



## Quantitative real-time PCR based evaluation and validation of reference genes in *Gossypium arboreum*

RAGHAVENDRA K P<sup>1</sup>, RAKESH KUMAR<sup>2</sup>, JOY DAS<sup>3</sup>, SANTOSH H B<sup>4</sup>, SACHIN A MORE<sup>5</sup>, RAMAKRISHNA N<sup>6</sup>, SHILPA G CHAWLA<sup>7</sup>, SANDHYA KRANTHI<sup>8</sup> and KESHAV RAJ KRANTHI<sup>9</sup>

ICAR-Central Institute for Cotton Research, Nagpur, Maharashtra 440 010, India

Received: 09 August 2017; Accepted: 04 July 2019

### ABSTRACT

Estimation of gene expression levels plays a crucial role in understanding the function of the target gene(s). Inter-sample variance in gene expression can be more precisely measured if transcripts levels are accurately normalized. Normalization is pre-requisite step prior to the determination of candidate gene expression by qPCR. In this study conducted at ICAR-Central Institute for Cotton Research, Nagpur during 2015–16, six candidate reference genes, viz. actin4 (ACT4), actin7 (ACT7), RNA Helicase (RNAH), Serine/threonine-protein phosphatase PP2A-1 (PP2A1), ubiquitin7 (UBQ7) and  $\alpha$  tubulin ( $\alpha$ TUB) were systematically analysed for their expression patterns in different tissues pertaining to three development stages of cotton namely seedling, early reproductive and fiber development. The study has identified actin-4/actin-7/ubiquitin-7 as the most ideal reference genes for fiber development stages whereas actin-4/ ubiquitin-7 and actin-7/RNA helicases for seedling and early reproductive development stages, respectively. Validation of identified reference genes for relative expression analysis of Gacobl9, a COBRA-like protein, demonstrated their usefulness in qPCR analysis in *Gossypium arboreum*.

**Key words:** Cotton, Gene Expression, *Gossypium arboreum*, Reference gene, qPCR

Cotton, popularly known as white Gold, is one of the world's richest source of natural fiber and also one of the most important cash crops belonging to genus *Gossypium* (Abd El-Moghny *et al.* 2017). Besides being a major source of raw material for textile industry, having a single elongated epidermal cell, it acts as a molecular reference system for studying cellular differentiation and cell patterning in plants (Kim and Triplett 2001). In the era of functional genomics, the study of gene expression profile plays a crucial role in understanding the function of the candidate gene(s) and to decipher complex gene regulatory network in biological systems (Vandesompele *et al.* 2002). Gene chip array, RNA blot, quantitative real-time PCR (qPCR), semi-qPCR, etc. are being routinely used by molecular biologists for quantitative estimation of transcript levels in biological

samples. However, qPCR exhibiting higher specificity and sensitivity with low sample volume requirement and amenability for high throughput applications remains undoubtedly the preferred and most widely used method (Czechowski *et al.* 2005). qPCR is also the most favoured choice for validation of large biological data generated through microarray, RNA-seq, and genotyping platforms (Jain *et al.* 2006). Despite being a robust and powerful technique, qPCR mediated gene expression profiling lacks accuracy if reliable and suitable reference genes are not used (Huggett *et al.* 2005). Normalization improves the precision of gene expression data by circumventing the limitations raised due to poor RNA quality and variability in reverse transcription & PCR efficiencies (Ermei *et al.* 2012). Ideally, an internal control must show invariable transcript levels both spatially and temporal across the experimental set-up (Marum *et al.* 2012). Housekeeping genes are generally considered to have constitutive expression as they are playing very crucial role in maintaining basic cells' surviving activities. However, several scientific studies have reported that these genes also showed variation in their expression may be due to their multiple roles in other cell's activities besides housekeeping functions (Czechowski *et al.* 2005). In addition to this, identified reference genes for one crop species may express inconsistently in another crop species even under identical experimental setup, thus negating their recurrent broad-spectrum use (Jain *et al.* 2006, Jian *et al.*

<sup>1</sup>Senior Scientist (kpraghavendra@gmail.com), <sup>2,3,4</sup>Scientist (rakesh.kumar8@icar.gov.in, Joy.Das@icar.gov.in, santoshgpb@gmail.com), <sup>5,6,7</sup>Senior Research Fellow (samknw2008@gmail.com, nallateegalaramakrishna@gmail.com, chawlashilpa11@gmail.com), <sup>8</sup>Principal Scientist (sandhya.kranthi@gmail.com) ICAR-Central Institute for Cotton Research (CICR), Shankar Nagar Post Office, Nagpur, Maharashtra 440 010, India; <sup>9</sup>Head, Technical Information Section (krkranthi@gmail.com), International Cotton Advisory Committee, Washington DC, USA.

2008). Reference genes have been systematically evaluated and validated for target gene expression normalization by using qPCR and statistical programs in many plant species under varied experimental conditions and developmental stages (Jian *et al.* 2008, Nicot *et al.* 2005, Reid *et al.* 2006, Hu *et al.* 2009). In cotton, scientific reports pertaining to selection and validation of reference genes are very limited and restricted with only tetraploid cotton *Gossypium hirsutum* (Tu *et al.* 2007, Artico *et al.* 2010, Wang *et al.* 2013, Raghavendra *et al.* 2014, Fausto *et al.* 2017). Thus, identified stably expressed and validated internal control genes in *Gossypium hirsutum* are generally in use for target gene expression analysis in *Gossypium arboreum*. Owing to the genomic complexity of the *Gossypium* spp. and availability of whole genome sequencing data containing uncharacterized gene sequences; it will be of immense value to systematic evaluation and validation of the reference genes in *G. arboreum*.

Through this study, six potential candidate internal control genes from *Gossypium arboreum* were statistically analyzed for their transcript expression stability to identify and validate the most suitable ones for target gene normalization by qPCR.

#### MATERIALS AND METHODS

**Plant material:** Total 11 samples each with three biological replicas were sourced from three different developmental stages (Seedling development, early reproductive stage and fiber developmental stage) of *G. arboreum* cv. Roja grown at ICAR-CICR, Nagpur during *kharif* 2015–16. Seedling development stage comprises of hypocotyl, cotyledonary leaves, and root tissues from 15 days old seedling raised on MS medium; early reproductive stage; consisted of sepals, petals, anthers and ovary from flowers at 0 Days Post Anthesis (DPA) and square at -3 DPA and samples set from fiber developmental stage comprising ovule with fibers at 3, 10 and 25 DPA, from plants grown

in the greenhouse.

**RNA isolation and cDNA synthesis:** Total RNA was extracted using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich St. Louis, MO, USA) by following the manufacturer's instructions. Traces of genomic DNA were eliminated by On-Column DNaseI digestion set (Sigma-Aldrich St. Louis, MO, USA). Total RNA was subjected to quality and quantity analysis using agarose gel electrophoresis and Qubit2.0 fluorometer (Invitrogen, USA) respectively. A total of 1 µg of total RNA was used for first-strand cDNA synthesis using Superscript III Reverse Transcriptase kit (Life Technologies, USA).

**Primer design and qPCR analysis:** qPCR primers for selected six candidate housekeeping genes and *Gacobl9* were synthesized from Integrated DNA Technologies (IDT) online tool (Supplementary Table 1). qPCR's were performed using a Mx 3005 Real-Time PCR System (Agilent Technologies). Each qPCR reaction comprised a final volume of 25 µl containing 2 µl of 1:10 diluted cDNA, 0.5 µl of each forward and reverse primer (0.4 µM), and 12.5 µl of 2x 2× Brilliant II SYBR Green PCR Master Mix, 0.375 µl of diluted reference dye and 11.125 µl nuclease free water (Agilent Technologies). The following thermal cycling profile was used for all qPCRs: initial hot start at 95°C for 10 min, 35 cycles of 95°C for 30 sec, 60°C for 45 sec and 72°C for 30 seconds and one segment of disassociation/melt curve. Reliability of qPCR run was confirmed by including non-template controls along with total six replication comprising three biological and two technical replications per sample.

**Reference genes transcript data analysis:** The reference genes relative expression data were systematically evaluated with geNorm v 3.0, NormFinder and BestKeeper programs (Vandesompele *et al.* 2002, Andersen *et al.* 2004, Pfaffl *et al.* 2004). Each software is programmed with a unique algorithm and calculates gene stability measure by taking threshold cycle values (Cq values) either in raw or in normalized form using a delta Ct method. The geNorm program ranks

Supplementary Table 1 List of candidate reference genes, their primer sequences selected for evaluation of expression stability in cotton (*Gossypium arboreum* cv. Roja)

| Reference gene | Accession number | Primer sequence                                              |
|----------------|------------------|--------------------------------------------------------------|
| RNAH           | XM_016862161.1   | F: GCCTGATCGATATGTGGAGGGAT<br>R: CAGGAAGGTTTGGCCATCTTGGA     |
| UBQ7           | DQ116441.1       | F: AGACAAGGAAGGCATTCCACCTGA<br>R: GCTTGATCTTCTTGGGCTTGGTGT   |
| ACT7           | NM_001327049.1   | F: TACAACGAACCTCCGTGTTGCTCCT<br>R: TCACCGGAATCCAGCACAAATACCT |
| αTUB           | EF151302.1       | F: GCTGAGAAGGCTTACCATGAGCA<br>R: GGTACATCAAACAGCAAGCCA       |
| PP2A1          | DT545658.1       | F: GATCCTTGTGGAGGAGTGGA<br>R: GCGAAACAGTTCGACGAGAT           |
| ACT4           | AY305726.1       | F: TTGCAGACCGTATGAGCAAG<br>R: ATCCTCCGATCCAGACACTG           |
| Gacobl9        | Cotton_A_03290*  | F: CTGCCACCACCTGATACTTAC<br>R: GGCTGTCAAACCGGAGAATA          |

\*Niu *et al.* 2015

the reference genes in an order of expression stability index-M value (lower the M value more the stability of expressed gene) derived based on repeated pairwise variance analysis. Besides this, a statistical parameter known as  $V_n/V_{n+1}$  variance is also calculated to identify the number of potential genes for more accurate expression normalization. The inclusion of additional gene becomes statistically insignificant if  $V_n/V_{n+1}$  variance scores below 0.15 cutoff. NormFinder algorithm workout both inter and intra-group transcript variance using repeated pairwise analysis by eliminating pseudo matched expression of candidate gene pair(s). Therefore, NormFinder analysis not only specifies the best individual internal control but also a pair of genes which may fall short in competence if analyzed individually. On the contrary, BestKeeper algorithm takes raw Ct values as input data and performs repeated correlation and regression analysis to find suitable reference gene(s).

## RESULTS AND DISCUSSION

**Comparative expression data of reference genes:** The data for (quantification cycle value) Cq values depicting the initial transcript abundance were collected for each reference genes (Supplementary Table 2, Fig 1.) Cq values vary from 20.66 to 24.46, where *ACT4* showed highest transcript abundance (mean Cq value=20.66) across all samples followed by *αTUB*, *UBQ7*, *ACT7*, and *RNAH*, whereas *PP2A1* gene showed lowest transcript levels (mean Cq value= 24.46). Interestingly, *ACT7* gene showed relatively stable expression profile with SD value 1.18 followed by 1.31, 1.64, 1.66, 1.84 and 2.91 for *PP2A1*, *UBQ7*, *ACT4*, *RNAH*, and *αTUB*, respectively.

**geNorm analysis:** geNorm analysis identified *ACT4/UBQ7* as most stable internal control gene pair (0.519) during the seedling developmental stage followed by *ACT7*, *PP2A1*, *RNAH* and *αTUB* (Fig. 2a). Similarly, *ACT4/ACT7* (1.37) and *UBQ7/ACT7* (0.64) were the most promising candidates for early reproductive and fiber development stages, respectively (Fig 2b, 2c). The results from the analysis also show that pairwise  $V_n/V_{n+1}$  variance is above 0.15 in all samples across the three developmental stages. It indicates that inclusion of next stably expressed genes are found to contribute significantly to the normalization of gene expression (Fig 3). The results from geNorm are

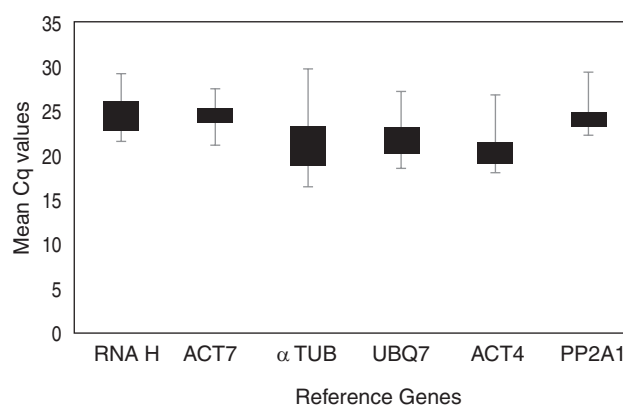


Fig 1 Expression levels (Mean Cq Values) of six candidate reference genes across the three developmental stages of *Gossypium arboreum* cultivar Roja. The boxes represent mean Cq values and bars correspond to the standard deviation (SD).

summarised in Supplementary Table 3 and Fig 2.

**NormFinder analysis:** Using NormFinder, *ACT4* (0.139) emerge out to be the promising internal control gene in seedling and early reproductive stage samples, whereas *PP2A1* (0.192) followed by *ACT4* (0.312) in fiber development stages respectively (Supplementary Table 4).

However, comparative analysis between geNorm and NormFinder softwares shows minor differences in normalization suitability ranking of reference genes (Supplementary Table 5).

**BestKeeper analysis:** Using BestKeeper software, three (*UBQ7*, *ACT4*, and *αTUB*), four (*ACT7*, *RNAH*, *ACT4* and *αTUB*) and two (*PP2A1* and *αTUB*) candidate genes showed significant correlation at p-value <0.001 with the BestKeeper Index in samples of seedling, early reproductive and fiber development stages respectively (Supplementary Table 6). SD data showed that *UBQ7* (1.31) is the most consistently expressed internal control gene among samples of seedling developmental stages. Similarly, *ACT7* (1.1) and *ACT4* (0.83) were found suitable for early reproductive and fiber development stages respectively.

**Validation of reference genes:** *Gacobl9*, a cotton COBRA-like gene was chosen as a target gene for the purpose to validate the qPCR normalization potential

Supplementary Table 2 Expression levels of six reference genes in ten biological samples across three experimental sets of cotton (*Gossypium arboreum* Cv. Roja)

| Reference genes | Seedling stage mean Cq±SD* | Early reproductive stage mean Cq ±SD | Fiber development stage mean Cq ±SD | Across the developmental stages mean Cq ±SD |
|-----------------|----------------------------|--------------------------------------|-------------------------------------|---------------------------------------------|
| RNAH            | 22.64±0.74                 | 24.9±1.71                            | 24.7±1.89                           | 24.3±1.84                                   |
| ACT7            | 24.66±0.92                 | 24.6±1.1                             | 22.69±0.92                          | 24.16±1.18                                  |
| ATUB            | 21.38±2.34                 | 21.67±3.4                            | 18.72±1.41                          | 20.84±2.91                                  |
| UBQ7            | 21.34±1.31                 | 22.51±1.59                           | 21.34±1.03                          | 21.8±1.64                                   |
| ACT4            | 20.05±1.39                 | 21.3±1.94                            | 20.04±0.83                          | 20.66±1.66                                  |
| PP2A1           | 24.26±0.71                 | 25.05±1.67                           | 23.26±0.75                          | 24.46±1.31                                  |

\*Geometric mean of Ct values observed across all biological samples along with Standard Deviation (SD) value

Supplementary Table 3 Order of stability for reference genes based on their M value calculated by geNorm software

| Seedling stage |                 | Early reproductive stage |                 | Fiber development stage |                 | Across the developmental stages |                 |
|----------------|-----------------|--------------------------|-----------------|-------------------------|-----------------|---------------------------------|-----------------|
| Ranking        | Stability value | Ranking                  | Stability value | Ranking                 | Stability value | Ranking                         | Stability value |
| ACT4           | 0.52            | ACT4                     | 1.37            | UBQ7                    | 0.65            | UBQ7                            | 1.08            |
| UBQ7           | 0.52            | ACT7                     | 1.37            | ACT7                    | 0.65            | ACT4                            | 1.08            |
| ACT7           | 0.84            | UBQ7                     | 1.52            | ACT4                    | 0.81            | ACT7                            | 1.29            |
| PP2A1          | 1.01            | RNAH                     | 1.77            | ATUB                    | 0.89            | PP2A1                           | 1.68            |
| RNAH           | 1.12            | PP2A1                    | 2.13            | PP2A1                   | 0.98            | RNAH                            | 1.87            |
| $\alpha$ TUB   | 1.43            | $\alpha$ TUB             | 2.47            | RNAH                    | 1.36            | $\alpha$ TUB                    | 2.13            |

(M values are listed from the most stable to the least stable)

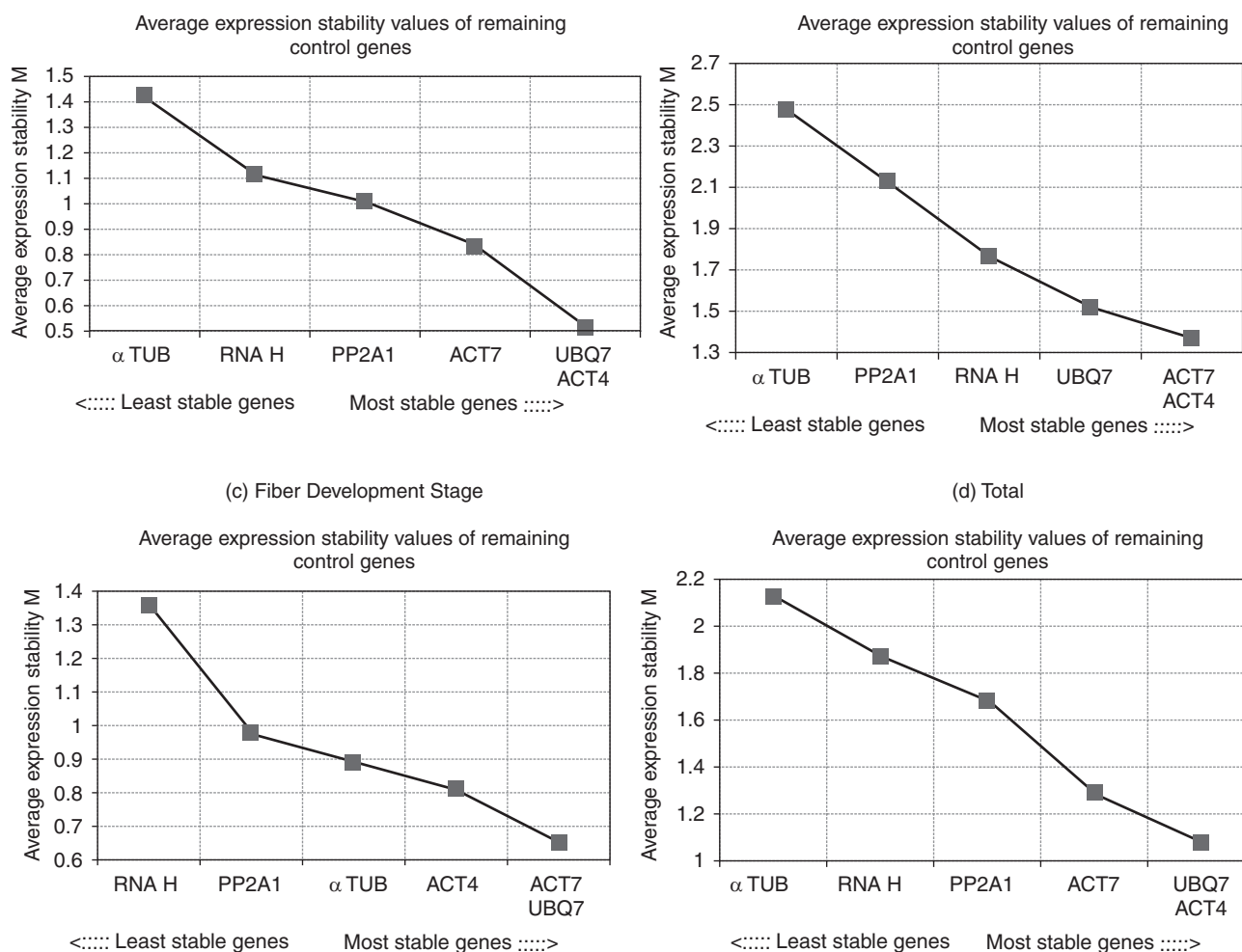


Fig 2 Average gene stability value (M value) and ranking of candidate reference genes by geNorm. Lower M value indicates more stable gene expression.

of identified reference genes. Relative expression of the *Gacobl9* during secondary cell developmental stages of cotton fibre (25 DPA) in relation to 10 DPA was quantified using *ACT4*, *ACT7*, and *UBQ7* as normaliser. We found the higher transcript abundance of target gene and it was in descending order for the normaliser *ACT4* (76.49), *ACT7* (39.05) and *UBQ7* (16.32) reference genes (Fig 4).

Allotetraploid *Gossypium* species including *G. hirsutum* (AD)1 and *G. barbadense* (AD)2 are the descendants of two diploid species *G. arboreum* (A genome) and *G.*

*raimondii* (D genome), resulted from genome union between A and D genomes and subsequent chromosome doubling (Yu-xiang *et al.* 2013). Of late, successful whole genome sequencing of cultivated diploid progenitor species has generated immense genomic data for exploration (Wang *et al.* 2012, Li *et al.* 2014.). Whole genome sequencing is an important and fundamental step for incisive scientific studies of wide dimensions, provide a valuable insights, vital for linkage relationship studies, identification of candidate genes, development of DNA based markers which can be



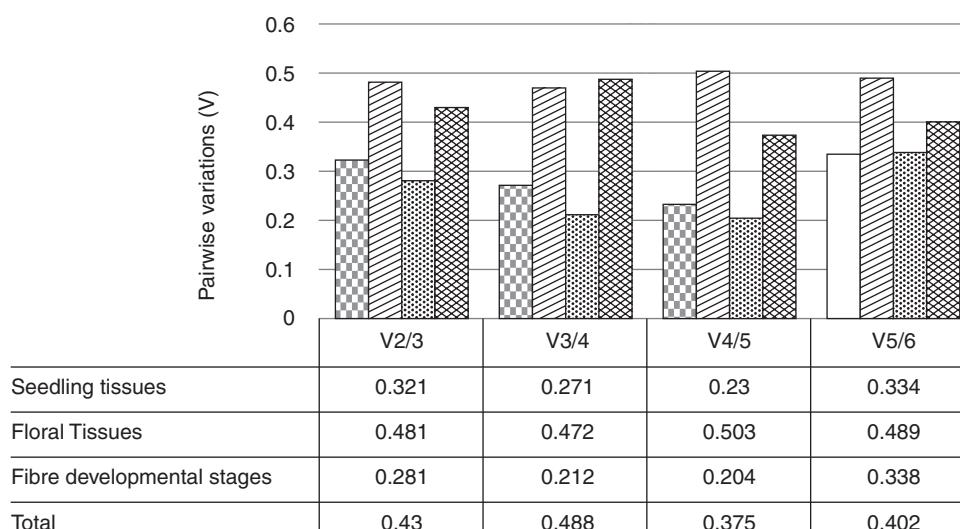


Fig 3 Pairwise variation ( $V_n/V_{n+1}$  variance) to determine the optimal number of control genes for an accurate normalization.

and abiotic stresses. With basic alphabets of *G. arboreum* genome decoded, assigning the function to the genome polymorphisms revealed in sequencing is most imperative and highly challenging task. Until and unless this sequence is converted to knowledge, the complete benefits of genome sequencing cannot be attained. The *G. arboreum* sequence information lays a strong platform for next generation genomics assisted desi cotton improvement. Hence, in present study, a total of six candidate reference genes were selected in *G.*

Supplementary Table 4 Reference genes expression stability values calculated by the NormFinder software

| Seedling stage   |                 | Early reproductive stage |                 | Fiber development stage |                 | Across the developmental stages |                 |
|------------------|-----------------|--------------------------|-----------------|-------------------------|-----------------|---------------------------------|-----------------|
| Ranking          | Stability value | Ranking                  | Stability value | Ranking                 | Stability value | Ranking                         | Stability value |
| ACT4             | 0.139           | ACT4                     | 0.192           | PP2A1                   | 0.312           | ACT4                            | 0.043           |
| UBQ7             | 0.17            | ACT7                     | 0.276           | ACT4                    | 0.344           | ACT7                            | 0.105           |
| PP2A1            | 0.393           | UBQ7                     | 0.366           | UBQ7                    | 0.469           | UBQ7                            | 0.15            |
| ACT7             | 0.482           | PP2A1                    | 0.438           | ACT7                    | 0.75            | PP2A1                           | 0.229           |
| RNAH             | 0.491           | RNAH                     | 0.46            | $\alpha$ TUB            | 1.168           | RNAH                            | 0.244           |
| $\alpha$ TUB     | 0.56            | $\alpha$ TUB             | 0.474           | RNAH                    | 1.182           | ATUB                            | 0.296           |
| Best combination |                 | Best combination         |                 | Best combination        |                 | Best combination                |                 |
| ACT4 & UBQ7      | 0.118           | RNAH & ACT7              | 0.187           | ACT7 & ACT4             | 0.298           | ACT4 & ACT7                     | 0.043           |

Stability values are listed from the most stable to the least stable

used in crop improvement either through plant breeding or through transgenic approach. With the recent and rapid developments in sequencing technologies, the time and cost aspects of sequencing have seen a great retrenchment. Having the sequence information of whole genome is just the beginning of long worthy journey in science.

*Gossypium arboreum* (Asiatic cotton) is well known for its tolerance to various climatic resilient features like biotic

Supplementary Table 5 Best combination of reference genes based on geNorm and NormFinder programs

| Seedling stage | Early reproductive stage | Fiber development stage | Across the developmental stages |
|----------------|--------------------------|-------------------------|---------------------------------|
| ACT4           | ACT7                     | ACT4                    | ACT4                            |
| UBQ7           |                          | UBQ7                    | ACT7                            |
| RNAH           |                          | ACT7                    | UBQ7                            |

Stability values are listed from the most stable to the least stable

*arboreum* based on their DNA sequence similarity to previously validated reference genes in *G. hirsutum* (Artico *et al.* 2010, Niu *et al.* 2015). A similar trend of candidate gene validation and selection is widely employed in many crop species (Reid *et al.* 2006, Wang *et al.* 2013). Three software packages were used for analysis and identification of suitable reference gene. The inherent differences in algorithms of these softwares resulted in variation in the ranking order of reference genes. For example, in our study, the BestKeeper outputs *ubiquitin7* gene, as most stable in seedling samples (Supplementary Table 6) while it ranked second by NormFinder (Supplementary Table 4). Likewise, *PP2A1* emerged as ideal gene during fiber development stage in NormFinder (Supplementary Table 4) while it was ranked four by geNorm (Supplementary Table 3). Similar results were also evidenced in other plant species (Chen *et al.* 2015, Wan *et al.* 2017). Our study revealed that *ACT4* and *ACT7* are the best combinations of internal control genes in all samples of three developmental stages (Supplementary Table 4). Actin filaments constitute a major

Supplementary Table 6 Expression stability analysis using BestKeeper

|                                 |          | RNAH  | ACT7  | $\alpha$ TUB | UBQ7  | ACT4  | PP2A1 |
|---------------------------------|----------|-------|-------|--------------|-------|-------|-------|
| Seedling stage                  | N        | 9     | 9     | 9            | 9     | 9     | 9     |
|                                 | GMCq     | 22.64 | 24.66 | 21.38        | 21.34 | 20.05 | 24.26 |
|                                 | $\pm$ SD | 0.74  | 0.92  | 2.34         | 1.31  | 1.39  | 0.71  |
|                                 | R        | 0.470 | 0.869 | 0.962        | 0.962 | 0.997 | 0.734 |
|                                 | p-value  | 0.201 | 0.002 | 0.001        | 0.001 | 0.001 | 0.024 |
| Early reproductive stage        | N        | 15    | 15    | 15           | 15    | 15    | 15    |
|                                 | GMCq     | 24.90 | 24.60 | 21.67        | 22.51 | 21.30 | 25.05 |
|                                 | $\pm$ SD | 1.71  | 1.10  | 3.40         | 1.59  | 1.94  | 1.67  |
|                                 | R        | 0.830 | 0.937 | 0.961        | 0.742 | 0.963 | 0.398 |
|                                 | p-value  | 0.001 | 0.001 | 0.001        | 0.002 | 0.001 | 0.143 |
| Fiber development stage         | N        | 9     | 9     | 9            | 9     | 9     | 9     |
|                                 | GMCq     | 24.70 | 22.69 | 18.72        | 21.34 | 20.04 | 23.26 |
|                                 | $\pm$ SD | 1.89  | 0.92  | 1.41         | 1.03  | 0.83  | 0.75  |
|                                 | R        | 0.692 | 0.863 | 0.953        | 0.844 | 0.915 | 0.699 |
|                                 | p-value  | 0.039 | 0.003 | 0.001        | 0.004 | 0.001 | 0.036 |
| Across the developmental stages | N        | n     | 30    | 30           | 30    | 30    | 30    |
|                                 | GMCq     | 24.30 | 24.16 | 20.84        | 21.80 | 20.66 | 24.46 |
|                                 | $\pm$ SD | 1.84  | 1.18  | 2.91         | 1.64  | 1.66  | 1.31  |
|                                 | R        | 0.609 | 0.856 | 0.943        | 0.835 | 0.957 | 0.564 |
|                                 | p-value  | 0.001 | 0.001 | 0.001        | 0.001 | 0.001 | 0.001 |

N= No. of samples; GMCq= Geometric mean of Cq; SD=Standard deviation; r=Coefficient of correlation

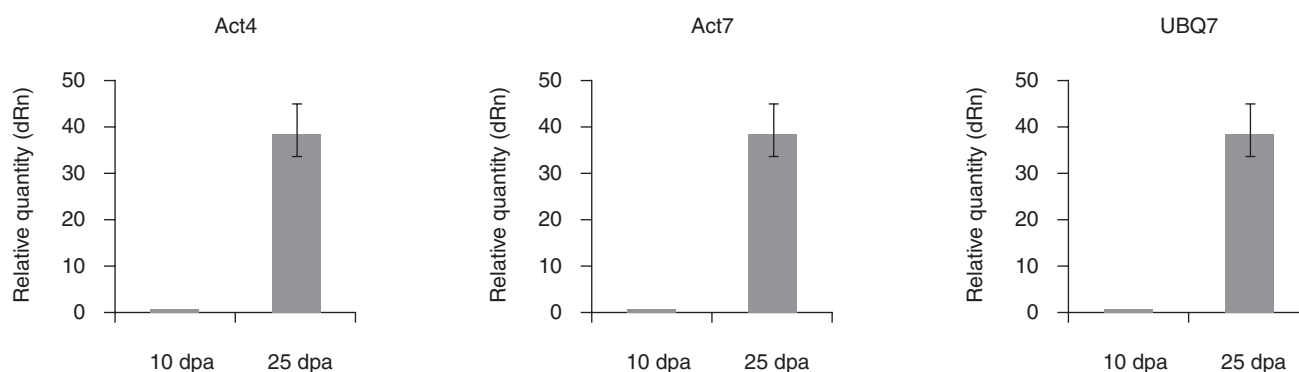


Fig 4 Relative quantity charts for Gacobl9 expression levels normalized with ACT4, ACT7 and UBQ7 during primary (10 DPA) and secondary cell wall deposition stage (25 DPA) of cotton fiber.

component of plant cell's cytoskeleton and are principle determinants of its shape and movements. *Actin* gene variants are traditionally being used and also reported as ideal internal control gene for qPCR in various plant species under different experimental conditions (Artico *et al.* 2010, Wan *et al.* 2017, Zhang *et al.* 2015). In addition, *ACT7* along with *RNAH* was found to be best for early reproductive development stage (Supplementary Table 5) which is also supported by BestKeeper analysis (Supplementary Table 6). During data analysis, we have observed that gene *RNAH* showed relative stable expression in early reproductive stage while poor and varied expression levels were observed in seedling and fiber development stages (Supplementary Table 5) suggesting tissue or stage-specific selection of

normalizer gene. Similarly, the weaker expression of most commonly used housekeeping gene such as  *$\alpha$ TUB* was also noticed (Fig 2, Supplementary Table 3 and 4). In congruence with our results, varied transcript abundance of *tubulin* as normalizer was also reported in crops like potato, grapes, and soybean (Jain *et al.* 2006, Nicot *et al.* 2005, Reid *et al.* 2006). Apart from actin, ubiquitin (*UBQ7*) was also emerged as one of the suitable internal control gene for samples representing seedling stage of cotton (Supplementary Table 5). *Ubiquitin*'s are classical enzymes play a very important role in post-translation modification of proteins. Stably expressed ubiquitin isoforms have also been identified and widely used (Ermei *et al.* 2012, Zhang *et al.* 2015). Similarly, Artico *et al.* (2010) identified and reported *ubq14*

has suitable housekeeping gene for all the tissues including plant organs and flower buds of *Gossypium hirsutum*.

Overall, the study has identified *actin4/actin7/ubiquitin7* as the most suitable reference genes for fiber development stages, whereas *actin4/ubiquitin7* and *actin7/RNA helicases* for seedling and early reproductive development stages, respectively. To validate the suitability of the identified internal control genes from the above study, tissue specific expression analysis of gene for COBRA like protein (*Gacobl9*) was performed. *Gacobl9* gene codes for a plant-specific glycosylphosphatidyl inositol (GPI) anchored protein and is reported to be expressed predominantly during secondary cell wall deposition stage of cotton fiber development (Niu *et al.* 2015). COBRA-like protein plays a key regulatory role in cellular microfibrillar orientation (Roudier *et al.* 2005). The massive accumulation of cellulose during secondary cell wall synthesis stage (25 DPA) of cotton fiber development necessitates the recruitment of the diverse group of proteins for the microfibrillar arrangement.

Relative expression of the *Gacobl9* during 25 DPA in relation to 10 DPA was quantified using *ACT4*, *ACT7*, and *UBQ7* as normaliser. Its expression during 25 DPA was found to be higher in comparison to 10 DPA. However, the relative transcript abundance of the *Gacobl9* gene was observed to be in descending order for *ACT4*, *ACT7*, and *UBQ7* which was in concurrence with their relative expression stability. The results obtained from our study are in line with the validation of identified reference genes reported elsewhere (Hu *et al.* 2009, Artico *et al.* 2010, Lin *et al.* 2014). In conclusion, our study identified and validated *actin4*, *actin7*, and *ubiquitin7* for their utilization as normalizers for gene expression studies in *Gossypium arboreum*. However, to avoid misleading outputs, a set of normalization genes should always be carefully evaluated and validated under specific experimental conditions.

#### REFERENCES

- Abd El-Moghny A M, Santosh H B, Raghavendra K P, Sheeba, J A, Singh S B and Kranthi K R. 2017. Microsatellite marker based genetic diversity analysis among cotton (*Gossypium hirsutum*) accessions differing for their response to drought stress. *Journal of Plant Biochemistry and Biotechnology* **26**: 366-70.
- Andersen C L, Jensen J L and Orntoft T F. 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 Research* **64**: 5245–50.
- Artico S, Nardeli S M., Brillhante O, Grossi-de-sa F M and Alves-Ferreira M. 2010. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biology* **10**: 49.
- Chen Y, Hu B, Tan Z, Liu J, Yang Z, Li Z and Huang. 2015. Selection of reference genes for quantitative real-time PCR normalization in creeping bentgrass involved in four abiotic stresses. *Plant Cell Reports* **34**: 1825–34.
- Czechowski T, Stitt M, Altmann T and Udvardi M K. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization. *Genome Analysis* **139**: 5–17.
- Ermei C, Shengqing S, Jianfeng L, Tielong C, Liang X, Xiuyan Y, Wenjuan, Y, Qian L and Zeping J. 2012. Selection of reference genes for quantitative gene expression studies in *Platycladus orientalis* (*cupressaceae*) using real-time PCR. *PLoS One* **7**: e33278.
- Fausto A K S, Silva T F, Romanel E and Vaslin M F S. 2017. microRNAs as reference genes for quantitative PCR in cotton. *PLoS ONE* **12**: e0174722.
- Hu R, Fan C, Li H, Zhang Q and Fu Y. 2009. Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Molecular Biology* **10**: 93.
- Huggett J, Dheda K, Bustin S and Zumla A. 2005. Real-time RT-PCR normalisation; strategies and considerations. *Genes & Immunity* **6**: 279–284.
- Jain M, Nijhawan A, Tyagi A K and Khurana J P. 2006. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications* **345**: 646–51.
- Jian B, Liu B, Bi Y, Hou W, Wu C and Han T. 2008. Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Molecular Biology* **9**: 59.
- Kim H J and Triplett B. 2001. Cotton fiber growth in planta and in vitro models for plant cell elongation and cell wall biogenesis. *Plant Physiology* **127**: 1361–6.
- Li F, Fan G, Wang K *et al.* 2014. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics* **46**: 567–72.
- Lin Y, Zhang C, Lan H, Lan H, Gao S, Liu H, Liu J, Cao M, Pan G, Rong T and Zhang S. 2014. Validation of potential reference genes for qPCR in maize across abiotic stresses, hormone treatments, and tissue types. *PLoS One* **9**: e95445.
- Marum L, Miguel A, Ricardo C P and Miguel C. 2012. Reference gene selection for quantitative real-time PCR normalization in *quercus suber*. *PLoS One* **7**: e35113.
- Nicot N, Hausman J F, Hoffmann L and Evers D. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany* **56**: 2907–2914.
- Niu E, Shang X, Cheng C, Bao J, Zeng Y, Cai C, Du X and Guo W. 2015. Comprehensive analysis of the COBRA-like (COBL) gene family in *Gossypium* identifies two COBLs potentially associated with fiber quality. *PLoS One* **10**: e0145725.
- Paterson A H, Wende J F, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D *et al.* 2012. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**: 423–7.
- Pfaffl M W, Tichopad A, Prgomet C and Neuvians T P. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnology Letters* **26**: 509–15.
- Raghavendra K P, Phanindra M L V and Kumar P A. 2014. Internal control gene for gene expression studies using real time quantitative RT-PCR during cotton boll development. *Cotton Research Journal* **6**: 32-6.
- Reid K E, Olsson N, Schlosser J, Peng F and Lund S T. 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology* **6**: 27.
- Roudier F, Fernandez A G, Fujita M, Himmelspach R, Borner G H, Schindelman G, Song S, Baskin T I, Dupree P, Wasteneys G O and Benfey, P.N. 2005. COBRA, an *Arabidopsis* extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically

- controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *Plant Cell* **17**: 1749–63.
- Tu L, Zhang X, Liu D, Jin S, Cao J, Zhu L, Deng F, Tan J and Zhang C. 2007. Suitable internal control genes for qRT-PCR normalization in cotton fiber development and somatic embryogenesis. *Chinese Science Bulletin* **52**: 3110–7.
- Vandesompele J, Preter K D, Pattyn F, Poppe B, Roy N V, Paepe A D, Roy N V, Paepe, A D and Speleman F. 2002. Accurate normalization of realtime quantitative RT–PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**: 0034.
- Wan D, Wan Y, Yang Q, Zou B, Ren W, Ding Y, Wang Z, Wang R, Wang K and Hou X. 2017. Selection of reference genes for qRT-PCR analysis of gene expression in *Stipa grandis* during environmental stresses. *PLoS One* **12**: 0169465.
- Wang K, Wang Z, Li F *et al.* 2012. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* **44**: 1098–1103.
- Wang M Wang Q and Zhang B. 2013. Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.). *Gene* **530**: 44–50.
- Yu-xiang W, Jin-hong C, Qiu-ling H and Shui-jin Z. 2013. Parental origin and genomic evolution of tetraploid *Gossypium species* by molecular marker and GISH analyses. *Caryologia* **66**: 368–74.
- Zhang C, Fu J, Wang Y, Bao Z and Zhao H. 2015. Identification of suitable reference genes for gene expression normalization in the quantitative real-time PCR analysis of sweet osmanthus (*Osmanthus fragrans* Lour.). *PLoS One* **10**: 0136355.