its own advantages and limitations. Among these techniques, qRT-PCR is the most widely used technique due to its robustness and accuracy. For accurate mRNA quantification, the qRT-PCR results are normalized using a reference gene/housekeeping gene (HKG) whose expression is stable across different experimental conditions, allowing for the quantification of changes that are due to biological differences rather than technical artifacts [8–10].

Urdbean, also known as blackgram (*V. mungo*), is the third most important grain legume after chickpea and pigeonpea in the Indian subcontinent. It is a nutrient-rich pulse possessing a rich blend of protein, fiber, and essential micronutrients like iron, zinc and potassium. Its adaptability to diverse agro-ecologies, shorter cropping duration, and ability to fix atmospheric nitrogen attract the small and marginal farmers [11]. As the world grapples with the burning issues like climate change, ever increasing population and nutritional security, the importance of the urdbean as a nutrient-rich and agriculturally important crop is being recognized recently. Despite this significance, it has long been overlooked by mainstream research and is treated as an “orphan crop”. Recently, efforts are being made to develop some genetic and genomic resources in urdbean. The availability of draft genome assembly, high density linkage map, along with the development of molecular markers such as SSRs and SNPs has enabled the identification of genes/QTLs linked with different traits. These genomic resources hold great promise for accelerating the urdbean improvement using the advanced genomics strategies [11]. Once the putative candidate genes are known their expression analysis needs to be performed, which requires a suitable HKG for the data normalization.

Several HKGs have been explored for normalization of the qRT-PCR data. In general, the HKGs used across the studies belong to central processes that are critical for the survival of the cell including, genes essential for the cell morphology, cell division, growth and development. Because these aforementioned genes are tied to fundamental and non-specialized processes, they are less likely to fluctuate in response to experimental variables that might alter the expression of genes involved in specific pathways. However, their stability should be validated for specific experiments, as no single gene is universally optimal for all situations. Therefore, in this work, we evaluated 14 genes of different functional classes (significantly reducing the chance that the genes will be co-regulated) using the qRT-PCR techniques in multiple stages, starting from cotyledon (germination) to pod (maturity) stages. The most suitable HKG for qRT-PCR in *V. mungo* was evaluated employing different algorithms developed to assess the stability of a normalizing gene through NormFinder, geNorm, BestKeeper, and the ΔCt method implemented in RefFinder software [12–16]. The most potential HKGs identified in the current study were further validated to demonstrate their suitability in normalizing the qRT-PCR data. The most stable HKGs identified in the current study will be highly useful for the researcher working on the advanced genomics experiments in Vigna species, including *V. mungo*.

Materials and methods

Plant materials

In the current work, an urdbean variety, “IPU13-1” released from the ICAR-Indian Institute of Pulses Research, Kanpur, was used as plant material. The variety has a potential yield of 10–12 q/ha with a maturation duration of 70–75 days. The seeds were sterilized for 30 seconds in 70% ethanol before sowing in clay pots filled with autoclaved soil. Four developmental stages, including germination [5 days after sowing (DAS)], the vegetative stage (30 DAS), the reproductive stage (45 DAS), and podding (60 DAS), were selected for the gene expression analysis. From the germination stage, plumule, radicle, and cotyledon tissues were harvested. While leaves, stems, and root tissue were harvested from all remaining three stages, including the vegetative stage, reproductive stage, and podding stage. Likewise, vegetative meristem, floral bud, and 15-day-old pods were harvested at the vegetative, reproductive, and podding stages, respectively. Further, the pods were also split into pod walls and seeds. All tissues were freshly harvested and snap-frozen into liquid nitrogen before storing them at -80°C until RNA isolation. All the samples were collected in three independent biological replicates.

Treatments

Four different treatments, including drought, salt, aluminium, and cold stress, were given to the sterilized seeds of IPU13-1 at the germination stage. The experiment for screening in salt, drought, and aluminium stress conditions was done in the customized growth chamber of REMI (CHM 12+) with a Floutronix grow light. The temperature was set at 35°C ± 2°C with relative humidity set to 65% and a light/dark period of 12 hours each. For salinity and aluminium stress, seeds were germinated in germination paper put in a petri-plate, followed by soaking with 120 mM NaCl and a 100 mM solution of $Al_2(SO_4)_3$, respectively [17, 18]. Likewise, the sterilized seedlings were germinated in MS media supplemented with 1% w/v of PEG 6000 for imposing drought stress [19]. Similarly, to collect tissues under cold stress, seeds were germinated in germination paper with the temperature set to 18°C ± 2°C with relative humidity set to 65% and a light/dark period of 12 hours each [20]. Tissues of plumule, radicle, and cotyledon from each of the aforementioned treatments were collected at 15 DAS in three independent biological replicates for the downstream analysis.

Total RNA extraction and cDNA synthesis

RNA isolation for each sample was done using the RNeasy Plant Mini Kit, Qiagen, Germany (Catalogue no. 74904), following the manufacturer’s protocol with slight modifications [21]. The DNase treatment was given to the isolated RNAs to remove the genomic DNA contamination. Total RNA integrity was checked in 2% agarose gel electrophoresis and the amount was quantified using a spectrophotometer (NanoDrop 2000c, Thermo-Scientific, USA). From the total quantity of extracted RNA, 1 µg was used for cDNA synthesis using Maxima H Minus Double-Stranded cDNA Synthesis Kit, Thermo-Scientific, USA (Catalogue no. K1612), following the manufacturer’s instructions. Finally, the cDNA was diluted with molecular-grade water in a ratio of 1:10 and used for qRT-PCR analysis [21].

Selection of housekeeping genes and primer synthesis

A total of 14 housekeeping genes were shortlisted based on the literature survey. The majority of these genes have not been tested in urdbean to date. The identifier for the selected genes was used to fetch the Arabidopsis protein sequences for the selected genes from TAIR (https://www.arabidopsis.org/). The reference protein sequences for the Urdbean genome were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_036885675.1/). Further, the NCBI blastp suit (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) was used for finding the urdbean homologs of the selected Arabidopsis protein sequences based on the similarity. The Urdbean protein with the maximum similarity for each Arabidopsis protein sequence was shortlisted for the downstream analysis. The protein sequences were further confirmed for functional conservation by analysis of conserved domains in both Arabidopsis and *Vigna mungo* homologs via NCBI-Batch CDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi; [22]). Primers from the selected genes were designed using the IDT's PrimerQuest tool (https://www.idtdna.com/PrimerQuest). PrimerQuest was chosen due to its robust primer design algorithms, which incorporate melting temperature, GC content, and secondary structure considerations. Further, the oligonucleotides were synthesized by Europhin Genomics, India. The primer sequences for these genes are provided in the supplementary file (Table S1).

The qRT-PCR analysis of the selected HKGs

Temperature gradient PCR was performed to optimize the specific Tm of each primer. Each 20 µL qRT-PCR reaction consisted of 10 µL master mix (Brilliant III Ultra-fast SYBR Green qRT-PCR Master Mix, Agilent Technologies), 0.3 µM forward and reverse primers, 7.4 µL PCR grade water and 50 ng cDNA as the template. The qRT-PCR reactions were carried out in an Agilent AriaMx Real-Time qRT-PCR system. The PCR conditions involved an initial denaturation at 94°C for 3 min, followed by 40 cycles of 95°C for 30 s, 60°C for 15 s, and 72°C for 30 s [21]. The amplicons were separated in 1.2% agarose gel using TBE buffer.

Expression stability analysis

For qRT-PCR data, raw Ct values (quantification cycle) were used for the analysis, and the mean Ct value of each biological replicate was considered to ensure the accuracy. For the expression stability analysis, the RefFinder web server (https://www.ciidirsinaloa.com.mx/RefFinder-master/; [16]) was employed. The RefFinder tool comprehensively utilizes several other gene stability analysis algorithms such as geNorm [13], Normfinder [12], BestKeeper [14], and the comparative ΔCt method [15]. The RefFinder tool utilizes only the original Ct value from quantitative real-time PCR (qRT-PCR) for the downstream analysis [16]. Accordingly, the mean values of each replicate were used for the analysis using the RefFinder tool.

The geNorm algorithms calculate the gene expression stability measure (M-value) based on pairwise variation analysis and also determine the optimal number of reference genes required for the normalization. Whereas, the NormFinder algorithm employs an ANOVA-based model to estimate intra- and inter-

group variation, providing a direct stability value for the HKGs tested. Likewise, the BestKeeper algorithm assesses the stability based on the standard deviation (SD) and coefficient of variation (CV) of the Ct values. The HKGs with SD > 1 are considered unreliable for normalization. Similarly, the ΔCt method compares the relative expression stability of the HKGs by analyzing the variation in their Ct values. The RefFinder algorithm comprehends an overall ranking based on the geometric mean of rankings from these four methods.

The statistical significance of candidate reference gene stability was determined based on rankings derived from the combined analysis. Further, to plot the mean stability value (geNorm), stability value (NormFinder), standard deviations (BestKeeper), arithmetic mean of the standard deviation (ΔCt method), and the geometric mean of the stability ranking (RefFinder) WPS office was used. The box-plot representing the distribution of the Ct values was also plotted using the GraphPad v10.4.1.

Gene normalizing efficiency of the candidate HKGs

Further, to validate the reliability of the most suitable HKGs (RPS34, RHA, and ACT2) identified in the current study, they were used as an internal control in the expression analysis of the 5 test genes previously reported by our group (Table S2) [23]. Parallelly, an external Actin (ACT) gene that has been previously reported and used in urdbean qRT-PCR was also used as an internal control. The control and infected leaf samples from Yellow Vein Mosaic (YMV) tolerant (DPU88-31) and susceptible (LBG17) genotypes were collected as previously reported by our group [23]. Further, the RNA isolation, cDNA library preparation, and qRT-PCR were performed using the aforementioned methodologies. The mean Ct values were calculated in Microsoft Excel, 2016 using all 4 candidate HKGs, and a one-way ANOVA followed by a Tukey's multiple comparisons test was performed at a significance level of $p \leq 0.05$.

Results

Vigna mungo is emerging as an important legume crop after chickpea and pigeonpea due to its high nutritive value, short cropping duration, and multiple cuisines made from it. It is an important blend in the vegetarian diets for the majority of the population in South Asia due to its high nutritive value [11]. Besides, legume crops are being targeted to achieve the second sustainable development goal of “Zero Hunger” [24]; urdbean can be a vital component in this battle. The Vigna species are highly versatile and can be accommodated to different cropping systems and environments, making them a good choice for advanced genomics experiments. Looking into the wide adaptability as well as utility of the *Vigna mungo*, it has attracted the farmers as well as researchers. The recent enrichment in genetic and genomic resources of urdbean has paved the way for the functional characterization of many candidate genes. And gene expression profiling is a key experiment in this regard, which further depends on HKGs for the normalization to get statistically robust results, taking care of the technical artifacts. Therefore, to identify the most suitable HKGs for normalization of qRT-PCR data in urdbean, we have conducted a comprehensive expression profiling of 14 candidate housekeeping genes in 17 different tissues.

Identification of the *V. mungo* orthologs of selected Arabidopsis HKGs

A total of 14 *V. mungo* candidate genes corresponding to each selected Arabidopsis gene were identified through NCBI blastp with the “max hit = 1” parameter. These 14 genes were considered for local gene expression analysis using the qRT-PCR technique in the 17 selected tissues (Table 1). The percent similarity between the gene orthologs of urdbean and Arabidopsis ranges between 47.337 and 99.738, among which 12 out of 14 have a percent similarity of over 65 percent, reflecting the sequence conservation (Table 1). All of the 14 genes were found to be annotated as hypothetical proteins, as the gene prediction pipelines are based on homology-based annotations but lack experimental validation. Since the genome of *V. mungo* was recently sequenced [3], this could be explained. The conserved domain analysis using NCBI-Batch CDD confirms the presence of the same domain as it was present in Arabidopsis ortholog (Figure S1). The higher percent similarity between the *V. mungo* HKGs and that of the Arabidopsis orthologs, as well as conserved domain analysis, advocates the functional as well as sequence conservation between them (Figure S1).

Sampling of diverse plant tissues

A total of 17 plant tissues representing different stages from germination to pod development were selected for the expression analysis. The tissues used in the current study capture all developmental stages in urdbean. The diversity in the type and stages of the tissues reinforces the utility of the HKGs identified in the current study to the urdbean research community (Figure 1).

RNA quality assessment and primer specificity

The quality of the RNA used in this work was evaluated to guarantee a proper estimate of gene expression. The 260/280 absorbance ratio of RNA samples was evaluated using the spectrophotometer and only the samples with absorbance between 1.9 and 2.1 were considered. This ensures the quality as well as purity of the RNA isolated. Sharp and clear bands matching 18S rRNA and 28S rRNA on the denaturing gel revealed intact, undamaged RNA molecules, therefore indicating strong RNA integrity (Figure 2a). Further, the temperature gradient PCR for each primer set was conducted to identify the optimal annealing temperature. The length of the amplified PCR product was verified using 2% agarose gel electrophoresis using a DNA ladder (100 bp) as a reference. In our study, the amplicon size was found to range between 85 bp and 260 bp (Figure 2b).

Gene expression stability analysis in different developmental stages

The qRT-PCR analysis of all selected 14 genes was conducted across the 17 tissues, representing the germination, vegetative, reproductive and pod development stages. The mean Ct (quantification cycle) values of all 14 genes from three biological groups were utilized for the estimation of stability across the three datasets (Figure 3). The candidate HKGs showed a large variation in mean Ct ranging from 10.84 (RBCS3B) to 39.82 (ISTL6 and CyP) while considering the total dataset (Figure 3). The whiskers represent the quartile where the Ct values for most of the HKGs fall, while the middle line represents

the middle (median) Ct value in the dataset. The broader whisker means more variability in Ct value, while the smaller whisker supports the stable gene expression across the tissues (figure 3). Further, the data generated was analysed through the RefFinder algorithm, which also has other gene stability analysis algorithms such as geNorm, Normfinder, BestKeeper, and the comparative ΔCt method implemented in it.

The geNorm algorithm calculates the mean stability value of the housekeeping genes by comparing their average pairwise variation with all other housekeeping genes considered. The geNorm algorithm relies on the principle that the expression ratio of two ideal internal control genes remains consistent across all samples, irrespective of experimental condition or cell type [25]. Therefore, the genes with the lowest average pairwise variation are considered to be the most stable. Using the geNorm algorithm, RPS34 and RHA genes were identified to be the most stable, with a mean stability value of 1.238, whereas the UFO gene was identified to be the least stable, with a mean stability value of 3.374. In the geNorm algorithm, any “M” value (geometric mean) below 1.5 is considered indicative of stable gene expression [13]. In our analysis, the top three genes i.e. RPS34, RHA and CyP were having a geometric mean stability value below 1.5. The mean stability value of all 14 genes as estimated through the geNorm algorithm is provided (Figure 4).

Similarly, in the comparative ΔCt method, the ΔCt values of genes are compared like geNorm. In the ΔCt comparison method, the arithmetic mean of the standard deviation in the pairwise comparison is considered. The lower the arithmetic mean, the higher will be the stability of the housekeeping gene. Similar to the geNorm method, in the comparative ΔCt method, RPS34 and RHA were found to be the most stable housekeeping genes, with the arithmetic mean and standard deviation of 2.51 and 2.73, respectively (Figure 5). Likewise, the UFO gene was found to be the least stable housekeeping gene with an arithmetic mean of standard deviation of 4.95 (Figure 5). The gene stability ranking of all 14 genes as estimated through the ΔCt method algorithm is provided (Figure 5).

In the BestKeeper algorithm, an index value based on the correlation analysis between the Ct values of the genes is calculated. Further, standard deviation, percent covariance explained, and power are calculated. Thereafter, the genes are ranked based on their stability metrics, with the most stable gene having the lowest standard deviation and highest correlation with others. In our analysis, the BestKeeper algorithm identified the USB1 gene to be the most stable, while EfTu was identified to be the least stable (Figure 6). The gene stability ranking of all 14 genes as estimated through the BestKeeper method algorithm is provided (Figure 6). The NormFinder algorithm estimates not only the overall variation of the candidate reference genes but also the variation between subgroups of samples. Unlike the geNorm and ΔCt comparison methods, NormFinder assesses the expression stability of each candidate independently. In our analysis, the RPS34 and ACT2 were found to be most stable using the NormFinder algorithm (Figure 7). The gene stability ranking of all 14 genes as estimated through the ΔCt method algorithm is provided (Figure 7).

The RefFinder algorithm compares the weights estimated for each reference gene through the aforementioned four algorithms to calculate the geometric mean of the weights and make a comprehensive ranking. Through the RefFinder, we identified RPS34 and RHA as the best and most stable housekeeping genes, while EfTu and RBCS3B were the least suitable (Figure 8). Among the four algorithms implemented in RefFinder, RPS34 was found to be the best housekeeping gene in the geNorm, ΔCt comparison method, as well as the NormFinder algorithm, except for the BestKeeper algorithm, where USB1 was found to be the best as depicted in the Venn diagram when the top three most stable genes were compared (Figure 9). Similarly, RHA was among the top three in geNorm, the ΔCt comparison method, and BestKeeper when the top three most stable genes from each of these algorithms are compared (Figure 9). Overall, the RPS34 and RHA genes were also found to be most stable and suitable to be used as control genes for normalization study through the RefFinder algorithm. This was also supported by the whisker plot of the Ct values (Figure 3). The gene stability ranking (Better--Good--Average) of the 14 genes in different developmental stages is provided (Table 2).

Gene expression stability analysis in different abiotic stress

The expression pattern of all of the 14 genes was further explored under four abiotic stresses *viz.* salt, drought, aluminium, and cold stress in the germination stage tissues (plumule, radicle and cotyledon). In the four abiotic stresses, RPS34 was found to be the most stable, followed by USB1, while RBCS3B was least stable using the geNorm algorithm (Figure S2). Likewise, using the ΔCt method, ACT2 and RPS34 genes were identified to be the most stable HKGs, while RBCS3B was found to be the least stable (Figure S3). In the four abiotic stresses, USB1 was found to be the most stable followed by DNAJ, while EfTu was least stable using the BestKeeper algorithm (Figure S4). Likewise, using ACT2 and DNAJ genes were identified to be the most stable HKGs, while RBCS3B was found to be the least stable using the NormFinder algorithm (Figure S5). Comprehensively, RefFinder identifies ACT2 and RPS34 as the most stable HKGs (Figure S6). The gene stability ranking (Better--Good--Average) of the 14 genes in different abiotic stresses in seedling stages is provided (Table S3).

Validation of the candidate HKGs identified in the current study

For validation of the best HKGs identified in the current study, we used the Ct value of RPS34, RHA and ACT2 HKGs to normalize the expression of 5 previously reported genes i.e., CAM, PR1, HSP70, NAC, and DEF genes, along with an external control gene, i.e., Actin (Table S2). For the experiment, the tissues and primers were generated and used as previously reported [23]. The gene expression pattern of all 5 test genes was in concurrence with the previous reports [23]. The larger deviation in the relative fold change value of the test gene when normalized with ACT as a control was clearly evident (Figure 10a-e). On the other hand, when the test genes were normalized with the HKGs identified in the current study, a substantial variation was observed. The normalized relative gene expression level clearly reflects the stability as well as the suitability of using RPS34, RHA, and ACT2 genes over ACT in the qRT-PCR experiment (Figure 10a-e).

Discussion

The gene expression is a highly spatiotemporal event influenced by both external and internal factors. The profiling of gene expression patterns is among the most sought-after preliminary experiments in both animal and plant sciences [26]. Majorly done by qRT-PCR, there are chances of error due to pipetting and

RNA quality affecting the imprecise estimation of gene expression level [27]. These probable errors are taken care of by using a good reference gene for the normalization of the data generated in the qRT-PCR experiments [13]. Specifically, the expression of the reference gene needs to be consistent across the spatiotemporal conditions and the different development stages. Therefore, while doing a gene expression analysis, selecting the most trustworthy reference genes has become a crucial research endeavor [28]. Limited efforts have been made to identify the suitable genes in legume crops, including *Lupin angustifolius*, *Vigna angularis*, *Lens culinaris*, *Vicia faba*, and *Cicer arietinum* [29–33].

So far only a handful of studies involving expression profiling have been conducted in *V. mungo* for different biotic and abiotic stresses [23, 24, 34]. The current study holds the potential to facilitate future research aiming to accurately explore the expression profiling of any selected gene across the developmental stages. To ensure the preciseness of the study, due care has been taken right from RNA isolation to qRT-PCR. Further, the majority of the candidate genes selected for the present study were novel to *V. mungo* and have not been explored to the best of our knowledge.

A clear consensus on which HKGs to be used for data normalization in *V. mungo* is lacking. Therefore, comparing the expression Ct values obtained from profiling multiples samples using different statistical algorithms will pave the way forward. Hence, the current study was designed to profile a set of 14 HKGs in 17 different tissues representing different stages right from seed germination to seed maturity. The genes used in the current study represent diverse metabolic and developmental pathways, including floral development, cellular cytoskeleton, RNA and protein processing, and protein trafficking. Diversity in the metabolic pathways and developmental stages used in the current study makes the study more reliable and of better utility for the *V. mungo* researcher community. Further, five different algorithms, including geNorm [13], NormFinder [12], BestKeeper [14], the comparative ΔCt method [15], and RefFinder were utilized in the current study [16].

Interestingly, two genes, RPS34 and RHA, were found to be the two most stable HKGs using the different statistical algorithms (Table 2), with the exception of the BestKeeper algorithm. Likewise, RBSC3, EfTu and UFO were found to be least stable across the different developmental stages in the current study (Table 2). The ribosomal proteins have been reported to be major HKGs in all domains of life [35]. Since RPS34 is an integral part of protein biosynthesis machinery, its stable expression could be explained. Surprisingly, any of the ribosomal proteins, although expressing stably, have not been explored much as a suitable HKGs for qRT-PCR normalization. The current study advocates to exploring the ribosomal proteins for gene normalization studies in other species. Likewise, RNA helicases are enzymes that are crucial for RNA metabolism. They unwind RNA secondary structures and/or displace RNA-binding proteins and are essential for various cellular processes involving RNA, including pre-mRNA splicing, stability and decay [36]. Since RNA helicase is ubiquitous and performs several functions vital for cell growth and development, their listing in the most stable HKGs is not surprising. Despite modest changes in the ranking of other genes, likely due to variation in the algorithms used, the findings clearly indicated that RPS34 and RHA consistently were most stable. Although ACT2 outperforms the RPS34 gene under all four abiotic stress conditions, RPS34 was still among the top stable HKGs.

Kundu and his coworkers put an effort to identify the best suitable HKGs in *V. mungo* under biotic (YMV) and abiotic stress (drought) but with only 7 HKGs in two stress conditions [34]. Comparably, our study not only uses more HKGs but also represents an array of developmental stages (17 tissues) along with the 4 stress conditions (salt, drought, aluminium and heat). The candidate HKGs identified in the current study were further demonstrated to be stably expressed and normalize the Ct value of the 5 test genes under independent experiment conditions. The HKGs identified in the current study i.e. RPS34, RHA, and ACT2 genes hold a greater potential to be widely used as HKGs in any gene expression experiment. Since this is the most comprehensive report on the identification of HKGs in Vigna species, the HKGs identified in the current study might be useful to the other Vigna species.

Conclusion

The current study explores the stability pattern of 14 HKGs representing diverse metabolic and development pathways in 17 different developmental stages from germination to seed maturity in *Vigna mungo*. The Ct values obtained were implemented to get a gene-wise ranking based on geNorm, NormFinder, BestKeeper, comparative ΔCt method and the RefFinder algorithm. Based on the comprehensive ranking, RPS34 and RHA were found to be most suitable for qRT-PCR data normalization in different developmental stages, while ACT2 and RPS34 were suitable under abiotic stress conditions. Based on the results obtained, the current study advocates using a combination of HKGs (RPS34 and RHA for developmental tissues and ACT2 and RPS34 for stress-treated tissues) for normalization of qRT-PCR data. The reference genes suggested here will ensure reliability and uniformity in qRT-PCR findings. Further, the expression of any gene is spatiotemporal and is influenced by both internal and external factors. And the genotypic background also plays an important role in it. The key findings of the current study need to be reconfirmed in independent background, including germplasm lines, to get more comprehensive insight. Parallelly, these genes further may also be tested and used in other related Vigna species such as *V. radiata*, *V. unguiculata* and *V. angularis*, which share closer ancestry to *V. mungo*.

Declarations

Author contributions

KK has conceptualized the whole study. PS, PK, SG, AAC, VK, KT performed the experiment and collected the plant tissues. PS, and AD performed the qRT-PCR. MR and KK performed statistical analysis. SKJ, KK, AD, MR, GPD wrote the draft manuscript. MR, GPD and KK edited the manuscript to prepare final draft. MR and GPD supervised and administered the whole study.

Acknowledgements

We acknowledge the support provided by the Director, ICAR-IIPR, Kanpur. The research work carried out in this manuscript is a part of Institute funded project.

Conflict of Interest

There is no conflict of interest, either financial or personal, with other people or organizations that could inappropriately influence or perceived to influence this work.

Data availability statement

The data that support the findings of this study are available in the supporting information of this article

References

1. Singh, N. K. et al. The first draft of the pigeonpea genome sequence. *Journal of plant biochemistry and biotechnology.* **21**:98-112 (2012). https://doi.org/10.1007/s13562-011-0088-8
2. Varshney, R. K. et al. Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nature biotechnology.* 31(**3**):240-246 (2013). https://doi.org/10.1038/nbt.2491
3. Junaid, A. et al. Evolutionary insights into 3D genome organization and epigenetic landscape of Vigna mungo. *Life Science Alliance.* 7(**1**), (2024). https://doi.org/10.26508/lsa.202302074
4. Gaikwad, K. et al. The chromosome-scale genome assembly of cluster bean provides molecular insight into edible gum (galactomannan) biosynthesis family genes. *Scientific Reports.* 13(**1**):9941 (2023). https://doi.org/10.1038/s41598-023-33762-3
5. Ameline-Torregrosa, C. et al. Transcriptomic approaches to unravel plant–pathogen interactions in legumes. *Euphytica.***147**:25-36 (2006). https://doi.org/10.1007/s10681-006-6767-1
6. Garg, R & Jain, M. Transcriptome analyses in legumes: A resource for functional genomics. *The Plant Genome.* 6(**3**) (2013). https://doi.org/10.3835/plantgenome2013.04.0011
7. Abdelrahman, M. et al. Legume genetic resources and transcriptome dynamics under abiotic stress conditions. *Plant, cell & environment.* 41(**9**):1972-1983 (2018). https://doi.org/10.1111/pce.13123
8. Nicot, N. et al. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany.* 56(**421**):2907-2914 (2005). https://doi.org/10.1093/jxb/eri285
9. Udvardi, M. K. Eleven golden rules of quantitative RT-PCR. *The Plant Cell.* 20(**7**):1736-1737 (2008). https://doi.org/10.1105/tpc.108.061143
10. Bustin, S. A. et al. MIQE precis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Molecular Biology.***11**:1-5 (2010). https://doi.org/10.1186/1471-2199-11-74
11. Manu, B. et al. Genomic designing for abiotic stress tolerance in mungbean and urdbean. In Genomic Designing for Abiotic Stress Resistant Pulse Crops (pp. 271-343). Cham: Springer International Publishing (2022). https://doi.org/10.1007/978-3-030-91039-6_8
12. Andersen, C. L. et al. 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(**15**):5245-5250 (2004). https://doi.org/10.1158/0008-5472.CAN-04-0496
13. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**:1-12 (2002). https://doi.org/10.1186/gb-2002-3-7-research0034
14. Pfaffl, M. W. et al. 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). https://doi.org/10.1023/b:bile.0000019559.84305.47
15. Silver, N. et al. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology.***7**:1-9 (2006). https://doi.org/10.1186/1471-2199-7-33
16. Xie, F. et al. RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes. *Functional & Integrative Genomics.* 23(**2**):125 (2023). https://doi.org/10.1007/s10142-023-01055-7
17. Dantas, B. F. et al. Germination, initial growth and cotyledon protein content of bean cultivars under salinity stress. *Revista Brasileira de Sementes.***29**:106-110 (2007). https://doi.org/10.1590/S0101-31222007000200014
18. Roso, R. et al. Effects of Aluminum and pH on Germination of Echium plantagineum L. *Journal of Agricultural Studies.* 8(**1**):313-327 (2020). https://doi.org/10.5296/jas.v8i1.16179
19. Islam, M. R. Evaluation of Polyethylene Glycol (PEG-6000) Induced Drought Stress Tolerant Mungbean Genotypes by using Correlation, Principal Component, Hierarchical Clustering and Multi-Trait Genotype-Ideotypes Distance Index Analysis. *Current Applied Science And Technology.* e0260871-e0260871 (2023). https://doi.org/10.55003/cast.2024.260871
20. Manasa, L. S. et al. Cold tolerance mechanisms in mungbean (Vigna radiata L.) genotypes during germination. *Agriculture.* 13(**2**):315 (2023). https://doi.org/10.3390/agriculture13020315
21. Kumar, K. et al. Identification and characterization of NHX gene family for their role under salt stress in Vigna mungo. *Physiologia Plantarum.* 176(**5**):e14563 (2024). https://doi.org/10.1111/ppl.14563
22. Marchler-Bauer, A. et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research.* 45(**D1**), D200-D203 (2017). https://doi.org/10.1093/nar/gkw1129
23. Kumar, S. et al. Gene expression and biochemical profiling of contrasting Vigna mungo genotypes against Mungbean Yellow Mosaic India Virus (MYMIV). *Journal of Food Legumes.* 35(**2**):107-116 (2022).

24. Kumar, K. et al. Single trait versus principal component based association analysis for flowering related traits in pigeonpea. *Scientific Reports.* 12(**1**):10453 (2022). https://doi.org/10.1038/s41598-022-14568-1
25. Vijayakumar S & Sakuntala M. Validation of reference gene stability for normalization of RT-qPCR in Phytophthora capsici Leonian during its interaction with Piper nigrum L. *Scientific Reports.* 14(1):7331 (2024). https://doi.org/10.1038/s41598-024-58139-y
26. Nolan, T. Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nature Protocols.* 1(**3**):1559-1582 (2006). https://doi.org/10.1038/nprot.2006.236
27. Hu, R. et al. Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Molecular Biology.* 10:1-12 (2009). https://doi.org/10.1186/1471-2199-10-93
28. de Vega-Bartol, J. J. et al. Normalizing gene expression by quantitative PCR during somatic embryogenesis in two representative conifer species: Pinus pinaster and Picea abies. *Plant Cell Reports*. 32:715-729 (2013). https://doi.org/10.1007/s00299-013-1407-4
29. Taylor, C. M. et al. Identifying stable reference genes for qRT-PCR normalisation in gene expression studies of narrow-leafed lupin (*Lupinus angustifolius* L.). *PLoS One.* 11(2):e0148300 (2016). https://doi.org/10.1371/journal.pone.0148300
30. Chi, C. et al. Selection and validation of reference genes for gene expression analysis in Vigna angularis using quantitative real-time RT-PCR. *PloS one.* 11(12):e0168479 (2016). https://doi.org/10.1371/journal.pone.0168479
31. Sinha, R. Sharma, T. R., Singh, A. K. Validation of reference genes for qRT-PCR data normalisation in lentil (Lens culinaris) under leaf developmental stages and abiotic stresses. *Physiology and Molecular Biology of Plants.* 25(**1**):123-134 (2019). https://doi.org/10.1007/s12298-018-0609-1
32. Gutierrez, N. et al. Assessment of candidate reference genes for expression studies in Vicia faba L. by real-time quantitative PCR. *Molecular Breeding.***28**:13-24 (2011). https://doi.org/10.1007/s11032-010-9456-7
33. Reddy, D. S. et al. Identification and validation of reference genes and their impact on normalized gene expression studies across cultivated and wild Cicer species. *PloS one.* 11(2):e0148451 (2016). https://doi.org/10.1371/journal.pone.0148451
34. Kundu, A. et al. Complex molecular mechanisms underlying MYMIV-resistance in Vigna mungo revealed by comparative transcriptome profiling. *Scientific Reports.* 9(**1**):8858 (2019). https://doi.org/10.1038/s41598-019-45383-w
35. Kumar J & Kumar V. Ribosome proteins—Their balanced production. In: Emerging Concepts in Ribosome Structure, Biogenesis, and Function (pp. 47-87). Academic Press (2021). https://doi.org/10.1016/B978-0-12-816364-1.00003-2
36. Bourgeois, C. F., Mortreux, F. & Auboeuf, D. The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nature Reviews Molecular Cell Biology.* 17(**7**):426-438 (2016). https://doi.org/10.1038/nrm.2016.50

Tables

Table 1: List of genes used in the current study and their function

S.no.	Arabidopsis gene	Gene name	Vigna mungo ortholog used in the current study	Percent homology with Arabidopsis homolog	Function
1.	AT1G30950.1	Unusual Floral Organ (UFO)	WVZ23601.1	66.743	Encodes Unusual floral organ genes which is required for the proper identity of the floral meristem and is constitutively expressed in the floral tissues
2.	AT5G62690.1	TUBULIN BETA CHAIN 2 (TUB2)	WVY95928.1	95.778	Encodes tubulin beta-2/beta-3 chain, which are components of microtubules in the cytoskeleton and play essential roles in various cellular processes
3.	AT3G18780.2	ACTIN 2 (ACT2)	WVY99582.1	92.042	Encodes an actin that is constitutively expressed in vegetative structures
4.	AT1G58050.1	RNA helicase (RHA)	WVZ12136.1	47.337	Encodes RNA helicase that unwind RNA secondary structures and/or displace RNA-binding proteins in an ATP-dependent manner, that is essential for various cellular processes involving RNA
5.	AT1G13340.1	ISTL6	WVZ09401.1	51.507	ISTL6 gene is crucial for the proper trafficking and degradation of membrane proteins within the cell. Proper functioning of this pathway is vital for maintaining cellular homeostasis and responding to environmental stressors
6.	AT2G36130.1	Cyclophilin (CyP)	WVZ05807.1	87.805	Cyclophilins are a family of highly conserved proteins with peptidyl-prolyl cis-trans isomerase (PPIase) activity, which catalyzes the isomerization of peptide bonds at proline residues that is crucial for protein folding, and function
7.	AT2G20560.1	DNAJ	WVZ01865.1	79.769	DNAJ heat shock family protein are molecular chaperones involved in various cellular processes and play essential roles in protein folding, stabilization, degradation, and trafficking
8.	AT5G38410.3	Rubisco small subunit 3b (RBCS3B)	WVZ13955.1	67.935	Encodes a member of the Rubisco small subunit (RBCS) multi-gene family
9.	AT5G58990.1	28S ribosomal S34 protein (RPS34)	WVZ02079.1	84.286	(RPS34) is a component of the small subunit of the ribosome in eukaryotes that are responsible for protein synthesis, and their subunits (small and large) work together to translate mRNA into polypeptides
10.	AT4G27320.1	Universal Stress Protein 21 (USP21)	WVZ04305.1	70.892	Protein that contains a universal stress protein domain. They are essential for plant stress tolerance as well as cellular homeostasis
11.	AT5G51170.1	U6 snRNA phosphodiesterase like protein (USB1)	WVZ06716.1	65.351	Encodes U6 snRNA phosphodiesterase-like protein, that are involved in RNA processing, specifically within the spliceosome machinery, where they contribute to pre-mRNA splicing
12.	AT1G07920.1	Elongation Factor-TU (EfTu)	WVY95914.1	96.197	GTP binding Elongation factor Tu family protein
13.	AT4G05320.2	Ubiquitin 10 (UBQ10)	WVZ24872.1	99.738	Encodes Ubiquitin 10 that is a member of the ubiquitin protein family, which plays a central role in post-translational modification of proteins through ubiquitination
14.	AT4G27960.2	Ubiquitin Conjugating Enzyme 9 (UBQE9)	WVY93904.1	94.667	Encodes UBIQUITIN CONJUGATING ENZYME 9 is a key enzyme in the ubiquitination pathway, primarily involved in SUMOylation

Table 2: Gene stability ranking of the 14 HKGs used in the current study through different algorithms using 17 developmental tissues

Ranking Order (Better–Good–Average)														
Method	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Delta CT	RPS34	RHA	CyP	ACT2	DNAJ	UBQC9	ISTL6	UBQ10	USB1	TUB2	USP21	EfTu	RBCS3B	UFO
BestKeeper	USB1	DNAJ	RHA	ACT2	UFO	CyP	RPS34	UBQC9	UBQ10	ISTL6	TUB2	USP21	RBCS3B	EfTu
NormFinder	RPS34	ACT2	DNAJ	CyP	RHA	UBQC9	ISTL6	UBQ10	USB1	TUB2	USP21	EfTu	RBCS3B	UFO
geNorm	RPS34	RHA	CyP	ACT2	DNAJ	ISTL6	UBQC9	UBQ10	USB1	TUB2	USP21	EfTu	RBCS3B	UFO
Comprehensive ranking	RPS34	RHA	ACT2	DNAJ	CyP	USB1	UBQC9	ISTL6	UBQ10	TUB2	UFO	USP21	EfTu	RBCS3

Figures

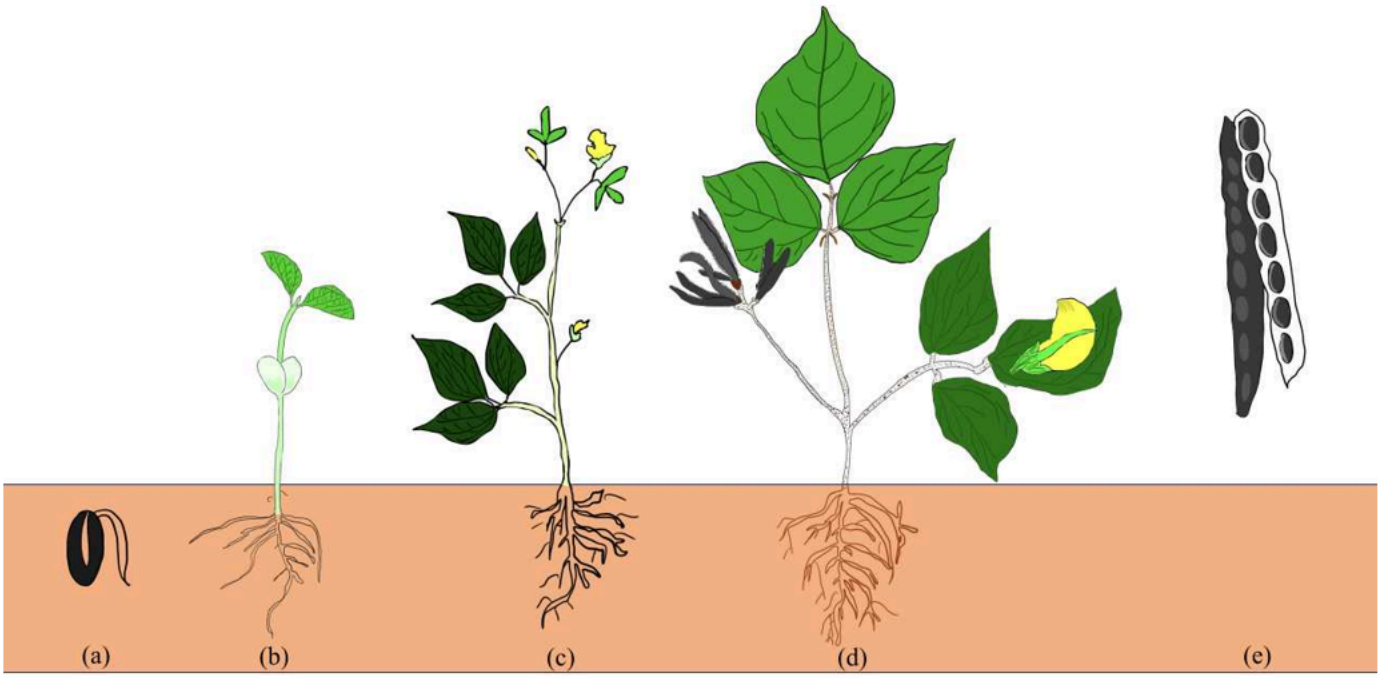

Figure 1

Tissues used in the identification of best HKGs for normalization of qRT-PCR data in blackgram (a) germination stage from which plumule, radicle and cotyledon tissues were collected at 5 days after sowing; (b) Vegetative stage from which vegetative leaves, meristem, stem, and roots tissues were collected at 30 days after sowing; (c) Reproductive stage from which reproductive buds, reproductive leaves, roots, and stem tissues were collected at 45 days after sowing; (d) Podding stage from which leaves, whole pod, stem and root tissues were collected at 60 days after sowing and (e) pods were split into pod wall and seed.

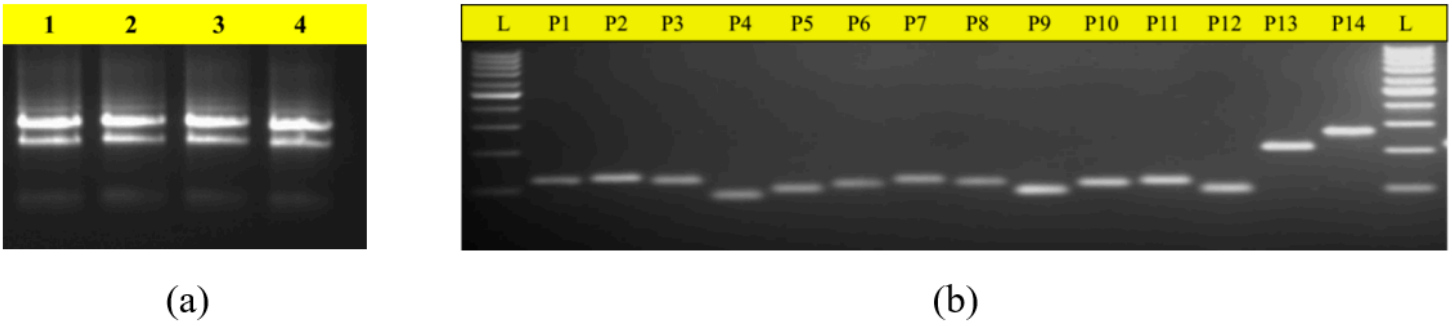

Figure 2

The agarose gel displaying the quality of the RNA isolated from the selected tissues (a), and the specific amplicon of the 14 housekeeping genes confirmed by their respective band sizes. Where P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, and L stands for UBQE9, TUB2, USP21, RPS34, USB1, RHA, ISTL6, EfTu, ACT2, DNAJ, CyP, RBCS3B, UFO, UBQ10 and 100 base pair Ladder, respectively

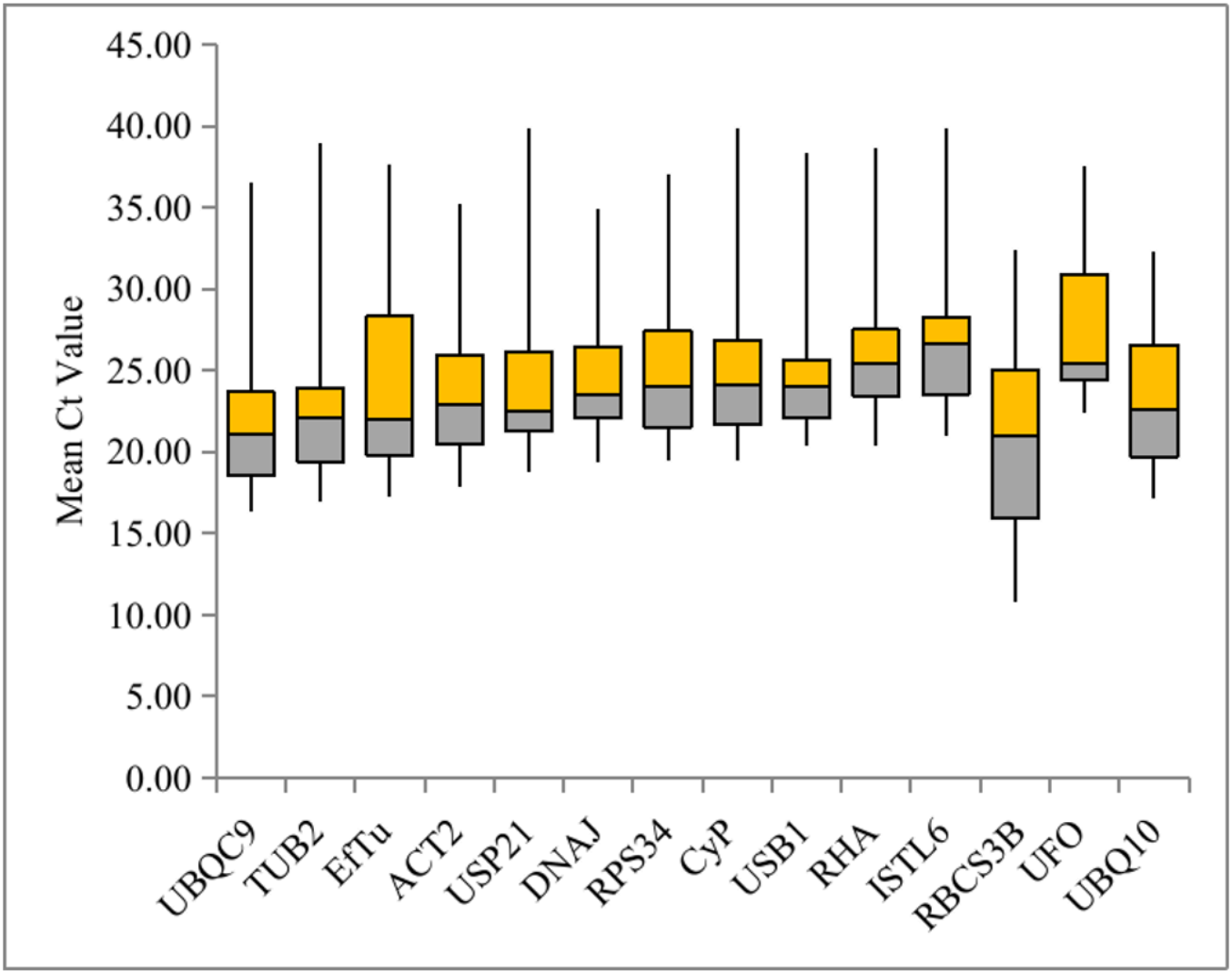

Figure 3

Variation in Ct values of all 14 HKGs considered for the current study across all the tissues. The central line within the box represents the mean value. The gray and yellow boxes represents the quartile where the Ct values for most of the HKGs falls. The middle line represents the middle (median) Ct value in the dataset. The larger vertical lines above the yellow and below gray whiskers represent the presence of outliers in the Ct values in some tissues.

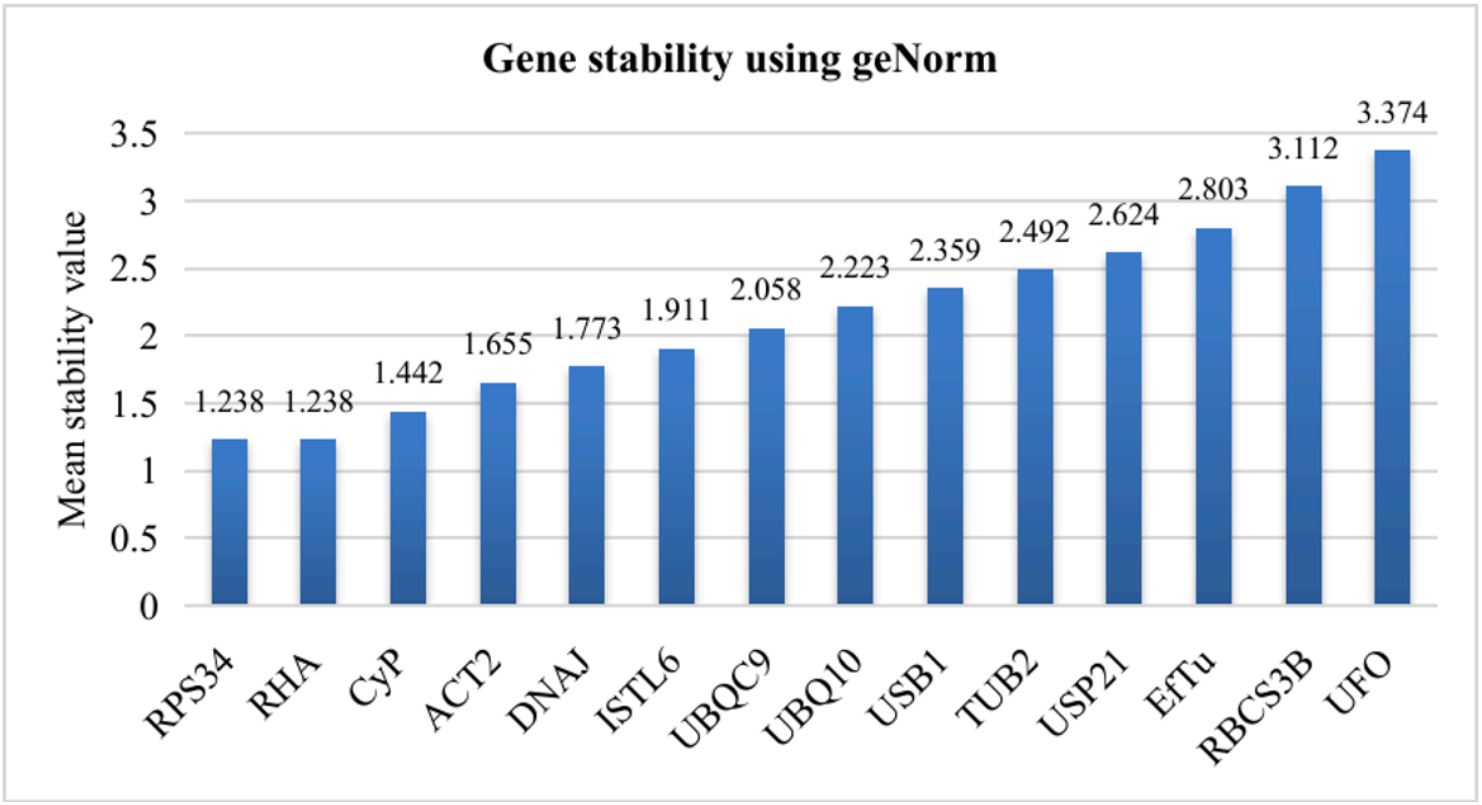

Figure 4

Gene stability analysis using the geNorm algorithm identifies RPS34 and RHA as the most suitable while RBCS3B and UFO as the least suitable for normalization of qRT-PCR data across different tissues.

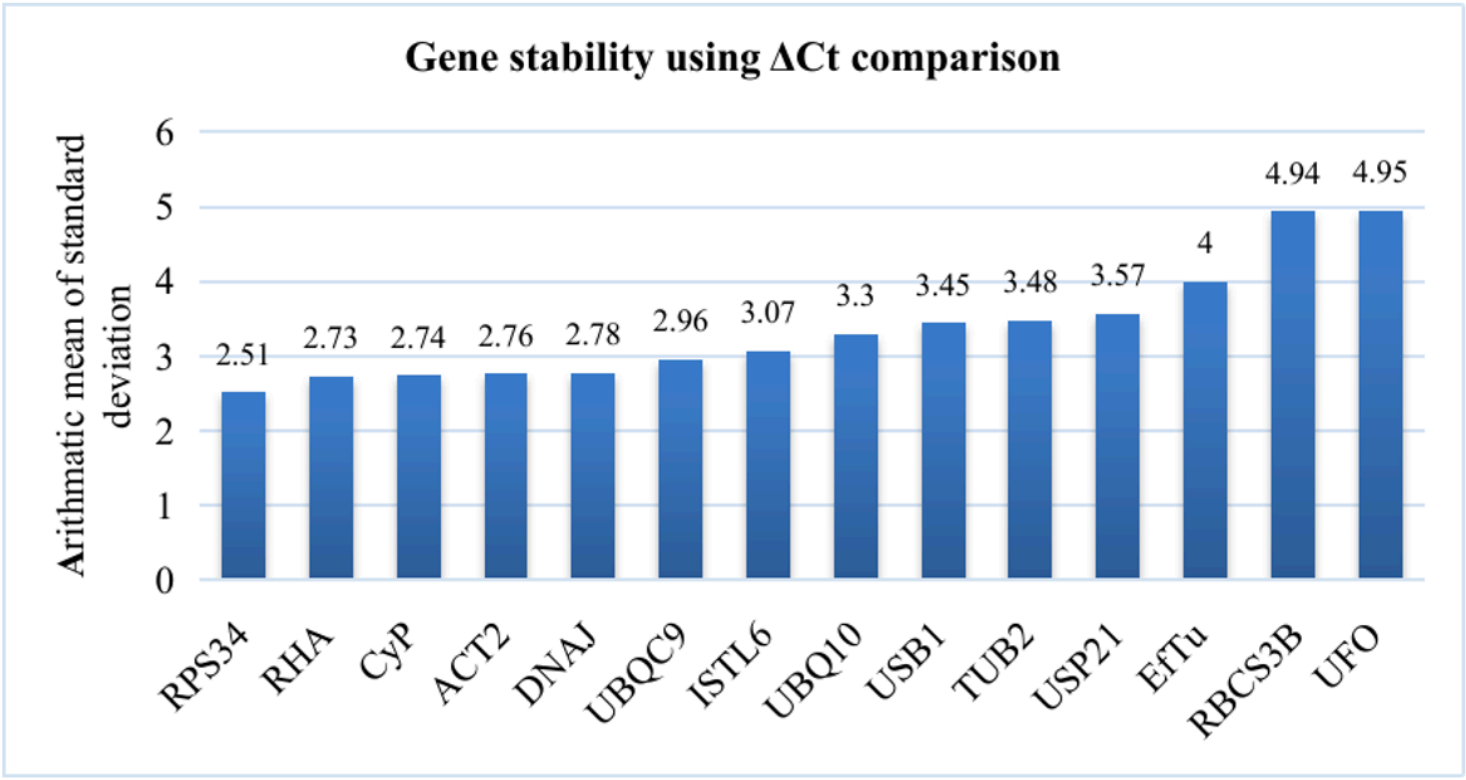

Figure 5

The ΔCt comparison method for gene stability analysis suggests RPS34 and RHA as the most suitable while RBCS3B and UFO as the least suitable for normalization of qRT-PCR data across different tissues.

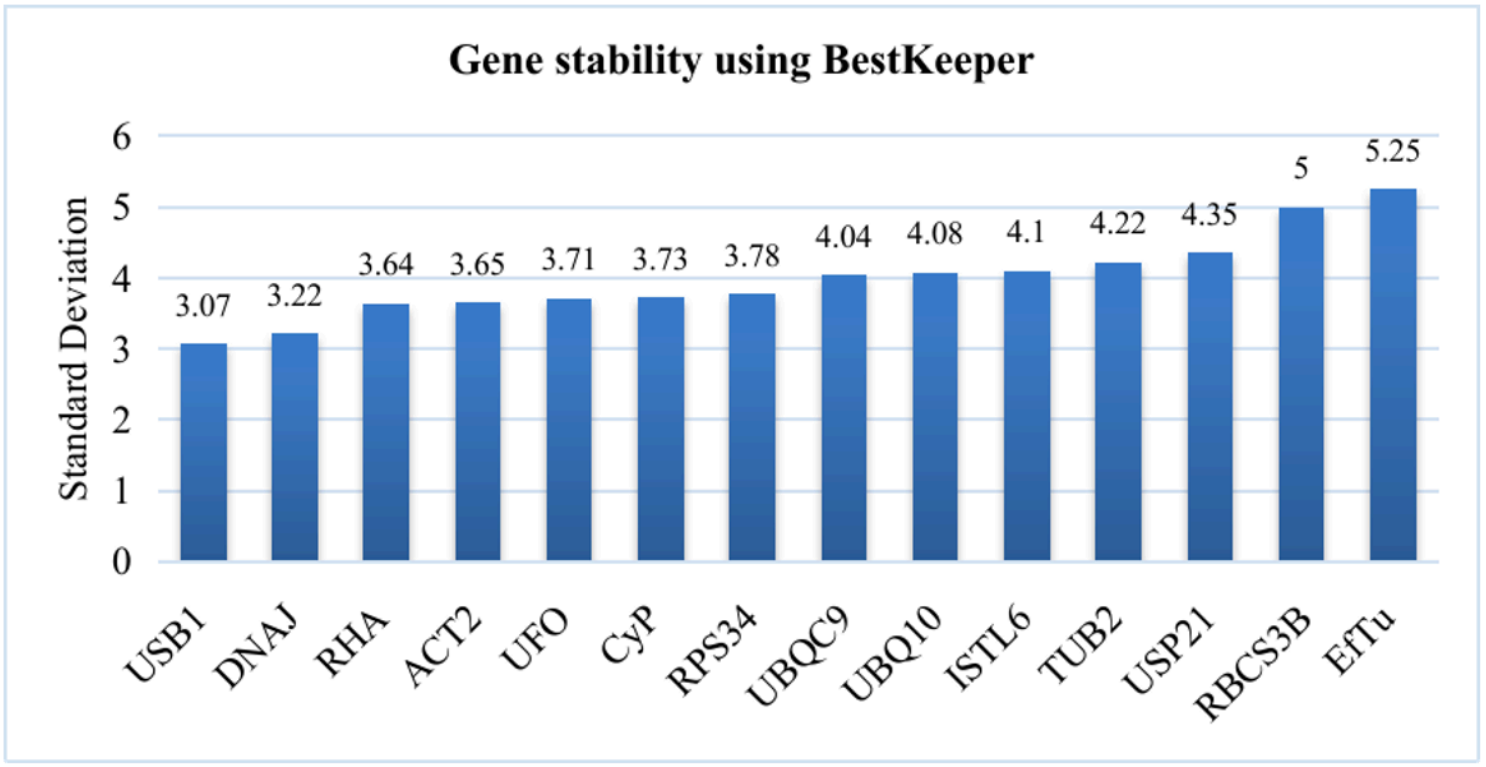

Figure 6

The BestKeeper algorithm for gene stability analysis identifies USB1 and DNAJ as the most suitable while RBCS3B and EfTu as least suitable for normalization of qRT-PCR data across different tissues.

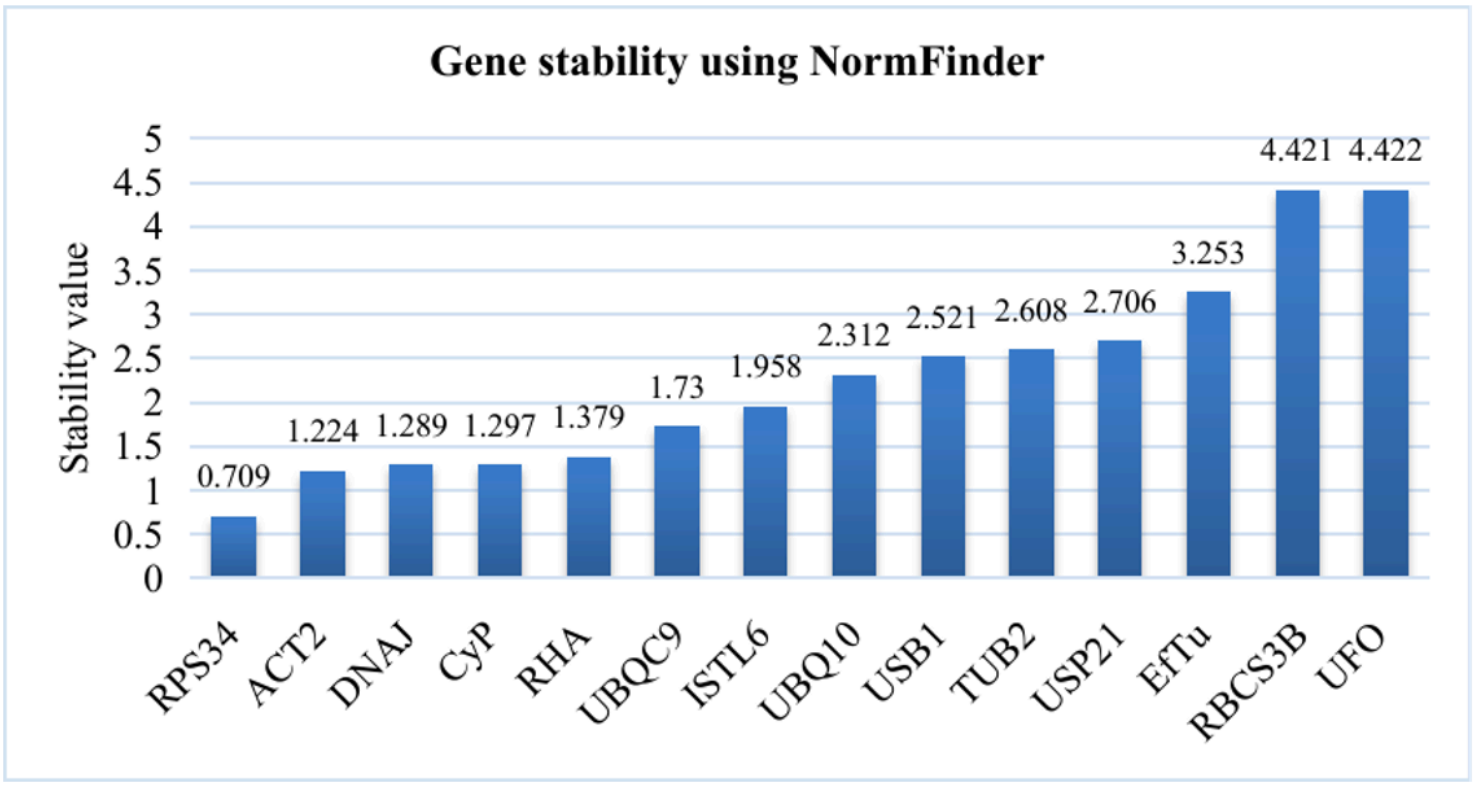

Figure 7

Gene stability analysis using the Normfinder algorithm infers RPS34 and ACT2 to be most suitable while RBCS3B and UFO as least suitable for normalization of qRT-PCR data across different tissues

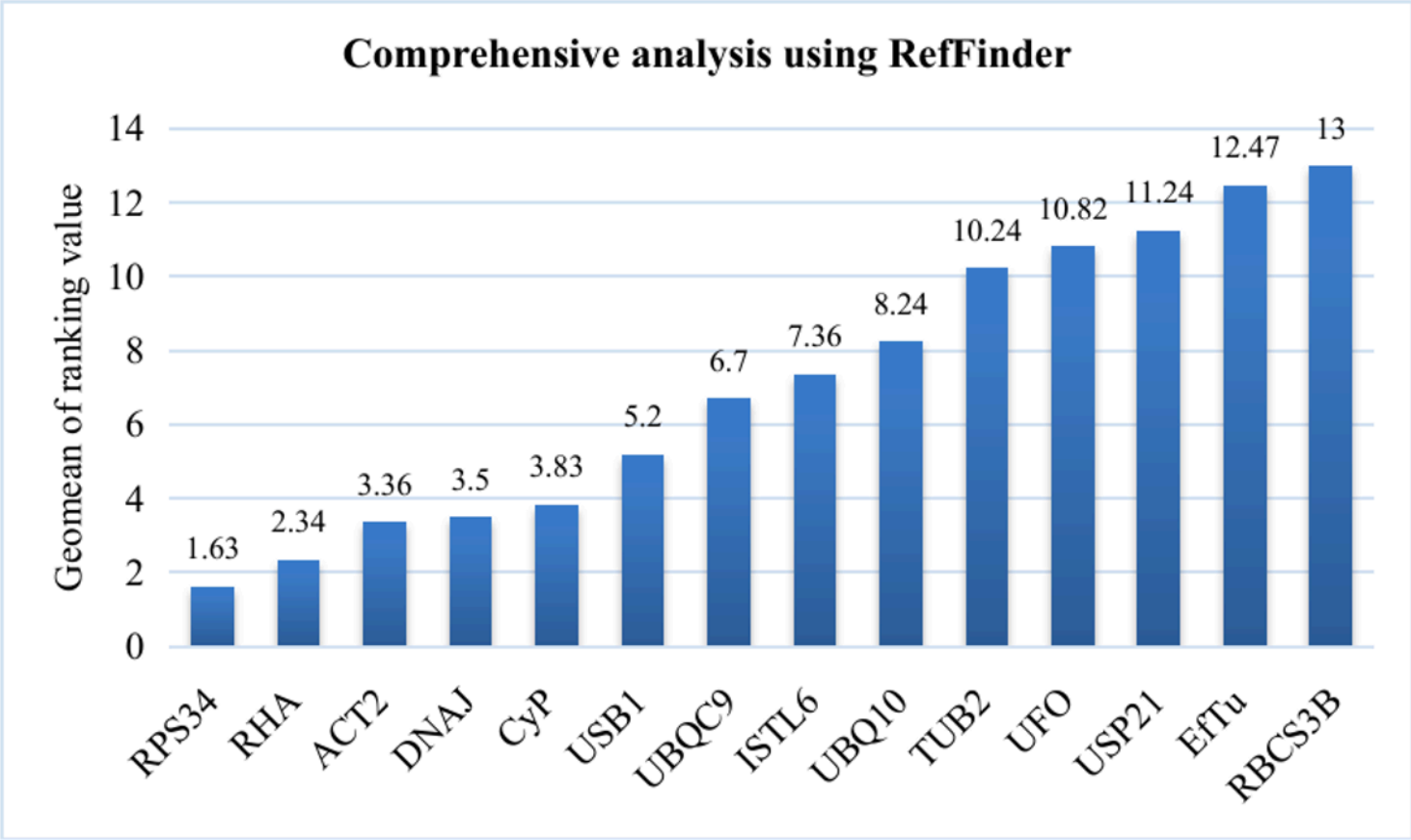

Figure 8

Gene stability analysis using the RefFinder algorithm infers RPS34 and RHA to be most suitable while EfTu and RBCS3B as least suitable for normalization of qRT-PCR data across different tissues

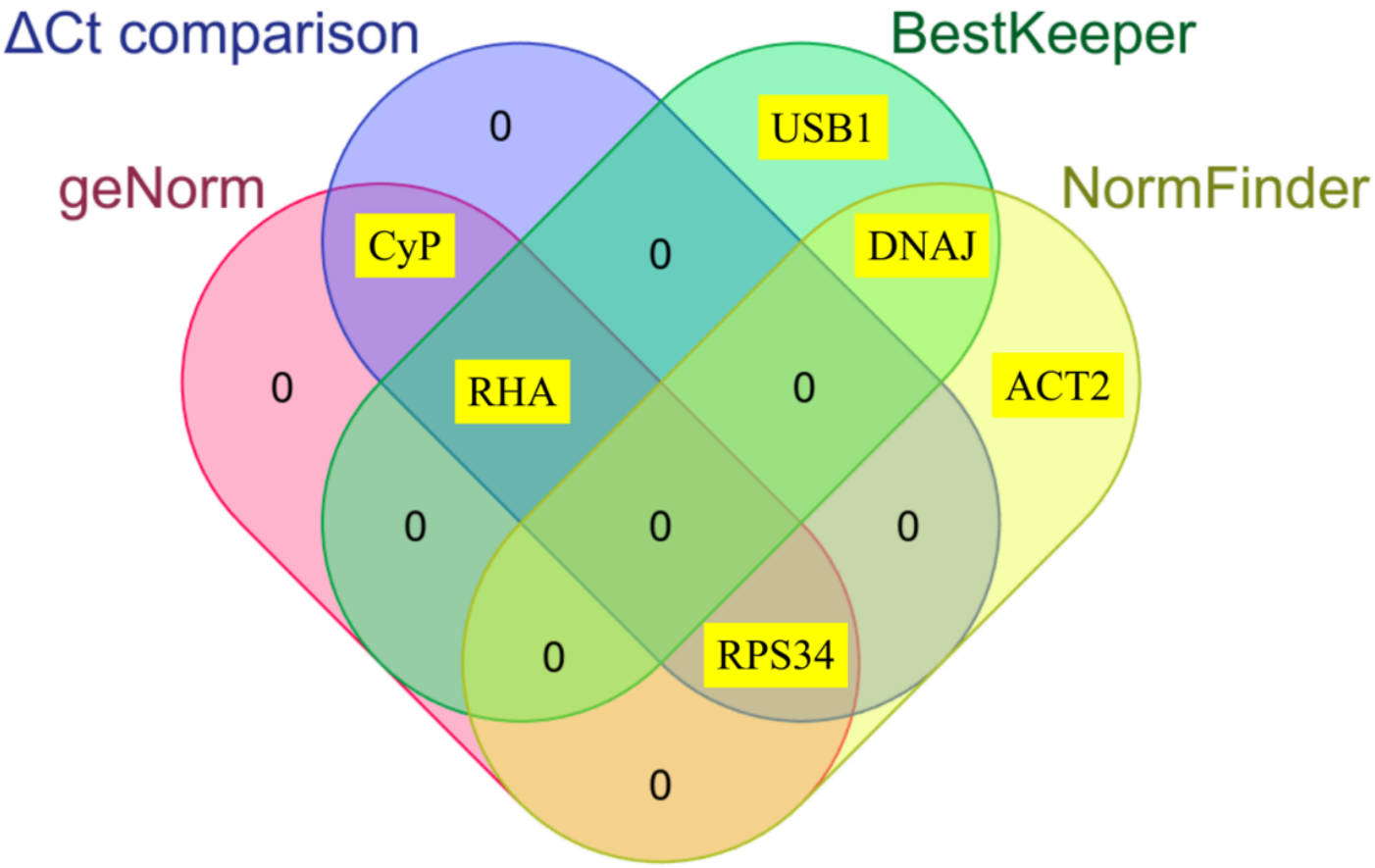

Figure 9

Consistency in the stability of the HKGs using multiple algorithms explained via Venn diagram. The RHA (ΔCt comparison, geNorm and BestKeeper) and RPS34 (BestKeeper, NormFinder and geNorm) genes were found to be most suitable via three algorithms whereas, CyP (ΔCt comparison and geNorm) and DNAJ (NormFinder and BestKeeper) were identified consistently via two algorithms.

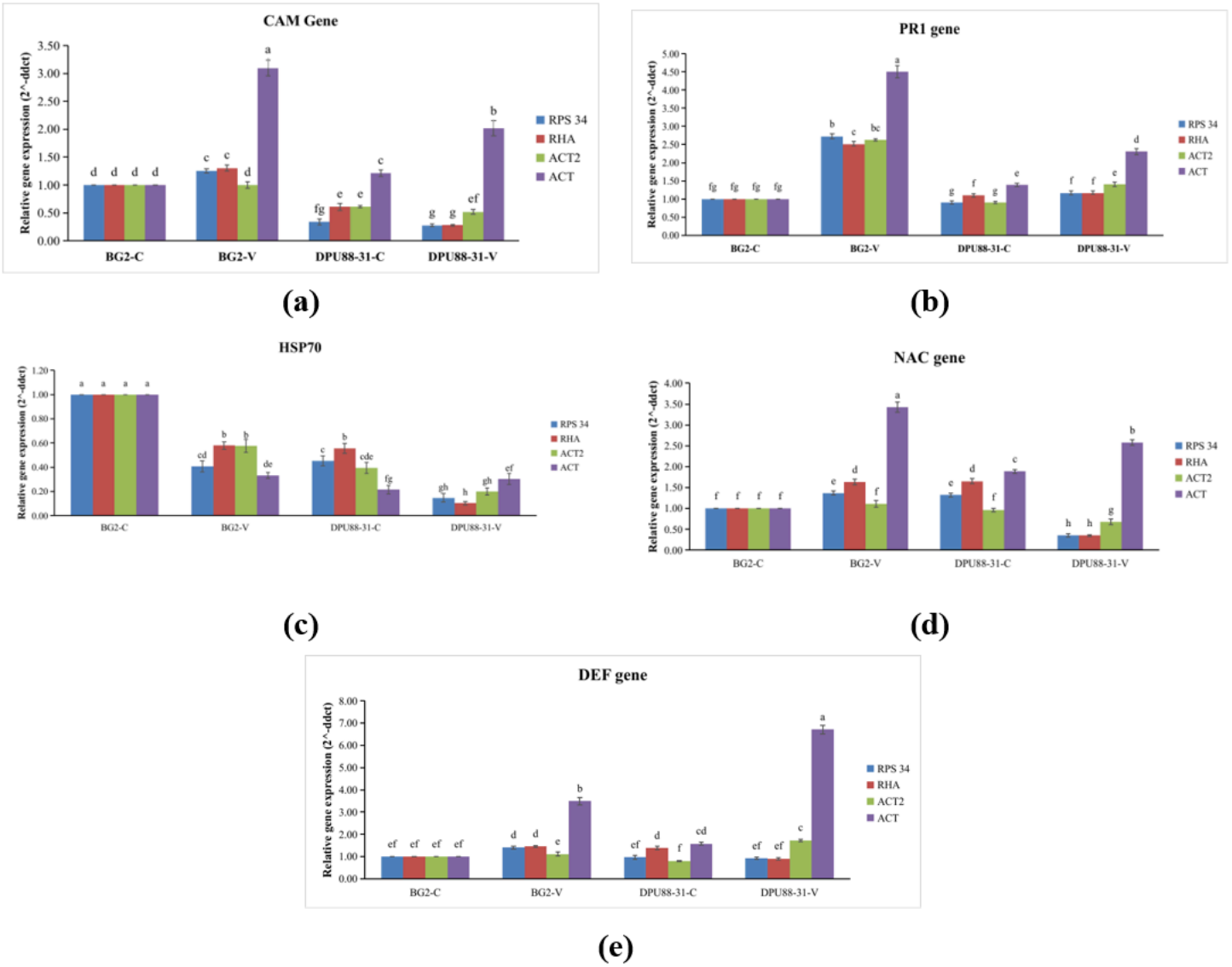

Figure 10

Relative expression level of 5 test gene i.e. CAM (a), PR1 (b), HSP70 (c), NAC (d) and DEF (e) when normalized with the RPS34, PHS and ACT2 along with the external control ACT. The relative expression level clearly reflect the consistency in the expression pattern of the test genes when normalized with the HKGS identified in the current study. A larger deviation can be seen in the expression pattern of the test genes when normalized with external control ACT gene.

Supplementary Files

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

- 6Supplementarytables.docx
- 5Supplementaryfigures.docx