Journal of Oil Palm Research Vol. 34 (3) September 2022 p. 439-452
DOI: https://doi.org/10.21894/jopr.2022.0002

DETERMINATION OF RELIABLE REFERENCE GENES FOR REVERSE TRANSCRIPTION QUANTITATIVE REAL-TIME PCR FROM OIL PALM TRANSCRIPTOMES

NADZIRAH AMIRUDDIN[1]; PEK-LAN CHAN[1]; KUANG-LIM CHAN[1]; PEI-WEN ONG[1]; PRISCILLA ELIZABETH MORRIS[1]; MEILINA ONG-ABDULLAH[1]; SUBHI SITI MASURA[1] and ENG-TI LESLIE LOW[1]*

ABSTRACT

*A set of reliable reference genes is essential for accurate quantification and interpretation of gene expression data using reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR). In this study, Roche-454 RNA-seq reads from 27 libraries of various oil palm tissues were systematically analysed to identify a set of potential reference genes. Eleven candidate reference genes were identified from the transcriptome data. These genes, together with three oil palm reference genes previously identified for tissue culture samples (*PD000380, PD00569, pOP-EA01332*) and five classical housekeeping genes* [glyceraldehyde-3-phosphate dehydrogenase (GAPDH), NAD5, TUBULIN, UBIQUITIN, ACTIN] *were analysed across samples collected from various tissues from mature oil palm (leaf, root, endosperm, mesocarp, female flowers) and stages of tissue culture (non-embryogenic callus, embryogenic callus, polyembryoids, and shoots from polyembryoids). The expression levels of these genes were compared and evaluated using geNorm. Three genes* (EgREF_5, EgREF_7 *and* EgREF_11) *were found to be appropriate reference genes for normalising gene expression data from both mature plant and tissue culture samples.*

Keywords: housekeeping genes, normalisation, RT-qPCR, transcriptome.

Received: 21 May 2021; **Accepted:** 2 November 2021; **Published online:** 17 January 2022.

INTRODUCTION

Gene expression analysis or the study of transcript abundance is essential in molecular biology research, especially in understanding the role of gene expression patterns in different biological processes (Dussert *et al.*, 2013; Kong *et al.*, 2021; Tranbarger *et al.*, 2011). It furthers our understanding and provides insights on the genetic and molecular mechanisms underlying developmental and cellular processes.

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is one of the most powerful and sensitive techniques to quantify gene expression levels and has been recognised as a key driver of gene expression analysis in numerous molecular biology applications (Kozera and Rapacz, 2013; VanGuilder *et al.*, 2008). It is still considered the method of choice to validate high-throughput gene expression data (Everaert *et al.*, 2017). This technique is sensitive enough to detect gene expression changes, for even low transcript levels (Bustin *et al.*, 2005; Nolan *et al.*, 2006).

Nevertheless, it is important to note that substantial experimental variability, such as initial material quality, different inhibitors in samples, primer design and reverse transcription efficiencies should be taken into account to accurately quantify gene expression (Ginzinger, 2002; Mahoney *et al.*, 2004). Technical variability could be added to

[1] Malaysian Palm Oil Board,
6 Persiaran Institusi, Bandar Baru Bangi,
43000 Kajang, Selangor, Malaysia.

* *Corresponding author e-mail: lowengti@mpob.gov.my*

the data by random pipetting errors (Bustin and Nolan, 2004; Fleige and Pfaffl, 2006). These factors can cause unreliability in the quantifications of gene transcripts. Despite being considered highly sensitive, accurate, and reproducible (Kubista *et al.*, 2006), this approach needs normalisation of the expression data to reduce the effects of variability in experimental data. Selection of an appropriate normalisation strategy is critical to acquire biologically meaningful data. Commonly, the level of gene expression is normalised by comparing messenger ribonucleic acid (mRNA) levels of target genes to endogenous controls, known as reference genes.

Reliable reference genes are required to interpret quantification data from RT-qPCR and compensate for any differences in the studied tissues or cells. An ideal reference gene is expressed at a constant level across various conditions and unaffected by experimental parameters (Guénin *et al.*, 2009; Schmittgen and Zakrajsek, 2000; Thellin *et al.*, 1999; Zhu *et al.*, 2013). In addition, they should not be co-regulated with the target gene but must be expressed in abundance with minimal variability (Radonic *et al.*, 2004). The reference and target genes should also have similar ranges of expression in the samples to be analysed (Cappelli *et al.*, 2008). The use of one or more reference genes is preferred for optimum normalisation (Bustin *et al.*, 2009; Chandna *et al.*, 2012; Guénin *et al.*, 2009; Vandesompele *et al.*, 2002; 2009). However, expression levels of reference genes can vary in response to changes in experimental conditions and/or tissue types, and that using unstable reference genes in the relative quantification of gene expression will lead to biases and inappropriate biological data interpretation (Artico *et al.*, 2010; Czechowski *et al.*, 2005; Dheda *et al.*, 2005; Li *et al.*, 2020; Thellin *et al.*, 1999).

The expression of classical housekeeping genes such as *glyceraldehyde-3-phosphate dehydrogenase* (GAPDH), *actin*, *tubulin* and *18S ribosomal RNA* were presumed to have constant levels of expression, being involved in basal cell metabolism, cytoskeleton or intracellular functions. These genes are widely used as reference genes. However, their expression levels have been demonstrated to vary under different experimental conditions (Barsalobres-Cavallari *et al.*, 2009; Exposito-Rodriguez *et al.*, 2008; Jian *et al.*, 2008; Karuppaiya *et al.*, 2017; Li *et al.*, 2020; Mallona *et al.*, 2010; Qi *et al.*, 2010; Reid *et al.*, 2006; Tang *et al.*, 2021).

Meta-analysis of transcript data has been shown to be an efficient method to survey for novel stably expressed genes from high-throughput technologies as an alternative to mine for reference genes. For example, microarray datasets have enabled the identification of novel reference genes from a variety of plants, including *Arabidopsis thaliana*, *Eucalyptus*, soybean and rice, as well as in human samples and cynomolgus monkeys (Chan *et al.*, 2014; Chang *et al.*, 2011; Cheng *et al.*, 2011; Czechowski *et al.*, 2005; de Oliveira *et al.*, 2012; Garrido *et al.*, 2020; Libault *et al.*, 2008; Narsai *et al.*, 2010; Park *et al.*, 2013).

Another high-throughput analysis method which uses deep-sequencing technologies known as RNA-seq, has provided an additional resource to microarrays. It is widely used for quantitative mRNA expression studies and sensitive enough to detect very low transcript expressions and their isoforms (Marioni *et al.*, 2008; Wang *et al.*, 2009). Furthermore, it also has the ability to identify novel transcripts and splice variants (Trapnell *et al.*, 2010). The technique is fast and generates replicated data with minimal variations (Marioni *et al.*, 2008; Mortazavi *et al.*, 2008; Nagalakshmi *et al.*, 2008; Wang *et al.*, 2009). Thus, RNA-seq is a feasible whole-transcriptome method for mining of stably expressed genes and the identification of novel reference genes for RT-qPCR normalisation. In recent findings, combinations of suitable reference genes for abiotic stresses in potato (Tang *et al.*, 2017) and *Arabidopsis pumila* (Jin *et al.*, 2019) were identified from transcriptome data.

RNA-seq derived transcriptome data has been widely explored in a number of oil palm studies (Beulé *et al.*, 2011; Bourgis *et al.*, 2011; Dussert *et al.*, 2013; Kong *et al.*, 2021; Shearman *et al.*, 2013). In oil palm, RNA-seq data was generated in studies on lipid accumulation (Tranbarger *et al.*, 2011) and carbon partitioning (Bourgis *et al.*, 2011) in mesocarp, response to *Ganoderma boninense* infection (Tee *et al.*, 2013), normal or mantled flowers and fruits (Shearman *et al.*, 2013), and phosphorus starvation in roots (Kong *et al.*, 2021). More researches were also conducted with the availability and increasing volume of African oil palm sequence data (Adam *et al.*, 2007; Bourgis *et al.*, 2011; Chan *et al.*, 2017; Jaligot *et al.*, 2011; Low *et al.*, 2008; 2014; Singh *et al.*, 2013). These datasets as a whole provide an invaluable pool of information that can be utilised to identify stably expressed genes that would greatly contribute to accurate and reliable quantification of RT-qPCR data. In oil palm, a number of reference genes for normalisation of gene expression has been identified in specific developmental stages, such as tissue culture (Chan *et al.*, 2014), stress-treated samples (Xia *et al.*, 2014), diverse sets of biological samples (vegetative and reproductive tissues, and developmental stages of mesocarp tissues) (Yeap *et al.*, 2014), as well as from young plantlets (Muhammad Afiq *et al.*, 2019). Nevertheless, improving the collection of reference genes would certainly help expedite validation of newly discovered genes involved in various tissues and developmental stages.

In this study, Roche-454 RNA-seq transcriptome libraries comprising 27 oil palm tissue samples were utilised to identify a set of novel reference

genes for normalisation of RT-qPCR expression data in oil palm. Nineteen candidate reference genes including three oil palm reference genes (*PD00380, PD00569, pOP-EA01332*) previously identified for tissue culture samples (Chan *et al*., 2014), five classical housekeeping genes (*GAPDH, NAD5, TUBULIN, UBIQUITIN, ACTIN*), and 11 genes from the transcriptome data were evaluated. Statistical analysis using geNorm identified the three most stable reference genes in combination, *EgREF_5, EgREF_7* and *EgREF_11*. This result demonstrates that stably expressed reference genes could be identified from mining and utilisation of transcriptome datasets.

MATERIALS AND METHODS

Plant Materials

Leaf (spear and mature), endosperm [10 weeks after anthesis (WAA) and 15WAA], mesocarp (10WAA and 15WAA) and female inflorescence (frond number F13: 8.5 cm in length and F17: 20.0 cm) were sampled from Malaysian Palm Oil Board (MPOB) Research Station, Kluang, Johor, Malaysia. White root from seedlings in polybags, primary root (10 months nursery palms) and lateral root (10 months nursery palms) were sampled from the oil palm nursery in MPOB, Bangi, Selangor, Malaysia. Embryogenic callus, non-embryogenic callus, polyembryoids, and shoots from polyembryoids were collected from the tissue culture laboratory at the Advanced Biotechnology and Breeding Centre, MPOB, Bangi, Selangor, Malaysia. All samples were frozen in liquid nitrogen and stored at -80ºC prior to RNA extraction.

Ribonucleic Acid (RNA) Extraction, Purification and Quality Assessment

Total RNA was extracted according to the method by Ong *et al*. (2019). The total RNA was purified using RNeasy Mini Kit with on-column RNase free DNase I treatment according to the manufacturer's instructions (Qiagen, USA, Valencia, CA, USA). The concentration and purity of total RNA were determined using Nanodrop ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc.), and the integrity was assessed by electrophoretic fractionation on an Agilent 2100 Bioanalyser using RNA 6000 Nano LabChip (Agilent Technologies, CA, USA).

Identification of Candidate Reference Genes in Oil Palm Transcriptome Data Sets

Candidate reference genes in oil palm were identified by performing differential expression analysis of Roche-454 RNA-seq reads from 27 libraries of various oil palm tissues from PRJNA201497 (leaf, root, seedling white root, pollen, mesocarp, endosperm) (Singh *et al*., 2013) and PRJNA345530 [pith (one day after anthesis, DAA), sepal (1DAA), fruit (1DAA), spikelet (1DAA), stalk (1DAA)] (Chan *et al*., 2017). Reads of each tissue were mapped to the published oil palm EG5 reference genome (Singh *et al*., 2013) using TopHat2 (Kim *et al*., 2013). Gene expression data was generated by cuffdiff (part of TopHat2 package) and normalised using the Fragments Per Kilobase per Million Reads method (FPKM). The average expression of genes expressed in each tissue was calculated and genes with a minimum FPKM of 40 in all tissues were selected as putative stably expressed genes. These identified genes were compared to the protein database using BLASTX (https://blast.ncbi.nlm.nih.gov) with default parameter 1 e^{-5}.

Primer Design and Efficiency Test

Primer pairs for candidate reference genes were designed using Primer3Plus software (http://www.bioinformatics.nl/primer3plus) (Untergasser *et al*., 2007) with melting temperatures between 60°C and 67°C, primer lengths 20-27 bases, GC content 40%-60%, and amplicon lengths of 100-150 bp (Chan *et al*., 2014). The transcript sequences of candidate reference genes used for primer design were aligned to the oil palm EG5 genome sequence to predict intron positions using Exonerate program (Slater and Birney, 2005). Primer pairs were designed on either different exons or spanning exon-exon junctions of the complementary deoxyribonucleic acid (cDNA) (Hu *et al*., 2009) to avoid co-amplification of the genes from genomic DNA. Primer search against the oil palm genome was performed to check the specificity of each designed primer. The HPLC-purified primers were purchased from Bio Basic Canada Inc. All primer pairs were tested for polymerase chain reaction (PCR) amplification efficiencies (Ex) and to check for the specificity of the amplicon.

Reverse Transcription Quantitative Real-time PCR (RT-qPCR)

First-strand cDNA synthesis was carried out using 2 μg of total RNA using the High-capacity cDNA Reverse-Transcription Kit according to the manufacturer's instructions (Applied Biosystems, Foster City, CA, USA). The synthesised cDNA strands were used as templates in a SYBR Green based RT-qPCR using the Eppendorf Mastercycler® ep *realplex* (Eppendorf, Germany). RT-qPCR was performed according to the method described by Chan *et al*. (2014). A 'no reverse transcriptase' (NRT) control and a 'non-template control' (NTC) were assigned as negative controls.

Data Analysis

Cycle threshold (Ct) values were retrieved using Realplex software version 2.2 (Eppendorf, Germany). Results were imported into Microsoft Excel and data analysis was carried out by calculating the average Ct values for three replicates. Subsequently, the results were transformed into expression quantities using the method described by Vandesompele *et al.* (2002), Ex^(minCt - sampleCt), (Ex = $[10^{(-1/slope)}-1] \times 100\%$, slope = slope of linear regression). The most stable reference genes across all samples were selected based on geNorm v3.4 using log-transformed data as input (Vandesompele *et al.*, 2002).

RESULTS

Selection of Candidate Reference Genes and Expression Analysis

A total of 11 putative stably expressed genes were identified through differential gene expression analysis of 27 oil palm transcriptome libraries. These genes are involved in various functions, such as DNA binding protein, macrophage migration inhibitory factor family protein, histone, deleted in split hand/splt foot protein, MKI67 FHA domain-interacting nucleolar phosphoprotein-like, NADH-ubiquinone oxidoreductase 13 kDa-B subunit, ribosomal protein, OB-fold nucleic acid binding domain-containing protein and ubiquitin. Of all the 11 genes, *EgREF_10*, which is a ubiquitin family protein (*Table 1*), showed the highest level of variation, with the highest expression in pollen and floret after anthesis (*Figure 1*). The difference between the highest and lowest was ~2.2 fold difference in expression while the average difference for the other genes was ~1.6 fold. Primer pairs were designed for the 11 genes from the transcriptome data, five classical housekeeping genes and three oil palm tissue culture samples reference genes (Chan *et al.*, 2014). Efficiency of designed primer pairs was evaluated by the presence of a single peak in the melting curve obtained after 40 cycles of amplification. Only primer pairs which showed a single amplified product were selected for further studies. Thirteen candidate genes were thus selected for further analysis (*Table 2*). The Ex of the selected primers ranged from 82% to 100%, and correlation coefficient (R^2) ranged from 0.979 to 0.999. Subsequently, SYBR Green detection-based RT-qPCR assay was carried out for transcript profiling of these genes. As shown in *Figure 2*, the genes have an average Ct value of 22.79. Most of the genes have expression values between 20 and 24. *UBIQUITIN* was the most abundant gene of the set (mean Ct: 20.18), whereas *PD00380* and *pOP-EA01332* were the least abundant genes (mean Ct: *PD00380* = 25.58, *pOP-EA01332* = 25.98). *PD00380* showed highest level of Ct variation, with a range between 33.91 and 22.23, of which the difference between 25th and 75th percentile is ~5 Ct.

Selection of Potential Reference Genes for Oil Palm

GeNorm analysis executed using a Visual Basic Application in Microsoft Excel (Vandesompele *et al.*, 2002) was used to select the most stably expressed reference gene across various oil palm tissues. Gene expression stability measure (M) of each reference gene was calculated using the relative expression values for each cDNA sample as input for the geNorm algorithm. The threshold proposed for stably expressed genes was M≤0.5 (Vandesompele *et al.*, 2002). Genes with the lowest M value are deemed to have the most stable expression. Two most stably expressed genes were obtained by eliminating the least stable gene in a stepwise manner. The algorithm ranked the potential reference genes based on their expression stability (*Figure 3*). *EgREF_5* and *EgREF_7* were the most stable genes having M≤0.5, while other genes (*EgREF_11, pOP-EA01332, PD00569, EgREF_1, ACTIN, GAPDH, EgREF_2, EgREF_5, UBIQUITIN*) had an M value between 0.5 and 1. The least stable genes, *TUBULIN* and *PD00380*, had M values of more than 1.

GeNorm was also used to determine the optimal number of reference genes required for accurate and reliable normalisation of expression data in the tested sample sets. Parameter V, defined as the pairwise variation (V_n/V_{n+1}) was calculated by geNorm between the two sequential normalisation factors (NF) of n and n + 1 genes of consecutively ranked reference genes. The purpose is to measure the effect of adding further reference genes on the normalisation factor. A cut-off threshold parameter V of 0.15 was recommended by Vandesompele *et al.* (2002), below which the addition of more reference genes is not required. As shown in *Figure 4*, the analysis showed that V3/V4 was 0.129, suggesting that the optimum number of reference genes is three. Hence, the best combination of reference genes is *EgREF_5, EgREF_7* and *EgREF_11*. The data also indicates that the new reference genes have greater expression stability than the conventionally used housekeeping genes, and can therefore provide more reliable normalised expression data.

DISCUSSION

RT-qPCR is one of the most sensitive tools that is commonly used to achieve rapid and reliable quantification of gene expression levels. However,

TABLE 1. SELECTED CANDIDATE REFERENCE GENES FROM TRANSCRIPTOME DATA

Query	Acc. num.	Annotation	E-value
EgREF_1	NP_001147321	DNA binding protein (*Zea mays*)	2.00E-11
EgREF_2	NP_195785	Macrophage migration inhibitory factor family protein (*Arabidopsis thaliana*)	2.00E-51
EgREF_3	XP_002319269	Histone 2 (*Populus trichocarpa*)	1.00E-43
EgREF_4	XP_002870907	Macrophage migration inhibitory factor family protein (*Arabidopsis lyrata* subsp. *Lyrate*)	3.00E-52
EgREF_5	NM_001153724	Deleted in split hand / splt foot protein (*Zea mays*)	1.00E-26
EgREF_6	NP_001241759	MKI67 FHA domain-interacting nucleolar phosphoprotein-like (*Zea mays*)	2.00E-66
EgREF_7	NP_001151372	NADH-ubiquinone oxidoreductase 13 kDa-B subunit (*Zea mays*)	9.00E-64
EgREF_8	NP_001146947	60S ribosomal protein L27 (*Zea mays*)	8.00E-59
EgREF_9	NP_178531	OB-fold nucleic acid binding domain-containing protein (*Arabidopsis thaliana*)	1.00E-35
EgREF_10	XP_002863802	Ubiquitin family protein (*Arabidopsis lyrata* subsp. *Lyrate*)	1.00E-33
EgREF_11	NP_001152653	40S ribosomal protein S14 (*Zea mays*)	2.00E-63

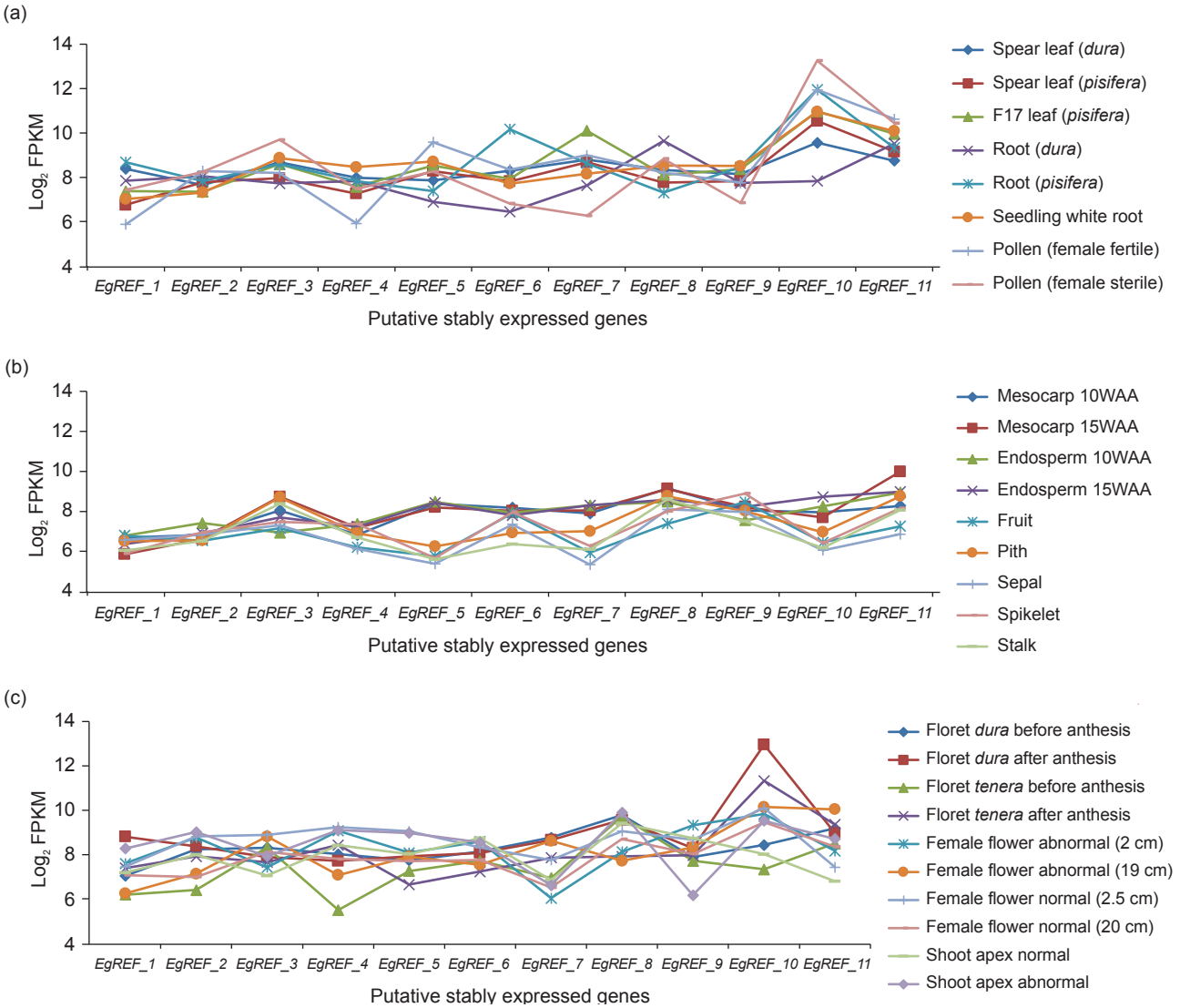

Figure 1. Expression of putative stably expressed genes in various oil palm tissues based on transcriptome data. (a) spear leaf (dura)*, spear leaf* (pisifera)*, F17 leaf* (pisifera)*, root* (dura)*, root* (pisifera)*, seedling white root, pollen (female fertile), pollen (female sterile) (b) mesocarp 10WAA, mesocarp 15WAA, endosperm 10WAA, endosperm 15WAA, fruit, pith, sepal, spikelet, stalk, and (c) floret* dura *after anthesis, floret* dura *after anthesis, floret* tenera *before anthesis, floret* tenera *after anthesis, female flower abnormal (2 cm), female flower abnormal (19 cm), female flower normal (2.5 cm), female flower normal (20 cm), shoot apex normal, shoot apex abnormal.*

TABLE 2. PRIMERS AND AMPLICON CHARACTERISTICS OF SELECTED REFERENCE GENES

Gene abbreviation	Gene description	GenBank ID	Primer sequences (F/R) (5′-3′)	Amplicon length (bp)	Annealing temperature (°C)	Amplification efficiency (%)	R^2
EgREF_3	Histone 2	Pr032825864	ATTTCTCAAGGCCGGCAAGTACG TGCCTCGGGACAATCCTAGTCTT	150	60	85	0.996
EgREF_4	Macrophage migration inhibitory factor family protein	Pr032825863	CCAAGACTGTCGCCAAGCTCATA GCCTCCAATGGAAACCAATTCGC	131	60	92	0.998
EgREF_5	Deleted in split hand / splt foot protein 1	Pr032825865	AGGACGCGAAGATCGACCTCTTT ATCATCCTCCCACTGCTGCATGA	116	60	87	0.997
EgREF_7	NADH-ubiquinone oxidoreductase 13 kDa-B subunit	Pr032825866	GCCAAGATGATCGAATGGGACCC GTCGGTGTTGGGGAACGTGTTTC	103	60	82	0.998
EgREF_10	Ubiquitin family protein	Pr032825867	ACAAGATCCGCATCCAGAAGTGGT GCTGTATTCTGATCAACCAGCC AACC	135	60	83	0.997
EgREF_11	40S ribosomal protein S17	Pr032825868	TCTCCCTCAAGCTTCAGGAGGAG GCATCTCGATGGTCTCCTTGTCG	111	60	87	0.995
PD00380	Predicted 40S ribosomal protein S27-2	EY397675	GATGGTTCTTCCGAACGATATTGA TCACATCCATGAAGAATGAGTTCG	113	60	87	0.995
PD00569	Manganese superoxide dismutase	EL682210	CACCACCAGACGTACATCACAAA GATATGACCTCCGCCATTGAACT	129	63	87	0.999
pOP-EA01332	Predicted protein IFH-1 like	EY406625	AAACGAAGGTACGGCAAGTACA AG CTTAGCACATGCAGAGCAGATGTT	111	63	100	0.979
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	DQ267444	GATCGAGAAATCAGCCACGTATG GTCACCAATAAAGTCGGTGGACA	124	63	87	0.998
TUBULIN	Alpha-tubulin 1	EL685625	CATGGCTTGCTGCCTTATGTATC AGGACACCAGTCAACAAACTGGA	109	63	95	0.999
UBIQUITIN	Polyubiquitin	EL689143	CCAGGCCAATCTCTCAGGATG GGGGGATGCCCTCTTTATCC	130	63	93	0.998
ACTIN	Actin	AY550991	TGCTGATCGTATGAGCAAGGAAA GAAATCCACATCTGCTGGAAGGT	147	60	83	0.999

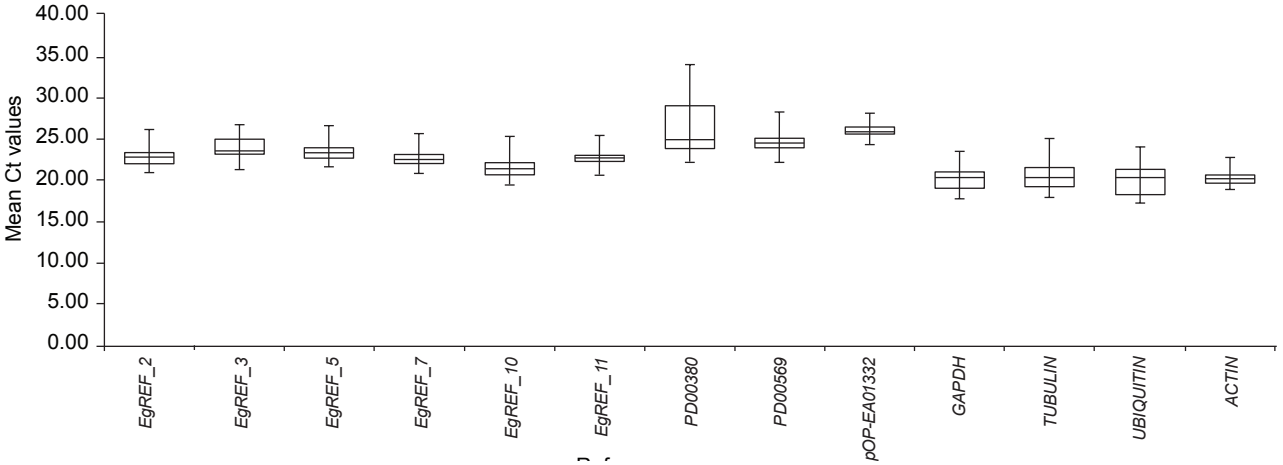

Figure 2. Mean Ct values of candidate reference genes across various oil palm tissues. The range of Ct values was exhibited in boxplot. The box indicates the 25th and 75th percentiles. Horizontal line inside the box is the median. Whiskers below and above the box represent minimum and maximum values. The 15 oil palm tissues tested are spear leaf, mature leaf, white root, lateral root, primary root, 10WAA kernel, 15WAA kernel, 10WAA mesocarp, 15WAA mesocarp, F17 inflorescence (20 cm), F13 inflorescence (8.5 cm), non-embryogenic callus, embryogenic callus, polyembryoids and shoots from the polyembryoids.

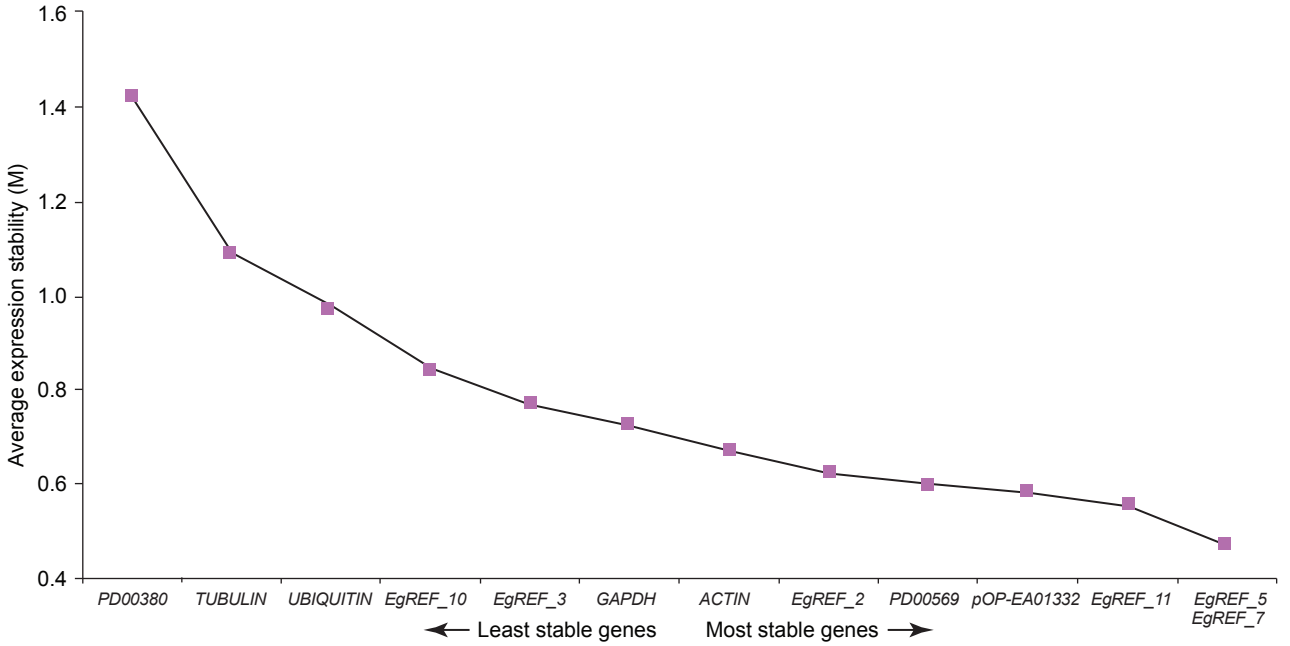

Figure 3. Determination of the most stably expressed reference genes using GeNorm software. Average expression stability values (M) were calculated for each candidate reference gene. Two most stable reference genes were obtained by excluding the least stable genes with higher M values in a stepwise manner.

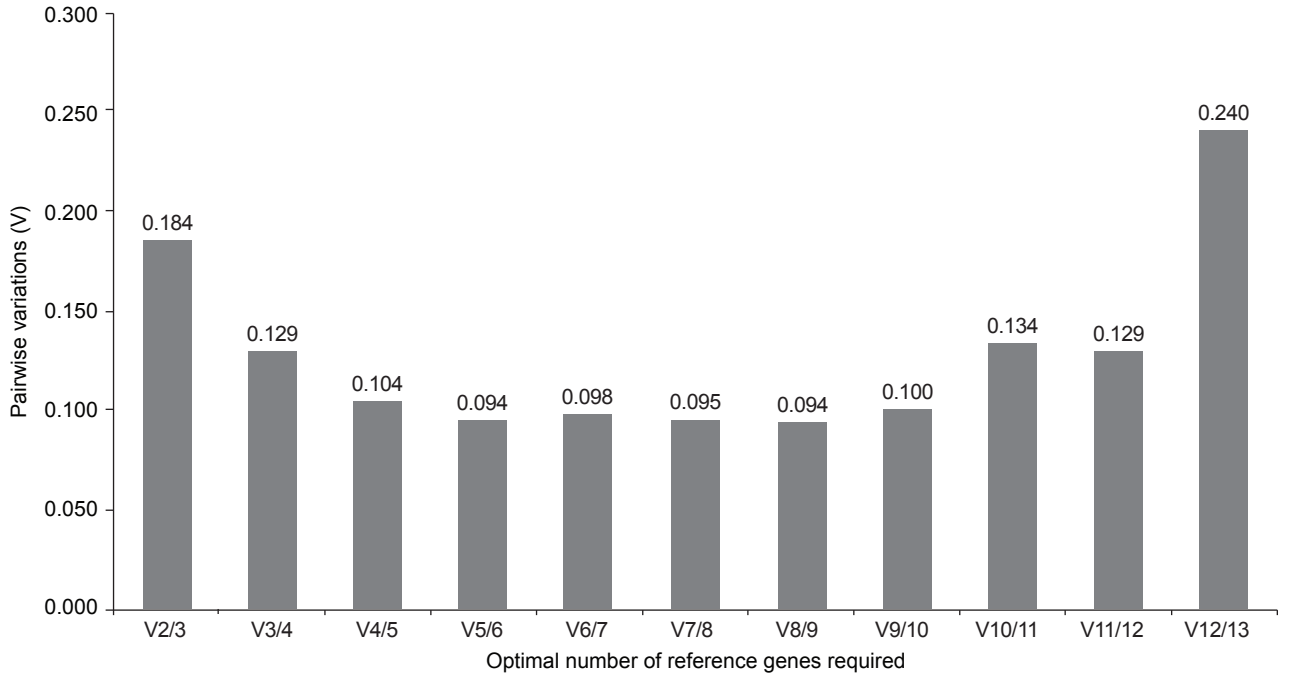

Figure 4. Determination of the optimal number of reference genes for accurate RT-qPCR data normalisation. Thirteen candidate reference genes were tested across 15 oil palm mature plant and tissue culture samples. Pairwise variation, $V_n/_{n+1}$ were calculated between the normalisation factors (NF) of n and n + 1 genes by GeNorm. Additional reference gene is not required if V is lower than the cut-off value of 0.15.

this approach is potentially affected by the quantity and quality of initial materials, first strand cDNA synthesis efficiency, primer performance and statistical analysis methods chosen (Maroufi *et al.*, 2010). Thus, there is a need to normalise raw expression data with stably expressed reference genes for accurate and reliable results. Normalisation is a critical step to compensate for technical variability caused by various steps of the experimental procedures which can affect interpretation of data from RT-qPCR. For this reason, it has become common practice to perform relative quantification by normalising RT-qPCR data to reference genes, as compared to absolute

quantification in which the input copy number is determined by relating the PCR signal to a standard curve for the particular genes of interest. It is also essential to understand that measurement of gene expression patterns may vary under different experimental conditions. Therefore, validation of reference genes should be done according to the panel that is specific for the chosen experimental conditions and tissue types under which the target gene is studied (Ruan and Lai, 2007; Selvey *et al*., 2001; Song *et al*., 2020; Suzuki *et al*., 2000; Thellin *et al*., 1999; Thorrez *et al*., 2008; Zhao *et al*., 2021).

In this study, we analysed transcriptome data from 27 libraries of various oil palm tissues to identify genes with low levels of variation in expression. A total of 11 genes with stable expression across all of the tissues tested were identified. From these genes, *EgREF_10*, annotated as ubiquitin, exhibited the highest level of variation, with highest levels of expression in pollen and floret. Ubiquitins, one of the most commonly used reference genes, are constitutively expressed to maintain cellular function. They are known as housekeeping genes whose protein products are involved in basic cellular processes (Bustin, 2002; Czechowski *et al*., 2005; Dheda *et al*., 2004; Hugget *et al*., 2005). Ubiquitin and small ubiquitin-related modifiers are reported to be involved in post-translational protein modifications by attaching to target proteins to alter their functions (Smalle and Vierstra, 2004). Other functions include protein regulatory activity known as ubiquitination, and ubiquitin-proteasome system, which involves 5% of *Arabidopsis* proteins (Hellman and Estelle, 2002).

In theory, housekeeping genes are assumed to have a constant level of expression and classical housekeeping genes are frequently used and utilised as reference genes for normalisation in RT-qPCR. A number of frequently used classical housekeeping genes that have been validated as suitable reference genes in many plants includes *ACTIN, TUBULIN, UBIQUITIN* and *GAPDH* (Garg *et al*., 2010; Gu *et al*., 2011; Hu *et al*., 2009; Maroufi *et al*., 2010; Nicot *et al*., 2005). *ACTIN*, a widely used reference gene, was selected as the most stable reference gene for normalisation of gene expression data in grapevine leaves (Gutha *et al*., 2010). In papaya, a number of gene expression studies using RT-qPCR had used *ACTIN* as a reference gene (Hernandez *et al*., 2007; Kouzaki *et al*., 2009; Yu *et al*., 2005; 2008).

TUBULIN was also deemed as an appropriate reference gene in banana (Podevin *et al*., 2012). This gene was also identified to be the most suitable reference gene for normalisation across various developmental stages of somatic embryos in two representative conifer species, *Pinus pinaster* and *Picea abies* (De Vega-Bartol *et al*., 2013). As for *UBIQUITIN*, it was found to have stable expression in *A. thaliana* and *Brachypodium* sp. (Czechowski *et al*., 2005; Hong *et al*., 2008). It was also considered a suitable reference gene in sugarcane leaf samples (de Andrade *et al*., 2017) and bioenergy plants (Cheng *et al*., 2019). *GAPDH* on the other hand has been used as a reference gene in citrus, Chinese wolfberry, cotton, chickpea, peanut, grapevine, coffee and sugarcane (Barsalobres-Cavallari *et al*., 2009; de Andrade *et al*., 2017; Garg *et al*., 2010; Mafra *et al*., 2012; Morgante *et al*., 2011; Reid *et al*., 2006; Wang *et al*., 2013a; 2013b).

Among the four classical housekeeping genes, the ranking from the most stable to the least stable gene was: *ACTIN* > *GAPDH* > *UBIQUITIN* > *TUBULIN*. *TUBULIN* and *UBIQUITIN* are two of the five frequently used classical housekeeping genes. However, in our range of samples, these genes performed poorly as well as in tissue culture materials evaluated by Chan *et al.* (2014). It has been reported that *TUBULIN* was found poorly ranked in bamboo, peanut, grape, potato and soybean (Chi *et al*., 2012; Fan *et al*., 2013; Jian *et al*., 2008; Nicot *et al*., 2005; Reid *et al*., 2006), while *UBIQUITIN* performed poorly as a reference gene in soybean (Jian *et al*., 2008) and grape (Reid *et al*., 2006).

In this study, *GAPDH* was ranked in the middle amongst the classical housekeeping genes. However, it had unstable expression across papaya fruit samples (Zhu *et al*., 2012), and was considered the least stably expressed gene during leaf and flower development in petunia (Mallona *et al*., 2010). It also displayed the biggest variation in leaves and roots of *Boehmeria nivea* L. (Yu *et al*., 2020). Based on EST data analysis, this gene was also identified as poorly ranked reference gene in tomato (Coker and Davis, 2003).

In our range of samples, *ACTIN* was observed to be the most stable gene among all classical housekeeping genes studied. However, two published reference genes and four genes identified from transcriptome data were found to be more stably expressed than *ACTIN*. Furthermore, expression stability of *ACTIN* gene family members was observed to differ in peanut in specific conditions (Chi *et al*., 2012; Morgante *et al*., 2011; Reddy *et al*., 2013). These results showed that the transcripts of housekeeping genes can vary in response to experimental conditions and tissue types. Thus, it is important to validate the expression stability of these genes prior to their use in normalisation for RT-qPCR.

Our results showed that three reference genes, *EgREF_5, EgREF_7* and *EgREF_11*, are suitable for normalisation of gene expression data from oil palm tissues. As reported by Szabo *et al.* (2004), valuable reference genes are expressed in several tissues with minimal variation in transcript level across experimental conditions. Genes with low variation in expression with coefficient variation of normalised relative quantities of less than 0.5

are suitable reference gene candidates (Hellemans *et al.*, 2007). The results suggest that it is possible to identify reliable candidate reference genes from multiple transcriptome datasets. *EgREF_5* was annotated as deleted in split hand/splt foot protein 1A, while the other two are NADH-ubiquinone oxidoreductase 13 kDa-B subunit (*EgREF_7*) and 40S ribosomal protein S17 (*EgREF_11*). Expression of these genes were found to be more stable than previously determined reference genes (Chan *et al.*, 2014) and classical housekeeping genes, in the set of diverse oil palm tissues within this study. In recent publications on oil palm reference genes, four genes namely *elF1, elF2, APT* and *cyc*, were identified from transcriptome datasets by Xia *et al.* (2014), of which *elF1* and *elF2* were selected as reference genes for cold, drought and salinity treatments. While *APT* and *cyc* were deemed stable reference genes for drought and salinity stress samples, two genes (*PD00380, PD00569*) were selected as reference genes for tissue culture samples (Chan *et al.*, 2014). These genes were also determined to be good reference genes for oil palm cultured leaf explants (Ooi *et al.*, 2012) and tissue culture samples (Chan *et al.*, 2010). However, in our range of samples, *PD00380* was observed to be the least stable expressed gene. Nevertheless, *pOP-EA01332* and *PD00569* were observed to have more stable expression and were ranked amongst the top five stably expressed genes. Other oil palm reference genes that have also been identified include *GRAS*, *Glutaredoxin, Cyp2* and *SLU7*, where *GRAS* and *Glutaredoxin* are suitable for reproductive tissues, vegetative tissues (*GRAS, Cyp2, SLU7*) and fruit development (*GRAS, Cyp2*) (Yeap *et al.*, 2014). Thus, validation of candidate reference genes is best done under specific experimental conditions and tissue types, as supported by the results of all of these studies.

CONCLUSION

This study shows that systematic analysis of transcriptome data can successfully guide the selection of reference genes for gene expression studies in oil palm, resulting in three housekeeping genes that can be used for RT-qPCR experiment on a wide range of mature plant and tissue culture samples. GeNorm statistical algorithm, identified a combination of three genes, *EgREF_5, EgREF_7* and *EgREF_11* as suitable reference genes for accurate and reliable normalisation across a diverse range of oil palm samples that consist of leaf, root, endosperm, mesocarp, female flowers, non-embryogenic callus, embryogenic callus, polyembryoids and shoots from polyembryoids. Hence, reference genes identified in this study could be used to generate reliable gene expression profile for a wide variety of tissues in oil palm.

ACKNOWLEDGEMENT

The authors would like to thank the Director-General of MPOB for permission to publish this article. We also wish to thank the staff of Genomics Laboratory, Tissue Culture Laboratory and Transgenic Technology Laboratory at the Advanced Biotechnology and Breeding Centre, MPOB for their assistance.

REFERENCES

Adam, H; Jouannic, S; Orieux, Y; Morcillo, F; Richaud, F; Duval, Y and Tregear, J W (2007). Functional characterization of MADS box genes involved in the determination of oil palm flower structure. *J. Exp. Bot., 58*: 1245-1259.

Artico, S; Nardeli, S M; Brihante, O; Grossi-de-sa, M F 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 Biol., 10*: 49.

Barsalobres-Cavallari, C F; Severino, F E; Maluf, M P and Maia, I G (2009). Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. *BMC Mol. Biol., 10*: 1.

Beulé, T; Camps, C; Debiesse, S; Tranchant, C; Dussert, S; Sabau, X; Jaligot, E; Alwee, S and Tregear, J W (2011). Transcriptome analysis reveals differentially expressed genes associated with mantled homeotic flowering abnormality in oil palm (*Elaeis guineensis*). *Tree Genet. Genomes, 7*: 169-182.

Bustin, S A (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. *J. Mol. Endocrinol., 29(1)*: 23-29.

Bustin, S A and Nolan, T (2004). Pitfalls of quantitative real-time reverse transcription polymerase chain reaction. *J. Biomol. Tech., 15*: 155-566.

Bustin, S A; Benes, V; Nolan, T and Pfaffl, M W (2005). Quantitative real-time RT-PCR – A perspective. *J. Mol. Endocrinol., 34*: 597-601.

Bustin, S A; Benes, V; Garson, J A; Hellemans, J; Hugget, J; Kubista, M; Mueller, R; Nolan, T; Pfaffl, M W; Shipley, G L; Vandesompele, J and Wittwer, C T (2009). The MIQE Guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem., 55*: 611-622.

Bourgis, F; Kilaru, A; Cao, X; Ngando Ebongue, G F; Drira, N; Ohlrogge, J B and Arondel, V (2011).

Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. Natl. Acad. Sci., 108*: 12527-12532.

Cappelli, K; Felicetti, M; Capomaccio, S; Spinsanti, G; Silvestrelli, M and Supplizi, A V (2008). Exercise induced stress in horses: Selection of the most stable reference genes for quantitative RT-PCR normalization. *BMC Mol. Biol., 9*: 1.

Chan, K L; Tatarinova, T V; Rosli, R; Amiruddin, N; Azizi, N; Ab Halim, M A; Sanusi, N S N M; Jayanthi, N; Ponomarenko, P; Triska, M; Solovyev, V; Firdaus-Raih, M; Sambanthamurthi, R; Murphy, D and Low, E T L (2017). Evidence-based gene models for structural and functional annotations of the oil palm genome. *Biol. Direct, 12*: 21.

Chan, P L; Ma, L S; Low, E T L; Shariff, E M; Ooi, L C L; Cheah, S C and Singh, R (2010). Normalized embryoid cDNA library of oil palm (*Elaeis guineensis*). *Electron. J. Biotechnol., 13*: 1-17.

Chan, P L; Rose, R J; Abdul Murad, A M; Zainal, Z; Low, E T L; Ooi, C L L; Ooi, S E; Yahya, S and Singh, R (2014). Evaluation of reference genes for quantitative real-time PCR in oil palm elite planting materials propagated by tissue culture. *PLoS ONE, 9(6)*: e99774.

Chandna, R; Augustine, R and Bisht, N C (2012). Evaluation of candidate reference genes for gene expression normalization in *Brassica juncea* using real-time quantitative RT-PCR. *PLoS ONE, 7*: e36918.

Chang, C W; Cheng, W C; Chen, C R; Shu, W Y; Tsai, M L; Huang, C L and Hsu, I C (2011). Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. *PLoS ONE, 6*: e22859.

Cheng, T; Zhu, F; Sheng, J; Zhao, L; Zhou, F; Hu, Z; Diao, Y and Jin, S (2019). Selection of suitable reference genes for quantitative real-time PCR normalization in *Miscanthus lutarioriparia*. *Mol. Biol. Rep., 46*: 4545-4553.

Cheng, W C; Chang, C W; Chen, C R; Tsai, M L; Shu, W Y; Li, C Y and Hsu, I C (2011). Identification of reference genes across physiological states for qRT-PCR through microarray meta-analysis. *PLoS ONE, 6*: e17347.

Chi, X Y; Hu, R B; Yang, Q L; Zhang, X W; Pan, L J; Chen, N; Chen, M; Yang, Z; Wang, T; He, Y and Yu, S (2012). Validation of reference genes for gene expression studies in peanut by quantitative real-time RT-PCR. *Mol. Genet. Genomics, 287*: 167-176.

Coker, J S and Davies, E (2003). Selection of candidate housekeeping controls in tomato plants using EST data. *Biotechniques, 35*: 740-749.

Czechowski, T; Stitt, M; Altmann, T; Udvardi, M K and Scheible, W R (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol., 139*: 5-17.

Dheda, K; Hugget, J F; Bustin, S A; Johnson, M A; Rook, G and Zumla, A (2004). Validation of housekeeping genes for transcript normalizing RNA expression real-time PCR. *Biotechniques, 37(1)*: 112-114, 116, 118-119.

Dheda, K; Hugget, J F; Chang, J S; Kim, L U; Bustin, S A; Johnson, M A; Rook, J A W and Zumla, A (2005). The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal. Biochem., 344*: 141-143.

De Andrade, L M; dos Santos Brito, M; Fávero Peixoto Junior, R; Marchiori, P E R; Nóbile, P M; Martins, A P B; Ribeiro, R V and Creste, S (2017). Reference genes for normalization of qPCR assays in sugarcane plants under water deficit. *Plant Methods, 13*: 28.

De Oliveira, L A; Breton, M C; Bastolla, F M; Camargo, S D; Margis, R; Frazzon, J and Pasquali, G (2012). Reference genes for the normalization of gene expression *Eucalyptus* species. *Plant Cell Physiol., 53(2)*: 405-422.

De Vega-Bartol, J J; Santos, R R; Simões, M and Miguel, C M (2013). Normalizing gene expression by quantitative PCR during somatic embryogenesis in two representative conifer species: *Pinus pinaster* and *Picea abies*. *Plant Cell Rep., 32*: 715-729.

Dussert, S; Guerin, C; Andersson, M; Joët, T; Tranbarger, T J; Pizot, M; Sarah, G; Omore, A; Durand-Gasselin, T and Morcillo, F (2013). Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiol., 162*: 1337-1358.

Exposito-Rodriguez, M; Borges, A A; Borges-Perez, A and Perez, J A (2008). Selection of internal control genes for quantitative real-time PCR studies during tomato development process. *BMC Plant Biol., 8*: 131.

Everaert, C; Luypaert, M; Maag, J L V; Cheng, Q X; Dinger, M E; Hellemans, J and Mestdagh, P (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci. Rep., 7(1)*: 1559.

Fan, C; Ma, J; Guo, Q; Li, X; Wang, H and Lu, M (2013). Selection of reference genes for quantitative real-time PCR in bamboo (*Phyllostachys edulis*). *PLoS ONE, 8*: e56573.

Fleige, S and Pfaffl, M W (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspects of Med., 27*: 126-139.

Garg, R; Sahoo, A; Tyagi, A K and Jain, M (2010). Validation of internal control gene for quantitative gene expression studies in chickpea. *Biochem. Biophys. Res. Commun., 396*: 283-288.

Garrido, J; Aguilar, M and Prieto, P (2020). Identification and validation of reference genes for RT-qPCR normalization in wheat meiosis. *Sci. Rep., 10*: 2726.

Ginzinger, D G (2002). Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Exp. Hematol., 30*: 503-512.

Gu, C; Chen, S; Liu, Z; Shan, H; Luo, H; Guan, Z and Chen, F (2011). Reference gene selection for quantitative real-time PCR in *Chrysanthemum* subjected to biotic and abiotic stress. *Mol. Biotechnol., 49*: 192.

Guénin, S; Mauriat, M; Pelloux, J; Wuytswinkle, O V; Bellini, C and Gutierrez, L (2009). Normalization of qRT-PCR data: The necessity of adopting a systematics, experimental conditions-specific, validation of references. *J. Exp. Bot., 60*: 487-493.

Gutha, L R; Casassa, L F; Harbertson, J F and Naidu, R A (2010). Modulation of flavonoid biosynthetic pathway genes and anthocyanins due to virus infection in grapevine (*Vitis vinifera L.*) leaves. *BMC Plant Biol., 10*: 187.

Hellman, H and Estelle, M (2002). Plant development: Regulation by protein degradation. *Science, 297(5582)*: 793-797.

Hellemans, J; Mortier, G; de Paepe, A; Speleman, F and Vandesompele, J (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol., 8(2)*: R19.

Hernandez, M; Cabrera-Ponce, J L; Fragoso, G; Lopez-Casilla, F; Guevara-Garsia, A; Rosas, G; León-Ramírez, C; Juárez, P; Sánchez-García, G; Cervantes, J; Acero, G; Toledo, A; Cruz, C; Bojalil, R; Herrera-Estrella, L and Sciutto, E (2007). A new highly effective anticysticercosis vaccine expressed in transgenic papaya. *Vaccine, 25*: 4252-4260.

Hu, R B; Fan, C M; Li, H Y; Zhang, Q Z and Fu, Y F (2009). Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol. Biol., 10*: 93.

Hugget, J; Dheda, K; Bustin, S and Zumla, A (2005). Real-time RT-PCR normalization: Strategies and considerations. *Genes Immun., 6(4)*: 279-284.

Hong, S Y; Seo, P J; Yang, M S; Xiang, F and Park, C M (2008). Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR. *BMC Plant Biol., 8*: 112.

Jaligot, E; Adler, S; Debladis, E; Beule, T; Richaud, F; Ilbert, P; Finnegan, E J and Rival, A (2011). Epigenetic imbalance and the floral development abnormality of the *in vitro*-regenerated oil palm *Elaeis guineensis*. *Ann. Bot., 108*: 1453-1462.

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 Mol. Biol., 9*: 59.

Jin, Y; Liu, F; Huang, W; Sun, Q and Huang, X (2019). Identification of reliable reference genes for qRT-PCR in the ephemeral plant *Arabidopsis pumila* based on full-length transcriptome data. *Sci. Rep., 9*: 8408.

Karuppaiya, P; Yan, X X; Liao, W; Wu, J; Chen, F and Tang, L (2017). Identification and validation of superior reference gene for gene expression normalization via RT-qPCR in staminate and pistillate flowers of *Jatropha curcas* – A biodiesel plant. *PLoS ONE, 12(5)*: e0177039.

Kim, D; Pertea, G; Trapnell, C; Pimentel, H; Kelley, R and Salzberg, S L (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol., 14(4)*: R36.

Kong, S L; Abdullah, S N A; Ho, C L; Musa, M H and Yeap, W C (2021). Comparative transcriptome analysis reveals novel insights into transcriptional responses to phosphorus starvation in oil palm (*Elaeis guineensis*) root. *BMC Genom Data, 22*: 6.

Kouzaki, H; O'Grandy, S M; Lawrence, C B and Kita, H (2009). Proteases induce production of thymic stromal lymphopoietin by airway epithelial cells through protease-activated receptor-2. *J. Immunol., 183*: 1427.

Kozera, B and Rapcz, M (2013). Reference genes in real-time PCR. *J. Appl. Genetics, 54*: 391-406.

Kubista, M; Andrade, J M; Bengstsson, M; Forootan, A; Jonák, J; Lind, K; Sindelka, R; Sjöback, R; Sjögreen, B; Ståhlberg, A and Zoric, N (2006). The real-time polymerase chain reaction. *Mol. Aspects Med., 27(2-3)*: 95-125.

Li, Y; Liang, X; Zhou, X; Wu, Z; Yuan, L; Wang, Y and Li, Y (2020). Selection of reference genes for qRT-PCR analysis in medicinal plant *Glycyrrhiza* under abiotic stresses and hormonal treatments. *Plants, 9*: 1441.

Libault, M; Thibivilliers, S; Bilgin, D; Radwan, O; Benitez, M; Clough, S J and Stacey, G (2008). Identification of four soybean reference genes for gene expression normalization. *Plant Genome, 1*: 44-54.

Low, E T L; Alias, H; Boon, S H; Shariff, E M; Tan, C Y A; Ooi, L C L; Cheah, S C; Raha, A R; Wan, K L and Singh, R (2008). Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: Identifying genes associated with callogenesis and embryogenesis. *BMC Plant Biol., 8*: 62.

Low, E T L; Rosli, R; Jayanthi, N; Mohd-Amin, A H; Azizi, N; Chan, K L; Maqbool, N J; Maclean, P; Brauning, R; McCulloch, A; Moraga, R; Ong-Abdullah, M and Singh, R (2014). Analyses of hypomethylated oil palm gene space. *PLoS ONE, 9(1)*: e86728.

Mafra, V; Kubo, K S; Alves-Ferreira, M; Ribeiro-Alves, M; Stuart, R M; Boava, L P; Rodrigues, C M and Machado, M A (2012). Reference genes for accurate transcript normalization in Citrus genotypes under different experimental conditions. *PLoS ONE, 7*: e31263.

Mahoney, D J; Carey, K; Fu, M H; Snow, R; Cameron-Smith, D; Parise, G and Tarnopolsky, M A (2004). Real-time RT-PCR analysis of housekeeping genes in human skeletal muscle following acute exercise. *Physiol. Genomics, 18*: 226-231.

Mallona, I; Lischewski, S; Weiss, J; Hause, B and Egea-Cortines, M (2010). Validation of reference genes for quantitative real-time PCR during leaf and flower development in Petunia hybrid. *BMC Plant Biol., 10*: 4.

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

Marioni, J C; Mason, C E; Mane, S M; Stephens, M and Gilad, Y (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res., 18(9)*: 1509-1517.

Morgante, C V; Guimarães, P M; Martins, A C Q; Araújo, A C G; Leal-Bertioli, S C M; Bertioli, D J and Brasileiro, A C (2011). Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. *BMC Res. Notes, 4*: 339-350.

Mortazavi, A; Williams, B A; McCue, K; Schaeffer, L and Wold, B (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods, 5(7)*: 621-628.

Muhammad Afiq, A H; Shaharuddin, N A and Zubaidah, R (2019). Identification of reliable reference genes for gene expression studies of oil pam plantlets using NormFinder and BestKeeper algorithms. *J. Oil Palm Res., 31*: 204-211.

Nagalakshmi, U; Wang, Z; Waern, K; Shou, C; Raha, D; Gerstein, M and Snyder, M (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science, 320(5881)*: 1344-1349.

Narsai, R; Ivanova, A; Ng, S and Whelan, J (2010). Defining reference genes in *Oryza sativa* using organ development, biotic and abiotic transcriptome datasets. *BMC Plant Biol., 10*: 56.

Nicot, N; Hausman, J F; Hoffman, L and Evers, D (2005). Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J. Exp. Bot., 56(421)*: 2907-2914.

Nolan, T; Hands, R E and Bustin, S A (2006). Quantification of mRNA using real-time RT-PCR. *Nat. Protoc., 1*: 1559-1582.

Ong, P W; Chan, P L and Singh, R (2019). Isolation of high quality total RNA from various tissues of oil palm (*Elaeis guineensis*) for reverse transcription quantitative real-time PCR (RT-qPCR). *J. Oil Palm Res., 31*: 195-203.

Ooi, S E; Choo, C N; Ishak, Z and Ong Abdullah, M (2012). A candidate auxin-responsive expression marker gene, *Eg*IAA9, for somatic embryogenesis in oil palm (*Elaeis guineensis* Jacq.). *Plant Cell Tiss. Organ Cult., 110*: 201-212.

Park, S J; Kim, Y H; Huh, J W; Lee, S R; Kim, S H; Kim, S U; Kim, J S; Jeong, K J; Kim, K M; Kim, H S and Chang, K T (2013). Selection of new appropriate reference genes for RT-qPCR analysis via transcriptome sequencing of cynomolgus monkeys (*Macaca fascicularis*). *PLoS ONE, 8*: e60758.

Podevin, N; Krauss, A; Henri, I; Swennen, R and Remy, S (2012). Selection and validation of reference genes for quantitative RT-PCR expression studies of the non-model crop *Musa*. *Mol. Breeding, 30*: 1237-1252.

Qi, J; Yu, S; Zhang, F; Shen, X; Zhao, X; Yu, Y and Zhang, D (2010). Reference gene selection for real-time quantitative polymerase chain reaction of mRNA transcript levels in Chinese cabbage (*Brassica rapa* L. ssp. pikenensis). *Plant Mol. Biol. Rep., 28*: 597-604.

Radonic, A; Thulke, S; Mackay, I M; Landt, O; Siegert, W and Nitsche, A (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochem. Biophy. Res. Commun., 313*: 856-862.

Reddy, D S; Bhatnagar-Mathur, P; Cindhuri, K S and Sharma, K K (2013). Evaluation and validation of reference genes for normalization of quantitative real-time PCR based gene expression studies in peanut. *PLoS ONE, 8(10)*: e78555.

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 Biol., 6*: 27.

Ruan, W and Lai, M (2007). Actin, a reliable marker of internal control? *Clin. Chim. Acta, 385(1-2)*: 1-5.

Schmittgen, T D and Zakrajsek, B A (2000). Effect of experimental treatment on housekeeping gene expression: Validation by real-time, quantitative RT-PCR. *J. Biochem. Biophys. Methods, 46*: 69-81.

Selvey, S; Thompson, E W; Matthaei, K; Lea, R A; Irving, M G and Griffiths, L R (2001). Beta-actin an unsuitable internal control for RT-PCR. *Mol. Cell. Probes, 15(5)*: 307-311.

Shearman, J R; Jantasurirayat, C; Sangsrakru, D; Yoocha, T; Vannavichit, A; Tragoonrung, S and Tangphatsornruang (2013). Trancriptome analysis of normal and mantled developing oil palm flower and fruit. *Genomics, 101*: 306-312.

Singh, R; Ong-Abdullah, M; Low, E T L; Abdul Manaf, M A; Rosli, R; Nookiah, R; Ooi, C L L; Ooi, S E; Chan, K L; Halim, M A; Azizi, N; Nagappan, J; Bacher, B; Lakey, N; Smith, S W; He, D; Hogan, M; Budiman, M A; Lee, E K; DeSalle, R; Kudrna, D; Goicoechea, J L; Wing, R A; Wilson, R K; Fulton, R S; Ordway, J M; Martienssen, R A and Sambanthamurthi, R (2013). Oil palm genome sequence reveals divergence of infertile species in old and new worlds. *Nature, 500*: 335-339.

Slater, G S C and Birney, E (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics, 6*: 31.

Smalle, J and Vierstra, R D (2004). The ubiquitin 26S proteasome proteolytic pathway. *Annu. Rev. Plant Biol., 55*: 555-590.

Song, H; Mao, W; Duan, Z; Que, Q; Zhou, W; Chen, X and Li, P (2020). Selection and validation of reference genes for measuring gene expression in *Toona ciliata* under different experimental conditions by quantitative real-time PCR analysis. *BMC Plant Biol., 20*: 450.

Suzuki, T; Higgins, P J and Crawford, D R (2000). Control selection for RNA quantification. *Biotechniques, 29(2)*: 332-337.

Szabo, A; Perou, C M; Karaca, M; Perreard, L; Quackenbush, J F and Bernard, P S (2004). Statistical modeling for selecting housekeeper genes. *Genome Biol., 5*: R59.

Tang, N; Zhang, W; Chen, L; Wang, Y and Tang, D (2021). Reference gene selection for real-time quantitative reverse-transcription polymerase chain reaction in flower buds of marigold. *J. Amer. Soc. Hortic. Sci., 146(5)*: DOI: 10.21273/JASHS05074-21.

Tang, X; Zhang, N; Si, H and Calderón-Urrea, A (2017). Selection and validation of reference genes for RT-qPCR analysis in potato under abiotic stress. *Plant Methods, 13*: 85.

Tee, S S; Tan, Y C; Abdullah, F; Ong-Abdullah, M and Ho, C L (2013). Transcriptome of oil palm (*Elaeis guineensis* Jacq.) roots treated with *Ganoderma boninense*. *Tree Genetics and Genomes, 9*: 377-386.

Thellin, O; Zorzi, W; Lakaye, B; De Borman, B; Coumans, B; Henne, G; Grisar, T; Igout, A and Heinen, E (1999). Housekeeping genes as internal standards: Use and limits. *J. Biotechnol., 75*: 291-295.

Thorrez, L; Van Deun, K; Tranchevent, L C; Van Lommel, L; Engelen, K; Marchal, K; Moreau, Y; Mechelen, I V and Schuit, F (2008). Using ribosomal protein genes as reference: A tale of caution. *PLoS ONE, 3(3)*: e1854.

Tranbarger, T J; Dussert, S; Joët, T; Argout, X; Summo, M; Champion, A; Cros, D; Omore, A; Nouy, B and Morcillo, F (2011). Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant Physiol., 156*: 564-584.

Trapnell, C; Williams, B A; Pertea, G; Mortazavi, A; Kwan, G; van Baren, M J; Salzberg, S L; Wold, B J and Pachter, L (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol., 28(5)*: 511-515.

Untergasser, A; Nijveen, H; Rao, X; Biselling, T; Geurts, R and Leunissen, J A M (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res., 35* (Web Server issue): W71-W74.

Vandesompele, J; De Preter, K; Pattyn, F; Poppe, B; Van Roy, N; De Paepe, A and Speleman, F (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol., 3*: research0034.1-0034.11.

Vandesompele, J; Kubista, M and Pfaffl, M W (2009). Reference gene validation software for improved normalization. *Real-time PCR: Current Technology and Applications* (Logan, J; Edward, S K and Saunders, N eds.). Caister Academic Press, London. p. 47-64.

VanGuilder, H D; Vrana, K E and Freeman, W M (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques, 44(5)*: 619-626.

Wang, L; Wang, Y and Zhou, P (2013a). Validation of reference genes for quantitative real-time PCR during Chinese wolfberry fruit development. *Plant Physiol. Biochem., 70*: 304-310.

Wang, M; Wang, Q and Zhang, B (2013b). Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.). *Gene, 530*: 44-50.

Wang, Z; Gerstein, M and Snyder, M (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet., 10*: 57-63.

Xia, W; Mason, A S; Xiao, Y; Liu, Z; Yang, Y; Lei, X; Wu, X; Ma, Z and Peng, M (2014). Analysis of multiple transcriptomes of the African oil palm (*Elaeis guineensis*) to identify reference genes for RT-qPCR. *J. Biotechnol., 184*: 63-73.

Yeap, W C; Loo, J M; Wong, Y C and Kulaveerasingam, H (2014). Evaluation of suitable reference genes for qRT-PCR gene expression normalization in reproductive, vegetative tissues and during fruit development in oil palm. *Plant Cell Tiss. Organ Cult., 116*: 55-66.

Yu, Q; Hou, S; Feltus, F A; Jones, M R; Murray, J E; Veatch, O; Lemke, C; Saw, J H; Moore, R C; Thimmapuram, J; Liu, L; Moore, P H; Alam, M; Jiang, J; Paterson, A H and Ming, R (2008). Low X/Y divergence in four pairs of papaya sex-linked genes. *Plant J., 53*: 124-132.

Yu, Q; Moore, P H; Albert, H H; Roader, A H K and Ming, R (2005). Cloning and characterization of a FLORICAULA/LEAFY ortholog, PFL, in polygamous papaya. *Cell Res., 15*: 576-584.

Yu, Y; Zhang, G; Chen, Y; Bai, Q; Gao, C; Zeng, L; Li, Z; Cheng, Y; Chen, J; Sun, X; Guo, L; Xu, J and Yan, Z (2020). Selection of reference genes for qPCR analyses of gene expression in ramie leaves and roots across eleven abiotic/biotic treatments. *Sci. Rep., 9*: 20004.

Zhao, Z; Zhou H; Nie, Z; Wang, X; Luo, B; Yi, Z; Li, X; Hu, X and Yang, T (2021). Appropriate reference genes for RT-qPCR normalization in various organs of *Anemone flaccida* Fr. Schmidt at different growing stages. *Genes, 12*: 459.

Zhu, J; Zhang, L; Li, W; Han, S; Yang, W and Chen, F (2013). Reference gene selection for quantitative real-time PCR normalization in *Caragana intermedia* under different abiotic stress conditions. *PLoS ONE, 8*: e53196.

Zhu, X; Li, X; Chen, W; Chen, J; Lu, W; Chen, L and Fu, D (2012). Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. *PLoS ONE, 7*: e44405.