Evaluation of Sorghum \([\text{Sorghum bicolor (L.)}]\) Reference Genes in Various Tissues and under Abiotic Stress Conditions for Quantitative Real-Time PCR Data Normalization


International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India

Accurate and reliable gene expression data from qPCR depends on stable reference gene expression for potential gene functional analyses. In this study, 15 reference genes were selected and analyzed in various sample sets including abiotic stress treatments (salt, cold, water stress, heat, and abscisic acid) and tissues (leaves, roots, seedlings, panicle, and mature seeds). Statistical tools, including geNorm, NormFinder and RefFinder, were utilized to assess the suitability of reference genes based on their stability rankings for various sample groups. For abiotic stress, PP2A and CYP were identified as the most stable genes. In contrast, EIF4α was the most stable in the tissue sample set, followed by PP2A; PP2A was the most stable in all the sample set, followed by EIF4α. GAPDH, and UBC1 were the least stably expressed in the tissue and all the sample sets. These results also indicated that the use of two candidate reference genes would be sufficient for the optimization of normalization studies. To further verify the suitability of these genes for use as reference genes, SbHSF5 and SbHSF13 gene expression levels were normalized using the most and least stable sorghum reference genes in root and water stressed-leaf tissues of five sorghum varieties. This is the first systematic study of the selection of the most stable reference genes for qPCR-related assays in Sorghum bicolor that will potentially benefit future gene expression studies in sorghum and other closely related species.

Keywords: qPCR, RefFinder, Sorghum bicolor, gene expression stability, reference gene, normalization

INTRODUCTION

Sorghum is the fifth most important cereal crop that is widespread in the semi-arid regions of the world with an annual production of 65.5 Mt (FAO, 2011), possesses strong drought-tolerance traits and a high forage value. Sorghum seeds are utilized for both human and animal feed, and cultivars are processed for syrup, sugar, and alcohol. Additionally, sorghum cultivars have great economic

Abbreviations: M ACP , Malonyl CoA-Acyl Carrier protein; 6PGDH, 6-Phosphogluconate dehydrogenase; SAMDC, S-adenosylmethionine decarboxylase; CYP, Cyclophilin/Peptidylprolyl Isomerase; PP2A, Serine/threonine-Protein Phosphatase 2A; qPCR, Quantitative Real-Time Polymerase Chain Reaction.
potential as they can be valorized for second-generation biofuels to produce environment-friendly energy (Vermerris, 2011; Calviño and Messing, 2012). Sorghum is related to many C4 plants including maize, sugarcane, foxtail millet and switchgrass, which tend to have larger polyploid genomes containing repetitive sequences (Price et al., 2005). In contrast, sorghum is a diploid with a small genome (750 Mbp) and possesses an extraordinary germplasm diversity that greatly aids gene discovery and analysis through comparative and functional genomics, making it highly useful as model cereal for structural and functional genomic studies aimed at improving agronomically important traits (Paterson et al., 2009).

Coupled with species phenotypes, functional genomics can provide important insights into complex biological processes including abiotic stress responses. Functional genomic studies are now being used to identify the roles of regulatory and structural genes. The importance of sorghum functional genomics has increased following the sequencing of its genome (Paterson et al., 2009). Despite a wide range of available experimental approaches for exploring gene expression at the transcriptional level, e.g., northern blotting, ribonuclease protection assays, reverse transcription PCR (RT-PCR), quantitative real-time PCR (qPCR) and DNA microarrays (Valasek and Repa, 2005). Among them, qPCR is the most efficient for quantification of gene expression levels due to its simplicity, sensitivity, accuracy and cost (Wong and Medrano, 2005); however, despite these advantages, the utility of this method is often limited due to the absence of reliable reference genes for qPCR data normalization.

Classically, most of the reference genes fall in the category of housekeeping genes, which have major roles in the maintenance of basic cellular metabolism irrespective of physiological conditions (Bustin, 2002). These reference genes have been widely used in qPCR assays in molecular and functional genomic studies as internal calibrators (Vandesompele et al., 2002; Eisenberg and Levanon, 2013); however, recent studies have shown that these reference genes are often not stably expressed under various experimental conditions (Thellin et al., 1999; Zhu et al., 2013; Gimeno et al., 2014). Consequently, the selection and validation of suitable reference genes that are uniformly and stably expressed across experimental conditions have become imperative (Jian et al., 2006; Li et al., 2012). Various statistical tools, e.g., geNorm, NormFinder, BestKeeper and RefFinder, have been developed to determine reference gene suitability for qPCR data normalization (Vandesompele et al., 2002; Andersen et al., 2004; Pfaffl et al., 2004). Stable reference gene validation has been performed in many plant species, including model species and crop plants such as Arabidopsis thaliana (Czechowski et al., 2005; Remans et al., 2008), Brassica napus (Yang et al., 2014a; Machado et al., 2015), Hordeum vulgare (Zmienko et al., 2015), Fagopyrum tataricum (Demidenko et al., 2011), Cajanus cajan (Sinha et al., 2015), Oryza sativum (Jain et al., 2006; Li et al., 2010; Li et al., 2014), Arachis hypogaea (Reddy et al., 2013), Pennisetum glaucum (Saha and Blumwald, 2014; Reddy et al., 2015b), Solanum tuberosum (Nicot et al., 2005; Mariot et al., 2015), Glycine max (Jian et al., 2008; Nakayama et al., 2014), Saccharum officinarum (Iskandar et al., 2004; Ling et al., 2015), Lycopersicon esculentum (Expósito-Rodríguez et al., 2008), Triticum aestivum (Paolacci et al., 2009), and Zea mays (Manoli et al., 2012; Lin et al., 2014).

Until now, sorghum qPCR assays were performed using a single reference gene to quantify target gene expression in response to various experimental conditions. Traditional reference genes including 18S RNA, Actin, EIF4a, Tubulin, and Ubiquitin have been used as calibrators for quantifying sorghum gene expression in response to abiotic stresses and during plant growth and development (Yang et al., 2004; Jain et al., 2008; Cook et al., 2010; Shen et al., 2010; Wang et al., 2010; Dugas et al., 2011; Ishikawa et al., 2011; Koegel et al., 2013; Li et al., 2013; Gelli et al., 2014; Shakoor et al., 2014; Yang et al., 2014b; Yin et al., 2014; Kebrom and Mullet, 2015; Li et al., 2015; Walder et al., 2015). Nevertheless, the reference genes used in these studies were randomly selected from various sources without any experimental validation. Moreover, single gene quantification qPCR assays are well known to frequently exhibit variability in gene expression under various experimental conditions (Dheda et al., 2005). The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines developed for the proper selection (Bustin et al., 2009) and validation of stable candidate reference genes for qPCR experiments highly recommend averaging data from more than 2 reference genes (Bustin et al., 2010). To date, only one study has evaluated reference genes in virus-infected Sorghum bicolor tissues (Zhang et al., 2013), and to the best of our knowledge, there are no systematic studies regarding the selection of suitable stable sorghum reference genes in various tissues and abiotic stress conditions.

This study tested 15 potential candidate reference genes, Acp1 Carrier Protein 2 (ACP2), ADP-Ribosylation Factor (ADPRF), alpha Tubulin (α-TUB), beta Tubulin (β-TUB), Eukaryotic Initiation Factor 4A-1 (EIF4a), Elongation Factor P (EF-P), Glyceraldehyde 3-phosphate Dehydrogenase (GAPDH), Malonyl Co-A-Acp Carrier Protein (MACP), Malate Dehydrogenase (MDH), Cyclophilin/Pepididylprolyl Isomerase (CYP), Protein Phosphatase 2C (PP2C), 6-Phosphogluconate Dehydrogenase (6PGDH), S-adenosyl methionine Decarboxylase (SAMDC), Serine/threonine-Protein Phosphatase (PP2A), and Ubiquitin Protein 1, Isoform c (UBC1). The stabilities of these reference genes were analyzed in sets of samples from treatments with various abiotic stresses (salt, cold, heat, water stress, and ABA stress) and tissues (seeds, leaves, panicles, roots and mature seeds). Additionally, to verify and support the identification of the best-ranked candidate reference genes, SbHSF5 and SbHSF13 gene expression levels were assayed. The stable reference genes identified in this study will be helpful in future qPCR-based molecular and functional studies in sorghum and related species.

**MATERIALS AND METHODS**

**Plant Materials and Abiotic Stress Treatments**

Sorghum bicolor genotypes (Parbani Moti, Phule Vasudha (PVS), S35, M35, and BTx623) seeds were collected from the
sorghum breeding unit of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in Patancheru, India. Seeds were sown in pots filled with soil mixture (3:2:1 clay:sand:manure) in a glasshouse. Various abiotic stress treatments (salt, cold, heat, water stress, and ABA stresses) were performed and various tissues (seedlings, leaves, panicles, roots, and mature seeds) were sampled according to (Reddy et al., 2015a). Leaf tissues were collected from sorghum cultivars including Parbani Moti, Phule Vasudha, S35, M35, and BTx623 were grown under control conditions and under progressive drought where at normalized transpiration ratio (NTR) of reached at 0.1 (10% of soil moisture remaining in the pot) (Vadez and Sinclair, 2001) then leaf samples were collected. Tissue samples were collected from three individual seedlings for three independent biological replicates, immediately snap frozen in liquid nitrogen and stored at −80°C until RNA extraction.

**Total RNA Isolation and cDNA Synthesis**

Total RNA was extracted from tissues using the Plant RNA mini spin kit (MACHEREY-NAGEL GmbH and Co. KG, Neumann-Neander-Straße 6-8, Duren, Germany) with the in-column DNase I treatment according to the manufacturer’s instructions. Sample quantity and purity were determined using a NanoVue Plus Spectrophotometer (GE Healthcare). RNA samples with an OD$_{260}$/OD$_{280}$ absorbance ratio between 1.9 and 2.2 were used for further analysis, and RNA integrity was assessed on 1% denaturing formaldehyde agarose gel. For each sample, 1 µg of total RNA was reverse transcribed using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen) and oligo (dT) primers according to manufacturer’s instructions. The reverse transcribed cDNAs were then diluted 1:12 with nuclease-free water and used for qPCR analysis.

**Candidate Reference Gene Selection and qPCR Primer Design**

Fifteen candidate reference genes were identified from the literature and the sequences of their corresponding homologs were extracted from the NCBI and the Phytozome databases (Table 1). The selected candidate reference genes included GAPDH, UBC1, MACP, ACP2, α-TUB, EF-P, 6PGDH, SAMDC, CYP, MDH, ADPFR, β-TUB, PP2A, EIF4α, and PP2C that were earlier reported to be stably expressed in monocots such as rice, pearl millet, barley, foxtail millet, wheat, and maize (Jain et al., 2006; Paolacci et al., 2009; Manoli et al., 2012; Kumar et al., 2013; Reddy et al., 2015b; Zmienko et al., 2015). Primers were designed using the Primer3 software (Untergasser et al., 2012) with the following parameters: 59–62°C annealing temperature, 20–22 bp primer length, 45–55% GC contents, and 90–150 bp amplicon length (Table 1). Pooled and diluted cDNA samples were used for qPCR and 2% agarose gel electrophoresis was used to check primer pair specificity prior to sequencing. The sequences amplified by each primer combination were compared with GenBank sequences using the BLASTN and BLASTX algorithms to verify amplicon specificity.

**Quantitative Real-Time RT-PCR (qPCR)**

The qPCR reactions were performed using a Realplex Real-Time PCR system (Eppendorf, Germany) and SYBR Green mix (Bioline) in 96 well optical reaction plates (Axxygen, USA) sealed with ultra-clear sealing film (Platex). The reactions were performed in a 10 µl total volume containing 5 µl of 2x SensiMix SYBR No ROX mix (Bioline), 400 nM of each primer, 1.0 µl of diluted cDNA and nuclease-free water. The reaction conditions were 95°C for 2 min, followed by 40 cycles of 15 s at 95°C and 30 s at 62°C with fluorescent signal recording. After amplification, melt curves were generated for each reaction to ensure specific amplification. All qPCR reactions, including the non-template control, were performed in biological and technical triplicates. The mean values obtained from the nine values (triplicates of each biological triplicate) were used to calculate the final quantification cycle values (Cq).

**Data Analysis**

The quantitative cycle (Cq) values were recorded using the RT-PCR system default settings in which the baseline was automatically corrected and threshold values were estimated using the noise band mode. Statistical analysis (mean and CV) of the Cq values was performed using a Microsoft Excel 2010 spreadsheet. PCR efficiencies (E) for candidate genes were evaluated using the dilution series method and pooled cDNA samples. The 12-fold diluted pooled cDNA sample was used for 2-fold serial dilutions. Five serially diluted cDNA samples were used as templates for the construction of standard curves for each primer pair using the above PCR composition and conditions. Standard curves were generated using linear regression based on the quantitative cycle (Cq) values for the dilution series. The correlation coefficients ($R^2$) and slope values were obtained from the standard curves, and the PCR amplification efficiencies ($E$) were calculated according to the following equation: $E = (10^{-1/slope} - 1)$.

**Statistical Tools for Normalization**

The genEX Professional software geNorm and NormFinder algorithms (MultiD Analyses AB, Sweden) were used to identify and analyze the stably expressed gene (s) in diverse experimental samples. The raw Cq values for each gene were corrected according to their PCR efficiencies and then converted into relative quantities. The mean values for the biological replicates were used as the input data for the geNorm and NormFinder analyses. The geNorm program calculates the candidate gene expression stability (M value) using pairwise comparisons and stepwise exclusion (Vandesompele et al., 2002). Pairwise variation analysis was performed to identify the optimal number of reference genes required for normalization in each sample set using geNorm and qBase plus software (v: 2.4; Biogazelle, Belgium) (Hellemans et al., 2007). NormFinder also measures candidate reference gene expression stability using stepwise exclusion by estimating the intra- and inter-group variation. A low stability value indicates low combined variation and high expression stability (Andersen et al., 2004). RefFinder is a web-based tool...
<table>
<thead>
<tr>
<th>S.No.</th>
<th>Gene symbol</th>
<th>Gene name (Operational isoform)</th>
<th>Accession number</th>
<th>Cellular function</th>
<th>Primers (F/R) (5′-3′)</th>
<th>Amplicon length (bp)</th>
<th>Tm (°C)</th>
<th>PCR efficiency</th>
<th>Regression coefficient (R²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GAPDH</td>
<td>Glyceraldehyde 3-phosphate-Dehydrogenase</td>
<td>XM_002439118</td>
<td>Glycolysis and gluconeogenesis</td>
<td>AAGGCCGGCATTGCTTTGAAT ACATGTGGCAGATCGGCTGA</td>
<td>107</td>
<td>84.1</td>
<td>1.00</td>
<td>0.998</td>
</tr>
<tr>
<td>2</td>
<td>UBC1</td>
<td>Ubiquitin Protein 1</td>
<td>XM_002452660</td>
<td>Protein degradation</td>
<td>AGAGGCTGACTCCGCTGGG CTCAGTTGACACCGGCACACAC</td>
<td>120</td>
<td>88.1</td>
<td>0.98</td>
<td>0.994</td>
</tr>
<tr>
<td>3</td>
<td>MACP</td>
<td>Malonyl CoA-Acyl Carrier protein</td>
<td>XM_002465363</td>
<td>Fatty acid biosynthesis</td>
<td>GCATTGAGAACATCGGGG CTTGATGGAAACCTCTTTCCA</td>
<td>139</td>
<td>82.8</td>
<td>0.98</td>
<td>0.997</td>
</tr>
<tr>
<td>4</td>
<td>ACP2</td>
<td>Acyl Carrier Protein-2</td>
<td>XM_002462703</td>
<td>Fatty acid and Polyketides biosynthesis</td>
<td>AGCAACCTGTTCGGCGGCAAGAG GAAACAGAGGATGCGGCTG</td>
<td>110</td>
<td>84.3</td>
<td>1.00</td>
<td>0.996</td>
</tr>
<tr>
<td>5</td>
<td>α−TUB</td>
<td>Alpha −Tubulin</td>
<td>XM_002466490</td>
<td>Cytoskeleton structure protein</td>
<td>GAGGGTGAGTTCTCTGAGGCCCTCCCTCATCTCCATCCTCGC</td>
<td>103</td>
<td>85.2</td>
<td>1.00</td>
<td>0.997</td>
</tr>
<tr>
<td>6</td>
<td>EF-P</td>
<td>Elongation Factor P</td>
<td>XM_002463651</td>
<td>Peptide bond synthesis</td>
<td>TGAAGGCTGIGAGAAGATTGT GTAAGCTCTAGCTCCTAGCC</td>
<td>114</td>
<td>80.3</td>
<td>1.00</td>
<td>0.999</td>
</tr>
<tr>
<td>7</td>
<td>6PGDH</td>
<td>6-Phosphogluconate dehydrogenase</td>
<td>XM_002449451</td>
<td>Involved in the pentose phosphate pathway</td>
<td>CAACATGACCAACGAAATGA</td>
<td>91</td>
<td>85.3</td>
<td>0.97</td>
<td>0.997</td>
</tr>
<tr>
<td>8</td>
<td>SAMDC</td>
<td>S-adenosylmethionine deacetylase</td>
<td>XM_002446695</td>
<td>Synthesis of polar monomers</td>
<td>GTGGGCGACCTCTCACTTACCCGAGTGTGCCTTCACAGAG</td>
<td>132</td>
<td>85.1</td>
<td>0.98</td>
<td>0.996</td>
</tr>
<tr>
<td>9</td>
<td>CYP</td>
<td>Peptidyld prolyl isomerase</td>
<td>XM_002453800</td>
<td>cis-trans isomerization of proline imide peptide bonds</td>
<td>GTATCTGTGCTGGCCGCTGCTCTTTCCAACCCGCTCTCCACCC</td>
<td>108</td>
<td>81.1</td>
<td>1.00</td>
<td>0.989</td>
</tr>
<tr>
<td>10</td>
<td>MDH</td>
<td>Malate dehydrogenase</td>
<td>XM_002467034</td>
<td>Citric acid cycle and gluconeogenesis</td>
<td>TGCACGGTGGAGATTGGAAAGGCTCTCCCTATTGGAGGAAGAA</td>
<td>103</td>
<td>82.5</td>
<td>1.01</td>
<td>0.972</td>
</tr>
<tr>
<td>11</td>
<td>ADP-RF</td>
<td>ADP-Ribosylation Factor</td>
<td>XM_002441244</td>
<td>Regulators of vesicular traffic and actin remodeling</td>
<td>GCCGTCTCTCTTCCGGCGGGAAGATG</td>
<td>136</td>
<td>84.5</td>
<td>0.96</td>
<td>0.998</td>
</tr>
<tr>
<td>12</td>
<td>β−TUB</td>
<td>Beta Tubulin</td>
<td>XM_002439758</td>
<td>Cytoskeleton structure protein</td>
<td>GCCGTGAGGTACCTGCGGTAGT CAGCAGACACCTCGAGATG</td>
<td>98</td>
<td>84.3</td>
<td>1.00</td>
<td>0.999</td>
</tr>
<tr>
<td>13</td>
<td>PP2A</td>
<td>Serine/threonine-Protein Phosphatase</td>
<td>XM_002453490</td>
<td>Control the specific dephosphorylation</td>
<td>AACCCCGAAACCCGAGACTATG TCGGTCGCGAGCTATTGGAAC</td>
<td>138</td>
<td>82.7</td>
<td>0.97</td>
<td>0.991</td>
</tr>
<tr>
<td>14</td>
<td>EBF4a</td>
<td>Eukaryotic Initiation Factor 4A</td>
<td>XM_002451491</td>
<td>Eukaryotic translation</td>
<td>CAATTCACTTCCGCGGCAAGATG TCGCGAAACCTTACCGCGGGAAC</td>
<td>144</td>
<td>84.8</td>
<td>1.02</td>
<td>0.993</td>
</tr>
<tr>
<td>15</td>
<td>PP2C</td>
<td>Protein Phosphatase 2C</td>
<td>XM_002468507</td>
<td>Signal transduction</td>
<td>GTCAACCTTTCCGGCGGAAAATCG GCGGCTGACTTTTTCGTG</td>
<td>140</td>
<td>85.4</td>
<td>0.94</td>
<td>0.992</td>
</tr>
</tbody>
</table>

Details of candidate reference genes, NCBI accession number, their putative functions, primer sequences, product size and amplicon characteristics, i.e., melting temperature, PCR efficiency and regression coefficient values.
(http://www.leonxie.com/referencegene.php) that integrates the four major computational programs currently available [geNorm (Vandesompele et al., 2002), NormFinder (Andersen et al., 2004), BestKeeper (Pfaffl et al., 2004), and the comparative ΔΔt method (Silver et al., 2006)], and calculates the geometric mean for comprehensive ranking.

**SbHSF5 and SbHSF13 Genes Expression**

Sorghum heat shock transcription factors *SbHSF5* and *SbHSF13* were selected as target genes for validating stabilities of the most and least stable reference genes by quantifying the gene expression levels in various experimental samples. Sample collections and experiments were performed as described above. *SbHSF5* and *SbHSF13* gene expression levels were normalized using the two most stable candidate reference genes, *PP2A* and *EIF4e*, as well as the two least stable reference genes in the all sample group, *UBC1* and *GAPDH*, individually and in combination. For reference gene validations, root tissues of variety Parbani Moti, and leaf tissues of five sorghum varieties under water stress conditions were used for estimating relative

![FIGURE 1](image-url) | Specificity of primer pairs for qPCR amplification. (A) Agarose gel (2.4%) electrophoresis showing PCR products of the expected sizes for 15 candidate genes. M: 50 bp DNA marker (NEB). (B) Dissociation curves of 15 candidate reference genes under various experimental conditions, each showing a single peak.
expression levels of SbHSF5 and SbHSF13 using REST software (Pfaffl et al., 2002).

**RESULTS**

**Primer Specificity and PCR Efficiency Calculations**

The candidate reference genes selected for this study represent various functional classes and gene families. Primer pairs for 15 candidate reference genes were used for qPCR amplification of sorghum cDNA and yielded single PCR products of the expected sizes (Figure 1A), as well as single melting curve peaks (Figure 1B). The amplification products were sequenced, and verification using the NCBI database BLASTN algorithm revealed 100% matches, demonstrating the gene specificity of the qPCR primer pairs. While the amplification efficiencies ($E$) of the candidate reference genes ranged from 0.94 ($PP2C$) to 1.03 ($\beta$-TUB), the regression coefficient values ($R^2$) ranged from 0.972 ($UBC1$) to 0.999 ($FE-P$ and $\beta$-TUB) (Table 1).

**Sorghum Reference Gene Expression Analysis**

A total of 15 candidate reference genes were selected for qPCR normalization. A SYBR Green-based qPCR assay was used for transcriptional profiling of the 15 candidate reference genes in 10 samples, including five from abiotic stress conditions (water stress, salt, ABA, cold and heat shock) and five different tissues (leaves, roots, panicles, mature seeds and seedlings). Candidate reference gene transcript levels were determined using the quantification cycle (Cq) values, and these genes varied in their abundance (Figure 2). The mean candidate gene Cq values ranged from 17.70 to 28.92, with most falling between 22 and 25. $GAPDH$ exhibited the lowest mean Cq value (17.70), indicating that $GAPDH$ was the most highly expressed, while $PP2C$, $ACP2$, $\alpha$-TUB, $6PGDH$, $MDH$, $ELFP$, $EIF4\alpha$, and $PP2A$ were moderately expressed (mean Cq 22.31–24.14), whereas $\beta$-TUB (mean Cq 26.69) and $UBC1$ (mean Cq 28.91) were expressed at low levels. Across the tested samples, the $GAPDH$ gene showed the least variation in expression levels (coefficient of variation 2.59%), while $\alpha$-TUB was the most variable (13.53%). The degree of variation in reference gene expression across the tested samples is shown in Figure 2. These results indicated that none of the reference gene expression levels were constant and that they varied from one assay to another.

**Reference Gene Ranking and Expression Stability Analysis**

The stability values for each reference gene or combination of reference genes may vary from one experimental set to another. The orders of candidate gene stability ranking for the three sample sets were determined separately to identify the most stable reference genes using three different statistical tools, geNorm, NormFinder and RefFinder.

**GeNorm Analysis**

GeNorm software was used to determine the expression stability of the 15 candidate sorghum reference genes; for this analysis, lower M values indicate greater gene expression stability. The software recommends a cutoff M value of 1.5 to identify sets of reference genes that are stably expressed. Figure 3 displays the expression stability rankings for the tested candidate genes; the lowest M value was calculated for the $PP2A$ and $EIF4\alpha$ pair ($M = 0.76$) and corresponded to the most stable expression in the all sample group, whereas the M values for $GAPDH$ and $UBC1$ were considerably higher than for the remaining candidate genes.
For the abiotic stress treatments, *PP2A* and *CYP* were the most stable reference genes, while *UBC1* and *β-TUB* were the least stable (Figure 3). In the tissue group, *SAMDC* and *PP2A* were the most stable candidate genes while *CYP* and *GAPDH* were the least stable (Figure 3). The optimal numbers of reference genes required for gene expression normalization for the various sample groups were determined using geNorm. As shown in Figure 4, the V2/3 values of all three-sample sets were below the threshold value of 0.15 (0.144 for all pooled samples, 0.112 for the abiotic treatments, and 0.14 for the various tissue sample sets), indicating that two reference genes were sufficient for sorghum gene expression data normalization.

**NormFinder Analysis**

NormFinder rankings differed slightly from the geNorm rankings (Figure 5). Both tools ranked *PP2A* and *EIF4α* as being the most stable genes in all the sample sets, while *GAPDH* and *UBC1* were the least stable genes (Figure 5); in contrast,
α-TUB and PP2A emerged as the most stably expressed genes in the abiotic stress sample set, while geNorm ranked α-TUB eighth (Figure 5). For the tissue sample set, EIF4α and PP2A reference genes occupied the top positions, while geNorm ranked EIF4α third (Figure 5). GAPDH and UBC1 were also the least stable reference genes in the abiotic stress and tissue sample sets (Figure 5).

**Reffinder Analysis**

The comprehensive ranking of the candidate reference genes for all three experimental sample sets using Reffinder was highly consistent with the findings of geNorm and NormFinder. Reffinder analysis revealed that PP2A and EIF4α were the most stable in all the sample set (Table 2). Similarly, PP2A and CYP were highly stable under abiotic stress conditions, while EIF4α and PP2A were the two most stable genes for the tissue sample set (Table 2). Comprehensive ranking revealed that GAPDH was the least stable gene in the all sample and tissue sample sets, while β-TUB was the least stable in the abiotic sample set. UBC1 was the second most stable gene in all three-sample sets (Table 2). Comprehensive ranking for the individual abiotic stress treatments revealed a pattern similar to that of the abiotic stress sample set in which PP2A was second for heat, salt and water stresses. Similarly, EF-P remained in the third position for the ABA and clod stresses (Supplementary Table 1).

**Validation of the Best and Least Ranked Sorghum Reference Genes**

To validate the selected candidate reference genes, the expression patterns of SbHSF5 and SbHSF13 were assayed in root tissues, water-stressed samples and different varieties (Figures 6A–C). In sorghum, SbHSF5 and SbHSF13 genes have been shown to express in different tissues and induced by abiotic stress treatments (Nagaraju et al., 2015). In this study, the two most stable reference genes from all the sample set, PP2A and EIF4α, were used for normalization: in roots under normal growth conditions, SbHSF5 and SbHSF13 expression were not significantly different. In contrast, the SbHSF5 and SbHSF13 expression patterns were very different when GAPDH and GAPDH and UBC1 indicated a 2- to 3-fold increase in expression, while normalization with GAPDH and GAPDH+UBC1 indicated downregulation of both genes in root tissues under normal conditions (Figure 6A). To further validate the selected reference genes, the transcript levels...
were quantified in five sorghum varieties under water stress conditions (Figure 6C). For water stress samples, SbHSF5 and SbHSF13 expression levels increased 1 to 12-fold under water stress conditions when normalized with the two most stable reference genes from all the sample set (PP2A and EIF4α; Figures 6B,C). As expected, normalization with the least stable genes (GAPDH and UBC1), either alone or in combination with UBC1 resulted in discrepancies for expression of both SbHSF5 and SbHSF13 gene expression (Figures 6B,C).

**DISCUSSION**

Accurate gene expression primarily relies on the stable expression of reference genes under diverse experimental conditions; however, thus far, no single reference gene has been shown to be stably expressed across experimental conditions (Olsvik et al., 2008; Cruz et al., 2009). The stability value of the most stable reference gene or the best combination of genes may vary from one experimental setup to another, meaning...
that the reference genes used for normalization should be validated under certain experimental conditions by statistically and experimentally. In the present study, 15 candidate reference genes were evaluated for expression stability among diverse experimental samples including various tissues and abiotic stress conditions. Expression analysis revealed that all 15 candidate reference genes varied significantly across the 10 tested samples; therefore, reference genes for optimal normalization were chosen from a set of candidate genes for each experiment either alone or in combination. To determine the comprehensive ranking of candidate reference genes in a sample set, we also used RefFinder, which considers the results of the four algorithms (geNorm, NormFinder, BestKeeper and ΔCq method) together. Based on their stability rankings, various candidate genes have been proposed for their suitability as reference genes depending on the experimental conditions, implying that the reference genes used for abiotic stress studies in sorghum might not be suitable for gene expression analysis across genotypes and species.

In all the sample set, all three algorithms identified two reference genes, PP2A and EIF4α, as the two most stable genes. The pairwise variation calculated by geNorm also indicated that the use of two genes would be sufficient for normalization (Figure 4). Based on these results, it is inferred that PP2A and EIF4α would be appropriate for qPCR data normalization for the sorghum all sample set. For the abiotic stress treatment sample set, the pairwise variation indicated that the top ranking two genes (PP2A and CYP) could be used for normalization (Figure 4). Similarly, PP2A and EIF4α would be sufficient for the normalization of the tissue sample set data (Figure 4). PP2A was identified as one of the two most stable genes in all three studied sample sets (Table 2). This result is in agreement with previous studies of maturing B. napus embryos (Chen et al., 2010), multiple stress conditions (Wang et al., 2014), development and under various environmental conditions in A. thaliana (Czechowski et al., 2005) and Caragana intermedia (Zhu et al., 2013), as well as for virus-infected Nicotiana benthamiana (Liu et al., 2012) and Z. mays (Zhang et al., 2013). Furthermore, PP2A was the most stably expressed gene among various Striga life stages (Fernández-Aparicio et al., 2013). EIF4α was second most stably expressed gene in the all sample and tissue sample sets (Table 2). Similar results were observed in Carica papaya, where EIF4α expression was stable under most of the experimental conditions tested (Zhu et al., 2012). EIF4α was also reported as being a stably expressed gene in sexual and apomictic Brachiaria brizantha (Silveira et al., 2009), and in sexual and apomictic accessions of Cenchrus ciliaris (Simon et al., 2013). In the present study, CYP was the third most stable gene; it was stably expressed under abiotic stress treatments and in the tissue sample set (Table 2). The results obtained in the present study are in agreement with our previous studies in peanuts, where CYP was the most stably expressed gene during vegetative stages and under abiotic stress conditions (Reddy et al., 2013) and was also stably expressed in various tissues of G. max (Jian et al., 2008), Vicia faba (Gutierrez et al., 2011) and salt-stressed S. tuberosum (Nicot et al., 2005). The CYP gene was moderately stable in developing and germinating G. max seeds (Li et al., 2012) and in bananas under various experimental conditions (Chen et al., 2011).
FIGURE 6 | Relative expression of SbHSF5 and SbHSF13 genes in root tissues (A) leaves during water stress treatment (B) selected varieties (C) using different reference genes selected by the RefFinder tool. SbHSF5 and SbHSF13 expression levels were normalized in a single and combined manner with either a stable or unstable reference genes. All samples were analyzed in triplicate, in three independent experiments.

contrast, our recent study in Cicer arietinum indicated that CYP was the least stably expressed gene under various experimental conditions (Reddy et al., 2016), consistent with a previous study of grapevine berry development (Reid et al., 2006). In conclusion, PP2A, EIF4α, and CYP are recommended as suitable reference genes for the qPCR-based normalization of gene expression in sorghum under abiotic stress and in various tissues.

To further validate the candidate reference genes selected from the RefFinder comprehensive ranking, the expression levels of two sorghum heat shock transcription factors, SbHSF5 and SbHSF13, were assessed. Plant heat shock transcription factors play major roles in higher plants under various abiotic stresses. In sorghum, SbHSF5 and SbHSF13 are expressed under abiotic stress conditions: SbHSF5 is moderately expressed under cold stress and upregulated under drought stress, and SbHSF13 is moderately expressed under cold and heat stress and upregulated under drought and high salinity stress (Nagaraju et al., 2015). In the present study, SbHSF5 and SbHSF13 expression levels were upregulated under water stress conditions and no significant changes were observed in root tissue gene expression levels when normalized to the most stable reference genes, PP2A and EIF4α (Figures 6A, B). While, SbHSF5 and SbHSF13 genes significantly upregulated under water stress conditions in leaf tissues of all five sorghum genotypes with relatively higher expression when
normalized with the sable reference genes (PP2A and EIF4A), the normalization was obscured when least stable reference gene (s) (UBC1 and GAPDH) were used (Figures 6B,C). The upregulation of ShHSF5 and ShHSF13 under this study is in agreement with another recent study in sorghum (Nagaraju et al., 2015) confirming the stability validation of the selected candidate reference genes and demonstrated the disadvantages of using unstable reference genes for normalization.

CONCLUSION

This study reports on the selection and validation of stable reference genes for qPCR-based gene expression studies in sorghum in response to various abiotic stresses and in different tissues. Three major statistical tools, geNorm, NormFinder and RefFinder, were used to analyze the suitability of reference genes. Some slight differences were observed between the statistical tools with regards to the selected stable reference genes. Overall, PP2A and EIF4A are the most stable candidate reference genes, and UBC1 and GAPDH were found to be least stable. The use of two reference genes would be optimal for the normalization of sorghum gene expression levels in various tissues and abiotic stress conditions. The most and least stable reference genes were further validated for their suitability as reference genes by normalizing ShHSF5 and ShHSF13 expression levels under various experimental conditions across five genotypes. This work will benefit future studies of gene expression in related experiments using different tissues and abiotic stress conditions in S. bicolor and related crops.

AUTHOR CONTRIBUTIONS

PS, KS, and VD designed the experiments. PS, DS, SK, and PB performed the experiments and data analysis. PS, DS and KS drafted the manuscript.

FUNDING

Financial support to PS through the INSPIRE Faculty Award (IFA11-LSPA-06) and the Young Scientist Scheme (SB/YS/LS-12/2013) provided by the Department of Science and Technology, Government of India, New Delhi is gratefully acknowledged. This work was performed as part of the CGIAR Research Program on Dryland Cereals.

ACKNOWLEDGMENTS

PS acknowledges INSPIRE Faculty Award (IFA11-LSPA-06) and the Young Scientist Scheme (SB/YS/LS-12/2013) from the Department of Science and Technology, Government of India, New Delhi. This work was performed as part of the CGIAR Research Program on Dryland Cereals. ICRISAT is a member of the CGIAR Consortium.

REFERENCES


Sudhakar Reddy et al.


**Conflict of Interest Statement** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Sudhakar Reddy, Srinivas Reddy, Sivasakthi, Bhatnagar-Mathur, Vadez and Sharma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.