Journal of Oil Palm Research Vol. 34 (2) June 2022 p. 261-275
DOI: https://doi.org/10.21894/jopr.2021.0036

IDENTIFICATION OF OIL PALM ROOT-SPECIFIC GENES THROUGH MINING OF RNA-SEQ DATA AND RT-qPCR ANALYSIS

SUBHI SITI MASURA[1,2,*]; OMAR ABD RASID[1]; NOOR AZMI SHAHARUDDIN[2, 3]; MAT YUNUS ABDUL MASANI[1]; KUANG-LIM CHAN[1]; ENG-TI LESLIE LOW[1]; SITI SURIAWATI BADAI[1]; AYUB NOR HANIN[1]; ABDUL RAHMAN SITI RAHMAH[1]; MOHD PUAD ABDULLAH[2]; AZZREENA MOHAMAD AZZEME[2] and GHULAM KADIR AHMAD PARVEEZ[1]

ABSTRACT

Identification of novel genes that are specifically expressed in root is essential for isolation and characterisation of root-specific promoters. Mining the transcriptome of various oil palm tissue-specific data generated from ribonucleic acid-sequencing (RNA-Seq) technology has enabled the discovery of root-specific genes. A total of seven candidates of root-specifically or preferentially expressed genes were selected from RNA-Seq analysis, and the gene expression profiles were validated using real-time quantitative polymerase chain reaction (RT-qPCR). The relative fold change of transcript expression in RT-qPCR was statistically analysed by comparing with root tissues at the in vitro *culture stage (RS1). Results showed that the transcript annotated as an oil palm metallothionine (*EgMT*) gene was significantly upregulated at around 7 to 170-fold across the different developmental stages of root tissues. A proline-rich protein* (EgPRP1) *transcript was also significantly upregulated by about 7 to 55-fold. Both* EgMT *and* EgPRP1 *transcripts had relatively low expressions in the other tissues studied. The high levels of expression of* EgMT *and* EgPRP1 *in roots highlighted the genes' promoter's potential to regulate a strong expression level of transgenes in a root-specific manner.*

Keywords: *Elaeis guineensis*, root-specific genes, RNA-Seq, RT-qPCR.

Received: 17 January 2021; **Accepted:** 18 June 2021; **Published online:** 30 August 2021.

INTRODUCTION

The study of gene expression profiles of cells and tissues using transcriptome data is essential in identifying novel genes. Transcriptome is a complete set of transcripts, including both coding and non-coding ribonucleic acids (RNAs) in a specific type of cell or tissue. Since the last decade, numerous technologies based on hybridisation and sequence-based methods have been developed for generating and quantifying transcriptome. With the advancement of the sequencing method, RNA sequencing (RNA-Seq) technology has emerged as one of the most potent transcript profiling techniques available to date. RNA-Seq provides a more accurate measurement of gene expression and enables the discovery of gene isoforms. Besides, RNA-Seq can also be used to investigate splice sites, alternatively spliced isoforms, small and non-coding RNA (Rivas *et al.,* 2011), single nucleotide polymorphism and post-transcriptional modification (Lalonde *et al.,* 2011). This technology

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

[2] Department of Biochemistry,
Faculty of Biotechnology and Biomolecular Sciences,
Universiti Putra Malaysia, 43400 UPM Serdang,
Selangor, Malaysia.

[3] Institute of Plantation Studies,
Universiti Putra Malaysia, 43400 UPM Serdang,
Selangor, Malaysia.

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

offers a rapid and comprehensive transcript profiling technique with considerably less time and a lower cost (Alpern *et al.*, 2019; Lister *et al.*, 2008; Oikonomopoulos *et al.*, 2020).

The RNA-Seq technology has been applied in numerous crops, including oil palm. As a high yielding source of vegetable oil, palm oil is widely consumed for edible purposes and used as feedstock for oleochemicals and biofuels production. Palm oil accounts for about one-third of world vegetable oil consumption (Kushairi *et al.*, 2018), and the demands will continue to increase due to the growth of the world's population. The effort to genetically engineer oil palm with aims to improve its oil for different oil quality or higher oil yield has been pioneered by the Malaysian Palm Oil Board (MPOB) since the mid-90s (Masani *et al.*, 2018; Masura *et al.*, 2017; Parveez *et al.*, 2015; Rasid *et al.*, 2020). MPOB has reported a breakthrough in oil palm research by deciphering the oil palm genome sequence of the *Pisifera* fruit form of *Elaeis guineensis* (Singh *et al.*, 2013). A total of ~1.5 Gb sequences of the 1.8 Gb genome with the size of 1.05 Mb were released to the public domain (Low *et al.*, 2017; Singh *et al.*, 2013). Transcriptome data were also generated, including those from leaf, inflorescence, pollen, mesocarps, kernel, roots and shoot (http://genomsawit.mpob.gov.my), and this endeavour aids the effort in discovering novel genes of important traits in various tissues of interest, including root.

Protection against drought, increased tolerance to salt, nutrients uptake and increased resistance to pests and diseases are among the useful traits that can be produced through modification of root systems (Li *et al.*, 2019; Potenza *et al.*, 2004). In oil palm, modification of the root system has been focused on producing plant resistance to diseases, particularly basal stem rot that has caused serious economic losses to the oil palm industry. The disease is caused by *Ganoderma boninense* fungus that develops from airborne spores and spreads in the soil through the root (Naher *et al.*, 2013). The adoption of genetic engineering is one of the biotechnological approaches to control or eradicate the spread of the disease. Targeting the expression of fungal resistant genes in oil palm roots could increase the plant defence system against the pathogen. In addition to the disease, the oil palm industry is also to anticipate the effects of climate change in the future that will result in a decline in crude palm oil production (Kushairi *et al.*, 2017). It will likely continue to influence soil properties, which may affect nutrient uptake by the palms (Rival, 2017). Therefore, the improvement of root traits through genetic engineering especially to maximise nutrients and water uptake, would increase crop yields, particularly in unfavourable environments such as under water shortage and low nutrient soils (Meister *et al.*, 2014; Wasson *et al.*, 2012).

To target genetic modification in the root, promoters or regulatory regions that regulate the expression of transgenes in a root-specific manner are required. Several plant root-specific promoters have been isolated and functionally characterised, including *TobRB7, Pyk10, RCc3, PsPR10, MsPRP2, GmPRP2, OsGRP7, Os03g01700* and *Os02g37190* (Chen *et al.*, 2014; Jeong *et al.*, 2010; Liu and Ekramoddoullah, 2003; Nitz *et al.*, 2001; Xue *et al.*, 2016; Yamamoto *et al.*, 1991). However, many studies have shown that the promoter efficiencies in the heterologous system vary considerably, probably due to the absence of some factors essential for promoter regulation (Hernandez-Garcia and Finer, 2014). In oil palm, two root inducible promoters, derived from metallothionine (*MT3-B*) and phosphate transporter (*EgPHT1*) genes have been isolated and characterised (Ahmadi *et al.*, 2018; Zubaidah and Siti Nor Akmar, 2005). *MT3-B* promoter's activity was induced by the presence of metal ions, while *EgPHT1* was induced under inorganic phosphate (Pi) deficiency. Although the strong inducible promoters can be of great benefit, this characteristic could limit their uses as root-specific promoters. As plants have several thousand genes with a vast range of functionalities, it is not surprising that an astonishingly high number of promoters and regulatory elements remain to be discovered (Hernandez-Garcia and Finer, 2014). This effort would lead to significant improvement in the regulation of numerous phenotypes and transgene expression, since a wide range of promoters is available for extensive genetic engineering works.

Prior to promoter isolation, the foremost prerequisite study is to identify genes that are specifically or preferentially expressed in the tissue of interest. The availability of oil palm transcriptome data generated from different tissues has enabled various analyses to be performed, including RNA-Seq analysis. The strategy allows quantification of differentially expressed gene, resulting in discovery of novel genes beneficial for genetic engineering work, including promoter isolation. This study identified two root-specific promoter candidates from the transcriptome data through the RNA-Seq analysis and real-time quantitative polymerase chain reaction (RT-qPCR) analyses. To our knowledge, this is the first study to identify oil palm root-specific or preferentially expressed genes through the mining of oil palm transcriptome data. The discovery would lead to the possession of a well-furnished toolbox of promoters necessary for gene stacking technologies to address more complex agronomic traits.

MATERIALS AND METHODS

Plant Materials

All samples used for RNA isolation were derived from *Elaeis guineensis* (*Tenera*). These include tissues from roots (at different developmental stages), mesocarp, kernel, green leaves, young leaves, inflorescences (male and female), callus, cabbage and plantlets. All samples were collected from an oil palm elite planting material, namely P456 clone. The P456 is a reclone of P164, an MPOB standard clone that produces high oil yield (8-10 t ha^{-1} yr^{-1}). The clone also has high success rates in tissue culturing and meagre mantling rates in the field (Zulkifli *et al.*, 2017).

RNA-Seq Analysis

To identify candidate genes specifically or preferentially expressed in roots, a total of 144 oil palm transcriptome libraries were used for differential expression analysis. The *in vitro* transcript analysis was performed using 27 tissue-specific transcriptome libraries from MPOB that were submitted to GenBank under BioProject PRJNA201497 and PRJNA345530 (Singh *et al.*, 2013), four root libraries from BioProject PRJEB7252 (Ho *et al.*, 2016), 51 libraries from mesocarp at different development stages (Morris *et al.*, 2020) (In-house project B), 28 libraries from different kernel development stages (In-house project B) (unpublished data) and 34 transcriptome libraries of roots (In-house project A). In-house project A contains datasets of roots that were infected with *G. boninense*, *G. boninense* and mycorrhiza, and controls (no infection) (unpublished data). In-house project B contains datasets from mesocarp at 5, 8, 10, 12, 15, 18, 20, 22 and 24 weeks after anthesis (WAA), and kernel at 8, 10, 12, 15 and 18 WAA. All transcriptome libraries were sequenced using Illumina sequencing technique except PRJNA201497 and PRJNA345530, which were generated using Roche/454 GS FLX Titanium (Roche/454) sequencing platform (*Table 1*).

The sequencing raw reads were trimmed using Trim Galore version 0.4.0 with Phred score >20 and length >30 bp. Read-mapping and expression analysis was performed using Tuxedo suite pipeline (Trapnell *et al.*, 2009). The reads from each library were mapped to *E. guineensis* P5-genome build (Singh *et al.*, 2013) using Tophat 2.0.9 with an intron length of 30 bases to 50 kb, followed by assembly using Cufflinks 2.2.1 with default parameters (Trapnell *et al.*, 2010). The assemblies of all the libraries were then merged using Cuffmerge 2.2.1, and the expression data were processed by Cuffdiff 2.2.1. Geometric fragment per kilobase per million mapped reads (FPKM) was calculated in Cuffdiff 2.2.1 to normalise the transcript expression levels. Systematic mining of the transcriptome data to select candidates for the root-specific promoter was carried out using Microsoft Excel. Transcripts that were smaller than 300 bp in length were removed from further analysis.

Isolation of Total RNA

Total RNA was extracted from 24 different oil palm tissues, including root tissues collected from 12 month-old tissue culture plantlets (RS1), 16 month-old plantlets that were planted in jiffy pots in the nursery (RS2), primary and lateral roots from 24 month-old oil palm seedlings (RS3), primary and lateral roots of 10 year-old oil palm (RS4), male inflorescences, female inflorescences, green leaves, young leaves, cabbage, callus, polyembryoids, plantlets (at tissue culture stage and not rooting yet), mesocarps and kernels. Isolation of total RNA of oil palm tissues was carried out using the method described by Zeng and Yang (2002). This is a simple method with modifications of the cetyltrimethyl ammonium bromide (CTAB) buffer and soluble polyvinylpyrrolidone (PVP). The utilisation of CTAB was suitable for RNA isolation of oil palm that contains a high level of phenolic compounds. The addition of PVP into the extraction buffer releases the RNA from lipids as the PVP forms complexes with polysaccharide and polyphenol compounds. The total RNA was dissolved in the

TABLE 1. OIL PALM TRANSCRIPTOME DATA

RNA-Seq Project	Number of trancriptome libraries	Sequencing platform	Layout	Number of reads
In-house project A	34	Illumina	paired-end	28.0-77.0M
In-house project B	79	Illumina	paired-end	15.0-89.0M
BioProject PRJEB7252	4	Illumina	1 paired + 3 single	16.0-19.0M
BioProject PRJNA201497	22	Roche/454	single-end	0.3-0.6M
BioProject PRJNA345530	5	Roche/454	single-end	1.1-1.3M

nuclease-free water and stored at –80ºC. Treatment of total RNA with RNase-free DNase and RNeasy Mini Kit (Qiagen USA, Valencia, CA) was carried out to remove deoxyribonucleic acid (DNA) contamination. NanoDrop ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc.) was used to quantify the RNA quantity and purity. Simultaneously, the RNA integrity and quality were assessed through the RNA Quality Number (RQN) using Fragment Analyser™ (Advanced Analytical Technologies, Inc.).

Real Time-quantitative PCR (RT-qPCR)

First-strand complementary DNA (cDNA) was synthesised using the High-capacity cDNA Reverse-Transcription Kit following the instruction described by the manufacturer (Applied Biosystems). Reverse transcription was carried out using 2 µg of total RNA, which yielded about 100 ng of cDNAs. PCR amplification efficiencies and correlation coefficient (R^2) of each primer pair were calculated using a standard curve generated using a two-fold serial dilution of the pooled cDNAs from roots (2, 4, 6, 8, 16 and 32 ng). The RT-qPCR based on SYBR Green was carried out using the CFX Connect™ Real-Time PCR Detection System (Bio-Rad) in 96-well plates. About 16 ng of cDNA was used in a 20 µL quantitative reaction mix containing 1x iTaq Universal SYBR Green Supermix (2X), 0.5 µM of forward primer and 0.5 µM of reverse primer. The protocol for RT-qPCR is as followed; 95.0°C, 30 min for one cycle and 30 s at 95.0°C and 30 s at 45.0°C (depending on the optimal primer annealing temperature) for 40 cycles. Each sample was analysed in three technical replicates (n=3). The melting curve for each amplicon was obtained from 65.0°C-95.0°C with a 0.5°C increase in temperature at each step. The relative fold difference of expression for each sample in each experiment was determined by normalising the mean cycle quantification (Cq) value for each gene to the mean Cq value of reference genes (Gibberellin-responsive protein 2 (*GRAS*), pre-messenger ribonucleuc acid (mRNA) splicing factor 7 (*SLU7)* and *PD00569*, and calculated relative to a calibrator using the $2^{-\Delta\Delta Cq}$ method (Livak and Schmittgen, 2001). The expression profiles of the transcripts were analysed using BioRad CFX Manager™ 3.0 software (BioRad). Significance fold change of expression for each gene was measured using the Student t-test with $p<0.01$.

Sequence Analysis

Nucleotide sequences were annotated using GenBank Plant Reference Sequence (RefSeq) Database (OLeary *et al.*, 2016) via BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al.*, 1997) with default parameters. Functional gene annotation was performed by searching the amino acid or protein sequence homology in a non-redundant RefSeq protein database (Pruitt *et al.*, 2005) by using BLASTx (https://blast.ncbi.nlm.nih.gov/Blast.cgi) alignments with an e-value threshold of 1^{e-5}. Nucleotide and amino acid sequences of the targeted gene and its counterpart from other plants that were retrieved from GenBank were deposited to Vector NTI® software (Thermo Fisher Scientific Inc.) (Lu and Moriyama, 2004). The open reading frame of the targeted gene was translated to the corresponding amino acid sequence by using a translation tool in the software. Then, multiple alignments of nucleotides or amino acid sequences of the targeted gene and its counterpart were performed using AlignX tool based on the ClustalW algorithm.

RESULTS AND DISCUSSION

Identification of Putative Root-specific Genes via RNA-Seq Analysis

RNA-Seq data from the root, shoot, fruit, inflorescence, pollen, leaf, mesocarp, kernel, pith, sepal, spikelet, and stalk were used to screen for candidates for root-specific or preferentially expressed genes. The transcripts were mapped to MPOB's *AVROS Pisifera* genome P5-build. A total of 51 889 genes with 165 751 isoforms were obtained from the RNA-Seq analysis. Transcripts smaller than 300 bp in length were removed, as these sequences could result in slight over-estimation of the expression abundance, which could lead to misinterpretation in the data analysis. The short reads could have arisen from incomplete contig assembly (Hsieh *et al.*, 2019) or belong to the small or non-coding RNA (Liu *et al.*, 2019). The analysis resulted in the discovery of genes that were predominantly but not specifically expressed in the root. Among the 159 490 filtered transcripts, seven transcripts (TCONS_00011027, TCONS_00000859, TCONS_00140324,TCONS_00083022,TCONS_00044801, TCONS_00110826, and TCONS_00034877) were found to be the most highly expressed in root tissues. These transcripts only had low levels of expression detected in other tissues (*Figure 1*) and were therefore chosen for further analysis. It is noteworthy from the findings of Evans *et al.* (1988) that showed that the efforts in isolating root-specific cDNA clones from pea were unsuccessful and that they concluded that root-specific mRNA species (if present) were only present at very low levels of abundance in root mRNA populations. Choosing a transcript that has low levels of expression may not be suitable for the identification of promoters to regulate transgene expression, as promoters that possess strong activity are desired for this purpose.

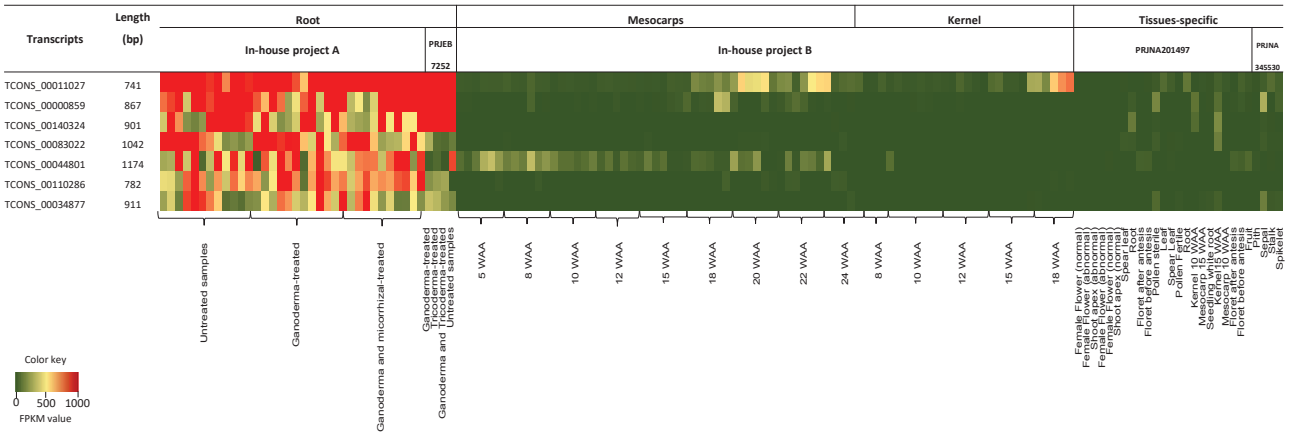

Figure 1. Mining of oil palm transcriptome data using the RNA-Seq approach. A total of seven candidate root-specific or root preferential expressed genes were selected from the in silico *analysis. Transcriptome libraries were derived from GenBank BioProject PRJNA201497, PRJNA345530, PRJEB7252, root (In-house project A), and mesocarp and kernel from In-house project B. Red and green shades indicate higher and lower expression, respectively. The colour key indicates the intensity associated with normalised expression value using the FPKM method.*

High-quality RNA for RT-qPCR

The expression pattern of the genes selected through RNA-Seq analysis was further validated using RT-qPCR analysis. To perform RT-qPCR, the isolation of high-quality total RNA is essential. In this study, total RNA was extracted from 24 different oil palm tissues including root tissues at different developmental stages (RS1, RS2, RS3 and RS4), male inflorescence, female inflorescence, green leaves, young leaves, cabbage, callus, polyembryoids, mesocarp and kernel. About 3.84-31.35 µg g^{-1} fresh weight tissue of total RNA was obtained. The total RNA was of high purity, as the A_{260}/A_{280} and A_{230}/A_{260} ratios for all samples were greater than 1.8, indicating the absence of protein and other organic compounds (Claros and Canovas, 1999). RNA integrity and quality were also good, as the RQN values were relatively high, ranging from 7 to 9 (*Table 2*).

Tissue Specificity Analysis through RT-qPCR Analysis

A standard curve of PCR amplification efficiency and R^2 was generated for each reference and target gene. The amplification efficiency value for all reference and target genes tested was within the range of 90% and 110%, while the R^2 value was >0.98, indicating a positive correlation between the amount of cDNA template and the cycle threshold (Ct) values (Bustin *et al.*, 2009). *Table 3* shows the primers used for target genes and the value of amplification efficiency and R^2 of the standard curve generated for RT-qPCR. The temporal and spatial expressions of targeted genes were evaluated across the 24 different oil palm tissues. The quantitative data for gene expression was normalised to the expression level of reference genes, namely *GRAS*, *SLU7* and *PD00569*, that have been documented as stable reference genes in oil palm (Chan *et al.*, 2014; Yeap *et al.*, 2014). The root tissue at the early development stage (RS1) was used as a calibrator. The selection of RS1 as a calibrator will give a better understanding of transcripts expression patterns in roots. The expression profiles of the genes can be measured from the earlier to the later stages of root development, while a significant comparison to other tissues (non-root) will give a good indication of their specificity.

Expression analysis of the putative root-specific genes using RT-qPCR is shown in *Figure 2*. The relative fold change of expression was measured in $\log_2$ ratio and statistically validated using t-test with $p<0.01$. The results showed that TCONS_00011027 was highly expressed in RS1 as the average Cq value was detected at 16. The transcript was highly expressed at the early stages of root development as no fold change in expression was observed in RS1, RS2 and lateral roots of RS3. Although the gene was significantly downregulated (~6 to 42-fold) at the later stages of root development [RS3 (primary root) and RS4], its expression was still relatively high, as the RT-qPCR amplification plots showed average Cq values at 16 to 22 cycles. Likewise, the transcript also showed a significant downregulation in the non-root tissues. However, the transcript's expression level in callus, polyembryoids, young and green leaves overlaps the expression in roots, indicates that it is not a root-specific or root-preferentially expressed gene.

TABLE 2. YIELD AND PURITY OF TOTAL RNA ISOLATED FROM VARIOUS OIL PALM TISSUES

Sample	Yield (μg g^{-1})	A260/A280	A260/A230	RNA quality number (RQN)
C	6.29	2.16	2.13	8.8
CB	19.90	2.06	2.01	8.9
FI	10.27	2.08	1.98	7.8
MI	9.19	2.10	1.95	8.2
GL	15.78	2.14	2.00	8.4
YL	13.92	2.19	2.18	9.0
K15	10.22	2.08	1.80	8.8
M15	19.20	2.06	1.96	9.0
PE	6.10	2.10	1.87	8.6
PL	19.60	2.07	2.11	9.0
RS1	3.84	2.05	1.80	7.6
RS2	4.78	2.05	2.10	7.0
RS3 (LR1)	11.40	2.15	2.27	7.9
RS3 (LR2)	28.58	2.08	1.83	7.6
RS3 (LR3)	30.73	2.06	2.03	7.7
RS3 (PR1)	21.08	2.11	2.27	8.8
RS3 (PR2)	31.35	2.02	2.18	9.0
RS3 (PR3)	25.95	2.10	2.23	8.6
RS4 (LR1)	15.80	2.14	1.81	8.4
RS4 (LR2)	24.22	2.09	2.29	7.4
RS4 (LR3)	31.06	2.02	2.21	8.6
RS4 (PR1)	24.47	2.10	2.28	7.4
RS4 (PR2)	12.51	2.15	2.25	7.0
RS4 (PR3)	14.70	2.14	1.80	7.4

Note: C - callus; CB - cabbage; FI - female inflorescence; MI - male inflorescence; GL - green leaves; YL - young leaves; K15 - kernel 15 WAA; M15 - mesocarp 15 WAA; PE - polyembryoids; PL - plantlets; RS1 - root from plantlets (12 months); RS2 - root planted in jiffy pot (16 months); RS3 (LR1), RS3 (LR2), RS3 (LR3), RS3 (PR1), RS3 (PR2) and RS3 (PR3) - lateral and primary roots in polybag (24 months); and RS4 (LR1), RS4 (LR2) and RS4 (LR3), RS4 (PR1), RS4 (PR2) and RS4 (PR3) - lateral and primary roots at field planting (10 years).

TABLE 3. PRIMERS OF PUTATIVE ROOT SPECIFIC GENES FOR RT-qPCR

Targeted gene	Primer sequence (5′-3′)	Amplification length (bp)	Annealing temperature (°C)	Amplification efficiency (%)	Regression coefficient (R^2)
TCONS_00011027	F-TTGGTTGTTGTAGTTCTTCATATTAG R-GGTGCTGGTCTTCTCAGCCA	140	48	96.8	0.992
TCONS_00000859	F-GCTTCGGCATTGTGACACT R-GCAGTTGGAGCTGCACTTGC	101	50	93.7	0.984
TCONS_00140324	F-TGGAAAATGGCTTCCAAGTC R-GATCCAGAAGAAGGTGAGGG	146	48	97.1	0.980
TCONS_00083022	F-GCCTAGGAAACAATCAAGTTTAACG R-ATACCGGCGGCGCTTGCGATGACATT	155	45	105.4	0.993
TCONS_00044801	F-CACAAATTTCAGACAAGCAGC R-CAGACTTTCTCACAGACACAGAACA	156	47	90.9	0.983
TCONS_00110826	F-AAGATGAGACGCCACAA R-CAGACTTTCTCACAGACACAGAACA	149	51	92.1	0.993
TCONS_00034877	F-CCGGACACAAACCACCAACT R-GGCTTCTCATACGGTTTGGG	139	45	91.1	0.989

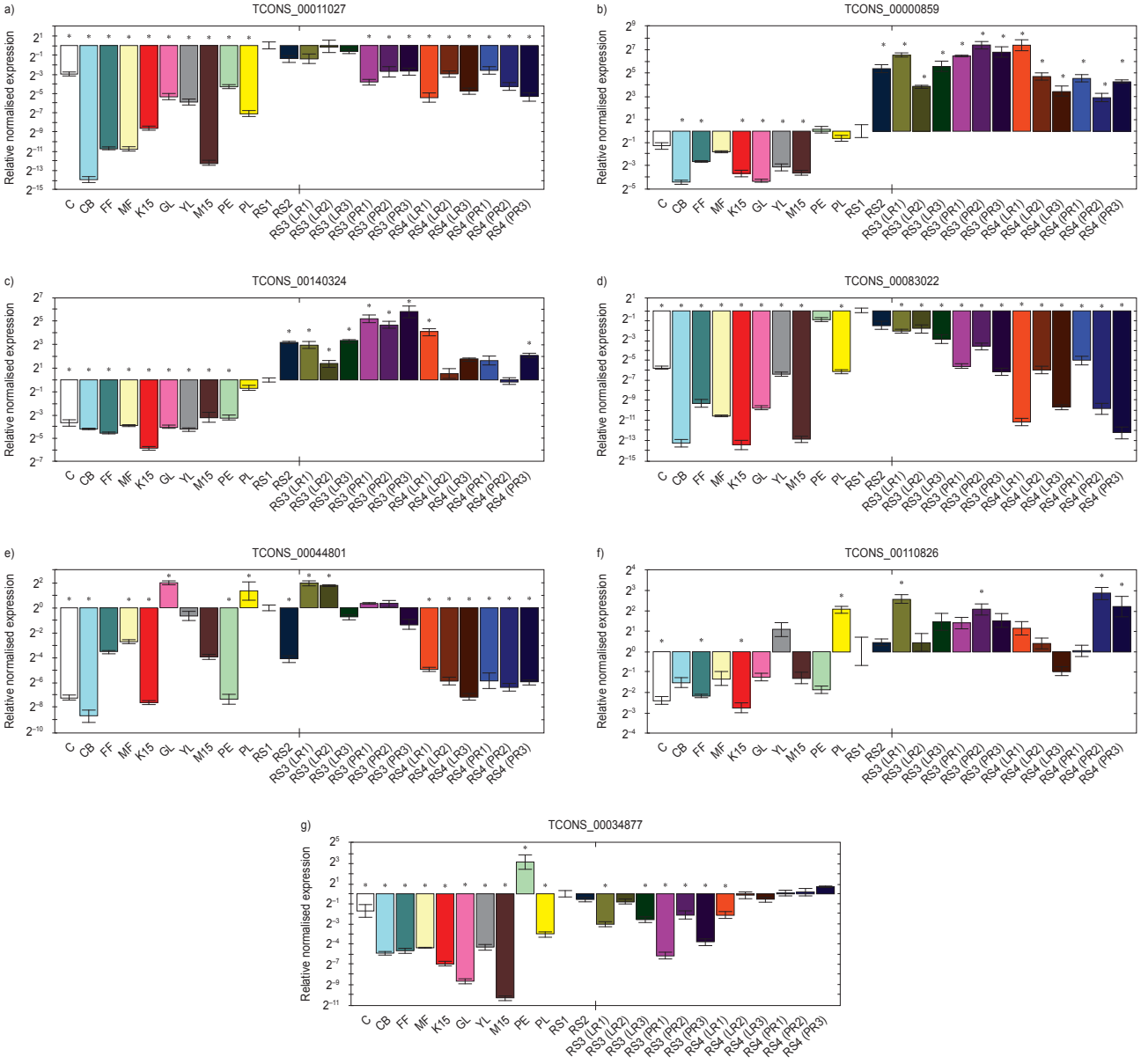

Figure 2. Expression analysis of putative root-specific genes using RT-qPCR. Graphs indicate the expression data of a) TCONS_00011027, b) TCONS_00000859, c) TCONS_00140324, d) TCONS_00083022, e) TCONS_ 00044801, f) TCONS_00110826, and g) TCONS_00034877. Y-axis shows the average relative expression of the transcripts calculated using $2^{-\Delta\Delta Cq}$ against reference genes. Expression levels of the transcripts were compared with RS1 as a calibrator. The average quantification cycle (Cq) value of RS1 for TCONS_00011027, TCONS_00000859, TCONS_00140324, TCONS_00083022, TCONS_00044801, TCONS_00110826, and TCONS_00034877 were 16, 27, 24, 19, 20, 27 and 16, respectively. Three technical replicates were used for each biological sample. Three biological samples were used for primary and lateral roots in polybag (24 months) [RS3 (PR1), RS3 (PR2), RS3 (PR3), RS3 (LR1), RS3 (LR2) and RS3 (LR3)] and primary and lateral roots at field planting (10 years) [RS4 (PR1), RS4 (PR2), RS4 (PR3), RS4 (LR1), RS4 (LR2) and RS4 (LR3)] and one biological sample used for root from plantlets (12 months) (RS1) and nursery stage (planted in jiffy pot (16 months) (RS2), callus (C), male inflorescence (MI), female inflorescence (FI), green leaves (GL), young leaves (YL), cabbage (CB), polyembryoids (PE), plantlets (PL), mesocarp at 15 WAA (M15) and kernel at 15WAA (K15). Error bars indicate the standard error of the mean of three technical replicates. Asterisks indicate the significance fold change of transcript expression compared to RS1 as a calibrator (t-test: $p<0.01$).

On the other hand, TCONS_00000859 had a low expression level in RS1, as the RT-qPCR amplification plots showed average Cq values at 27. However, the gene expression was markedly increased across the different developmental stages of root with a significant upregulation of around 7 to 170-fold. It was notable that the expression of the transcript in non-root tissues was relatively low compared to most of the root tissues. All the non-root tissues had either no significant fold change in comparison to RS1 (callus, male inflorescence, polyembryoids and plantlets) or was significantly downregulated by 6 to 21-fold (cabbage, female inflorescence, kernel, leaves, and mesocarp). A similar gene expression pattern was seen for TCONS_00140324, which had a moderate

level of expression in RS1 (average Cq value of 24). In addition to a comparable expression of TCONS_00140324 in the later stages of root development [RS4(LR2), RS4(LR3), RS4(PR1) and RS4(PR2)], a significant upregulation of the transcript at around 7 to 55-fold was observed across other developmental stages of the root. Apart from plantlets that showed no differential expression with RS1, the transcript was downregulated in callus, polyembryoids, cabbage, inflorescence, kernel, mesocarp and leaves by about 9 to 57-fold. These results were generally in concordance with RNA-Seq data that showed that TCONS_00140324 and TCONS_00000859 transcripts were observed in abundance in roots, with low or barely detectable expression in other tissues.

Over and above, we found that the TCONS_00083022 and TCONS_00044801 transcripts were highly expressed in RS1 as an average amplification plot of Cq was detected at 19 and 20 cycles, respectively. However, TCONS_00083022 was significantly downregulated across all tissues tested, except in polyembryoids that showed no significant fold change in expression. For TCONS_00044801, the transcript was only upregulated in green leaves, plantlets and RS3 (LR1 and LR2) by about 4-fold while significantly downregulated in other tissues, including RS2 and RS4. Overall, the expression levels of these two transcripts in root tissues overlap with the expression levels in the non-root tissues, indicating that the genes are not preferentially expressed in root tissues.

For TCONS_00110826, a noticeable low expression of the transcript was observed in RS1 with an average Cq value of 27. The transcript was significantly upregulated in root tissues of RS3 (LR1 and PR2), RS4 (PR2 and PR3) and plantlets at around 4 to 6-fold, while significant downregulation was observed in callus, female inflorescence and kernel. No significant fold change in expressions was observed in the other tissues studied, indicating that the gene expression was constant but relatively low, even in the root development samples. For TCONS_00034877, the gene was expressed in RS1 with an average Cq value of 16. Although the RT-qPCR profiles in root had indicated either downregulation of around 4 to 33-fold or no significant fold change in expression, the expression levels of the gene in root tissues overlaps with the majority of the non-root tissues. The gene seems to be constitutively expressed, suggesting a possible housekeeping role in oil palm tissues.

Although the RNA-Seq and RT-qPCR data were generally in agreement, particularly for expression profiles of TCONS_00000859 and TCONS_00140324, some results differed. Discrepancies between the RNA-Seq and RT-qPCR data were observed, particularly for TCONS_00083022, TCONS_00044801, TCONS_00110826, and TCONS_00034877. The differences could be attributed to the different biological samples or materials used in both platforms. The background of the biological samples used for the RNA-Seq experiments varies, as the samples came from *Tenera, Dura* or *Pisifera* palms. In contrast, the RT-qPCR experiments were conducted using tissue culture-derived ramets of P456 clone (*Tenera* palm) (Zulkifli *et al.*, 2017). The variability of biological materials derived from different genetic backgrounds might contribute to the variation in expression quantification.

Further validation of the transcript expression profiles through RT-qPCR is essential and fundamental as this method is highly sensitive for gene quantification and can be highly sequence-specific (Costa *et al.*, 2013). To further validate and obtain an accurate result, the RT-qPCR was conducted using biological materials derived from the same genetic background with three technical replicates for each sample. This has increased RT-qPCR precision, improved experimental variation, and served to improve confidence as a better estimation of the mean is provided by the technique (Sanders *et al.*, 2014).

Selection of Putative Root-specific Genes

Based on spatial and temporal expression measured by RT-qPCR, the high abundance of TCONS_00011027 and TCONS_00034877 transcripts were not only observed in the root but also the other tissues studied. Based on the annotation to the non-redundant RefSeq protein database in Genbank, TCONS_00011027 and TCONS_00034877 were similar to metallothionine (XM_010924034.3) and early nodulin-75-like genes (ENOD2) (XM_010942463.1), respectively. In legume, the early nodulin-75-like ortholog was involved in the *early* stages of root nodule development (Franssen *et al.*, 1987). The gene codes for a proline-rich protein, which is a part of the structural protein component of the plant cell wall. Strong expression of the gene in oil palm may coincide with its role as structural wall protein involved in important developmental processes, such as vascular differentiation, wound healing, or defence response against pathogens (Wilson *et al.*, 1994). However, it is notable that early nodulin-like proteins belong to a multigene family. Although nodulin genes had already been cloned and sequenced, the functions for many of them are sparsely described (Tikhonovich and Provorov, 2007).

For TCONS_00083022, the gene showed higher accumulation in tissue culture samples (callus, polyembryoids and RS1) and the earlier stages of root development (RS2 and RS3). While lower

levels of expression were detected in RS4 and the other tissues studied. The gene is similar to pathogenesis-related protein 1 (*PR-1*) (Genbank accession no: XM_010940037.3) that belongs to a multigene family. Pathogenesis-related proteins (PR) are a structurally diverse group of plant proteins that are toxic to invading fungal pathogens (Agrios, 2005). PR-1 protein has been reported to have antimicrobial activity.

In addition to the pathogen attack response, *PR-1* genes are also responsive to abiotic stimuli, suggesting their important roles in abiotic stress response (Breen *et al.*, 2017). The elevated expression of TCONS_00083022 in tissue culture samples (callus, polyembryoids) RS1, RS2 and RS3 may be associated with abiotic stimuli. Controlled stress in *in vitro* cultures, such as mechanical injuries, oxidative stress, and high plant growth regulator concentrations, may stimulate stress-related genes, such as *PR-1*. At the RS2 and RS3 stages, oil palm seedlings in polybags may also encounter growth stresses, such as water deficit, dry heat and low humidity that could interfere with the root elongation process. In contrast to RS4, which is planted in peat soil, the high-water table and loose soil structure of peat soil may have made it easier for root growth and elongation, likely reducing the induction of stress-related genes. Interestingly, as the gene shows expression patterns related to plant defence response, its promoter may be inducible, making it useful to fine-tune the expression of transgenes in response to pathogen attack or abiotic stresses. On the other hand, TCONS_00044801 showed varying expression levels at different fold changes across all the tissues tested, while TCONS_00110826 showed relatively low levels of expression. The roles and regulation of these genes are yet to be understood as they are classified as proteins of unknown function.

This study showed that TCONS_00000859 and TCONS_00140324 transcripts were highly expressed in the root and barely detected in other tissues. As the results demonstrate the potential of their promoters to drive strong expression of transgenes in a root-specific manner, the sequences of TCONS_00000859 and TCONS_00140324 were analysed in detail. The first gene, named *EgMT* codes for the TCONS_00000859 transcript, had significant similarity to an oil palm metallothionein sequence (Genbank accession no: MK557924.1). The gene belongs to the Class II metallothionein (MT) gene family. *EgMT* has a 192 bp open reading frame that encodes a 63 amino acid polypeptide with a theoretical molecular mass of 6.58 kDa and a pI of 4.65. Multiple amino acid alignments conducted on *EgMT* and its counterparts from other plants such as from *Asparagus officinalis* (XP_020267036.1), *Fritillaria agrestis* (AAB95221.1), *Dracaena cambodiana* (ASR83111.1), *Ananus comosus* (OAY84410.1) and *Metroxylon sagu* (ABA43635.1), as shown in *Figure 3*, indicates that sequence similarity was around 64.06%-70.77%. A detailed comparison analysis was also carried out between *EgMT* and *MT3-B* sequences (another oil palm root-specific promoter). Using a pairwise sequence alignment tool (https://www.ebi.ac.uk/Tools/psa/emboss_needle/), results showed that these genes share about 89.60% similarity in their coding regions and 21.70% similarity in both their 5′ and 3′ non-coding regions. This data indicates that *EgMT* belongs to another family of oil palm MT genes. Based on nucleotide search to the National Center for Biotechnology Information (NCBI), oil palm has at least five MT genes that belong to different types of MT (data not shown). The detailed analysis of *EgMT* indicated that it contains the C-X-C motif, which is essential for effective metal binding. This cysteine-rich metal-binding protein is vital as MTs are involved in various cellular functions such as protection against oxidative stress, zinc and copper homeostasis, and buffering against toxic heavy metals (Joshi *et al.*, 2016; Shabb *et al.*, 2017).

Determination of spatial and temporal expression of *EgMT* through RT-qPCR indicated that *EgMT* is one of the MT types that are preferentially expressed in the root. The expression of *EgMT* was in contrast to a metallothionine coded by TCONS_00011027 that showed high abundance in root and other tissues including callus, polyembryoids, young and green leaves. Cobbett and Goldsbrough (2002) reported that MTs of higher plants were classified into several types (Foley and Singh, 1994; Guo *et al.*, 2003; Reid and Ross, 1997). The diverse expression patterns of different MT genes suggest that plant MT isoforms may differ in sequence and in the functions they perform in specific tissues (Cobbett and Goldsbrough, 2002). Although many genes encoding MTs have been isolated and characterised, their precise functions and regulation are not entirely understood (García-Hernández *et al.*, 1998). Interestingly, in addition to oil palm *MT3-B* that has been documented as an inducible root promoter, Dong *et al.* (2010) also reported *OsMT-I-4b* as an inducible root promoter from rice. The potential of *EgMT* promoter to regulate transgenes in a root-specific manner is evident by the abundance of its transcript in root based on RT-qPCR and RNA-Seq. However, further characterisation of the promoter via deletion analysis may reveal the critical regulatory regions of *EgMT* that are essential for root-specific regulation.

TCONS_00140324, coded by a gene named *EgPRP1*, is similar to a 14kD proline-rich protein (Genbank accession no: XM_010908304.3). In general, proline-rich proteins belong to the hybrid proline- or glycine-rich protein (HyP/GRP) gene family that functions as plant-specific and putative

cell-wall/plasma membrane-associated proteins (Fujino *et al.*, 2014). The EgPRP1 protein consists of 132 amino acids, including 14 (11.45%) proline residues, with a theoretical molecular mass of 13.44 kDa and a pI of 9.27. The protein also consists of a hydrophobic domain at the N terminus that represents a signal sequence. The EgPRP1 protein sequence is most similar (63.91%-93.02% identical) to a set of putative cell wall-localised proline-rich proteins isolated from several plant species, such as *Phoenix dactylifera* (XP_008808738.1), *A. comosus* (XP_020085092.1), *Glycine max* (XP_003525817.1), *Citrus sinensis* (XM_006477171.2) and *Sesamum indicum* (XP_011081229.1) (*Figure 4*).

Figure 3. Multiple sequence alignment of EgMT *with various homologous sequences retrieved by BLASTx analysis. The* EgMT *gene showed homology to metallothionein genes from* Asparagus officinalis *(XP_020267036.1)[AoMT],* Fritillaria agrestis *(AAB95221.1) [FaMT],* Dracaena cambodiana *(ASR83111.1)[DcMT],* Ananus comosus *(OAY84410.1) [AcMT] and* Metroxylon sagu *(ABA43635.1)[MsMT] with sequence similarity of around 64.06%-70.77%. Multiple sequence alignment showed conserved C-X-C motifs (highlighted in box). Gaps introduced for best alignment are shown as dash.*

Figure 4. Multiple sequence alignment of EgPRP1 *with various homologous sequences retrieved by BLASTx analysis. The* EgPRP1 *gene showed homology to proline-rich protein genes from* Ananas comosus *(XP_020085092.1)[AcPRP],* Phoenix dactylifera *(XP_008808738.1)[PdPRP],* Glycine max *(XP_003525817.1)[GmPRP],* Citrus sinensis *(XM_006477171.2) [CsPRP]) and* Sesamum indicum *(XP_011081229.1)[SePRP], with sequence similarity of around 63.91%-93.02%. Gaps introduced for best alignment are shown as dash.*

A similar expression pattern as *EgMT* was observed in the RT-qPCR results. High levels of expression of *EgPRP1* were markedly detected in roots. The high activity *of EgPRP1* in roots may correlate to the PRP function involved in cell and root elongation. Several studies have described the discovery of PRP genes with specific or preferential expression in root. For examples, the specific expression of *Nicotiana tobaccum* and soybean hydroxyproline-rich glycoprotein in the endodermal cells of the zone that lateral roots emerge, are necessary to provide the mechanical strength required for penetrating through the main root, as the genes are considered to be involved in cell wall reformation (Ahn *et al.,* 1996) and hardening (Keller and Lamb, 1989). Several studies also reported that rice *RCc2* and *RCc3*, alfalfa *A9*, maize *ZRP3*, carrot *PRP1*, bean *PVR5*, *Vitis vinifera PRP2*, *Medicago sativa PRP2*, soybean *PRP1* and *PRP2* genes, which all encode a proline-rich protein, are expressed preferentially in the root (Choi *et al.,* 1996; Ebener *et al.,* 1993; Jeong *et al.,* 2010; Winicov *et al.,* 2004; Xu *et al.,* 1995). This indicates that the findings of *EgPRP1* are in agreement with these previous studies. On the other hand, a unique expression of *EgPRP1* has also been detected in the plantlets. The long duration of plantlet cultivation in tissue culture media triggered various stresses that may lead to the expression of *EgPRP1*. The result is in accordance with previous researches that showed abiotic and biotic stresses influenced the expression of the PRP genes (Bhattacharya *et al.,* 2013; Srinath *et al.,* 2017; Suzuki *et al.,* 1993). Nonetheless, the comprehensive studies of PRPs have revealed that PRPs act in various plant growth processes to meet the functional and physical requirements of different cell types at different developmental stages (Josè-Estanyol *et al.,* 1992; Showalter and Rumeau, 1990). For example, in addition to the exclusive expression of *AtPRP1* and *AtPRP3* in roots, other *Arabidopsis* PRPs, namely *PRP2* and *PRP4* transcripts, were abundant in the plant's aerial organs (Bernhardt and Tierney, 2000; Fowler *et al.,* 1999). PRPs' promoter's ability in driving the gene of interest or reporter genes has also been studied. Several reports documented that *RCc3, AtPRP1, AtPRP3, MsPRP2* and *GmPRP2* can drive the expression of transgenes in a root-specific manner (Chen *et al.,* 2014; Fowler *et al.,* 1999; Jeong *et al.,* 2010; Li *et al.,* 2019; Winicov *et al.,* 2004).

CONCLUSION

The transcript profiling studies through RNA-Seq have accelerated the discovery of novel root-specific or preferential genes for promoter isolation. RNA-Seq data analysis identified seven candidate genes that potentially have root-specific or preferential expression in roots. Validation experiments showed that in general, the RNA-Seq expression profiles are correlated well with the RT-qPCR data. Some of the discrepancies observed were possibly due to the different biological materials used in both transcript profiling platforms. Out of the seven selected candidates, we found two novel oil palm transcripts, *EgMT* and *EgPRP1*, that were highly expressed in the root and barely detectable in other tissues. The results greatly suggest the potential of *EgMT* and *EgPRP1* promoters in regulating the strong expression of transgenes in a root-specific manner.

ACKNOWLEDGEMENT

We would like to thank the MPOB management for permission to publish this article. We also would like to acknowledge members of the Advanced Biotechnology and Breeding Centre, MPOB especially the Bioinformatics Unit and Clonal Propagation Group (CPG) for helpful discussion and technical support. Special thanks to Nadzirah Amiruddin for her kind help. We would also like to extend our thanks to the staff of Transgenic Technology Group (TTG) for their helpful assistance.

REFERENCES

Agrios, G N (2005). Chapter 6 - How plants defend themselves against pathogens. *Plant Pathology* (Agrios, G N ed.) 5th edition. Academic Press, San Diego. p. 207-248.

Ahmadi, F; Akmar Abdullah, S N; Kadkhodaei, S; Ijab, S M; Hamzah, L; Aziz, M A; Rahman, Z A and Rabiah Syed Alwee, S S (2018). Functional characterization of the gene promoter for an *Elaeis guineensis* phosphate starvation-inducible, high affinity phosphate transporter in both homologous and heterologous model systems. *Plant Physiol. Biochem., 127*: 320-335.

Ahn, J H; Choi, Y; Kwon, Y M; Kim, S-G; Choi, Y D and Lee, J S (1996). A novel extensin gene encoding a hydroxyproline-rich glycoprotein requires sucrose for its wound-inducible expression in transgenic plants. *Plant Cell, 8*: 1477-1490.

Alpern, D; Gardeux, V; Russeil, J; Mangeat, B; Meireles-Filho, A C A; Breysse, R; Hacker, D and Deplancke, B (2019). BRB-seq: Ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol., 20(1)*: 71. DOI: 10.1186/s13059-019-1671-x.

Altschul, S F; Madden, T L; Schäffer, A A; Zhang, J; Zhang, Z; Miller, W and Lipman, D J (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res., 25(17)*: 3389-3402. DOI: 10.1093/nar/25.17.3389.

Bernhardt, C and Tierney, M L (2000). Expression of AtPRP3, a proline-rich structural cell wall protein from *Arabidopsis,* is regulated by cell-type-specific developmental pathways involved in root hair formation. *Plant Physiol., 122*: 705-714.

Bhattacharya, R; Koramutla, M K; Negi, M; Pearce, G and Ryan, C A (2013). Hydroxyproline-rich glycopeptide signals in potato elicit signalling associated with defense against insects and pathogens. *Plant Sci., 207*: 88-97.

Breen, S; Williams, S J; Outram, M; Kobe, B and Solomon, P S (2017). Emerging insights into the functions of pathogenesis related protein 1. *Trends Plant Sci., 22(10)*: 871-879. DOI: 10.1016/j.tplants.2017.06.013.

Bustin, S A; Benes, V; Garson, J A; Hellemans, J; Huggett, 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(4)*: 611-622.

Chan, P L; Rose, R J; Abdul Murad, A M; Zainal, Z; Leslie Low, E T; Li Ooi, L C; Eng Ooi, S; Suzaini, Y 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. DOI: 10.1371/journal.pone.0099774.

Chen, L; Jiang, B; Wu, C; Sun, S; Hou, W and Han, T (2014). *GmPRP2* promoter drives root-preferential expression in transgenic *Arabidopsis* and soybean hairy roots. *BMC Plant Biol., 14*: 245.

Choi, D W; Song, J Y; Kwon, Y M and Kim, S G (1996). Characterization of cDNA encoding a proline-rich 14 kDa protein in developing cortical cells of the roots of bean (*Phaseolus vulgaris*) seedlings. *Plant Mol. Biol., 30*: 973-982.

Claros, M G and Canovas, F M (1999). RNA isolation from plant tissues: A practical experience for biological undergraduate. *Biochem. Educ., 27*: 110-113.

Cobbett, C and Goldbrough, P B (2002). Phytochelatins and metallothionein: Roles in heavy metal detoxification and homeostasis. *Annu. Rev. Plant Biol., 53*: 159-182.

Costa, C; Giménez-Capitán, A; Karachaliou, N and Rosell, R (2013). Comprehensive molecular screening: From the RT-PCR to the RNA-seq. *Transl. Lung Cancer Res., 2(2)*: 87-91. DOI: 10.3978/j.issn.2218-6751.2013.02.05.

Dong, C J; Wang, Y; Yu, S S and Liu, J Y (2010). Characterization of a novel rice metallothionein gene promoter: Its tissue specificity and heavy metal responsiveness. *J. Integr. Plant Biol., 52(10)*: 914-924. DOI: 910.1111/j.1744-7909.2010. 00966.x.

Ebener, W; Fowler, T J; Suzuki, H; Shaver, J and Tierney, M L (1993). Expression of DcPRP1 is linked to carrot storage root formation and is induced by wounding and auxin treatment. *Plant Physiol., 101*: 259-265.

Evans, I M; Swinhoe, R; Gatehouse, L N; Gatehouse, J A and Boulter, D (1988). Distribution of root mRNA species in other vegetative organs of pea (*Pisum sativum* L.). *Mol. Gen. Genet., 214(1)*: 153-157. DOI: 10.1007/BF00340194.

Foley, R C and Singh, K B (1994). Isolation of a *Vicia faba* metallothionein like gene: Expression in foliar trichomes. *Plant Mol. Biol., 26*: 435-444. DOI: 410.1007/BF00039552.

Fowler, T J; Bernhardt, C and Tierney, M L (1999). Characterization and expression of four proline-rich cell wall protein genes in *Arabidopsis* encoding two distinct subsets of multiple domain proteins. *Plant Physiol., 121*: 1081-1091.

Franssen, H J; Nap, J P; Gloudemans, T; Stiekema, W; Van Dam, H; Govers, F; Louwerse, J; Van Kammen, A and Bisseling, T (1987). Characterization of cDNA for nodulin-75 of soybean: A gene product involved in early stages of root nodule development. *Proc. Natl. Acad. Sci. USA, 84(13)*: 4495-4499. DOI: 10.1073/pnas.84.13.4495.

Fujino, K; Obara, M and Sato, K (2014). Diversification of the plant-specific hybrid glycine-rich protein (*HyGRP*) genes in cereals. *Front. Plant Sci., 5*: 489-489. DOI: 10.3389/fpls.2014.00489.

García-Hernández, M; Murphy A and Taiz, L (1998). Metallothioneins 1 and 2 have distinct but overlapping expression patterns in *Arabidopsis*. *Plant Physiol., 118(2)*: 387-397. DOI: 10.1104/pp.118.2.387.

Guo, W J; Bundithya, W and Goldsbrough, P B (2003). Characterization of the *Arabidopsis* metallothionein gene family: Tissue-specific expression and induction during senescence and in response to copper. *New Phytol., 159*: 369-281.

Hernandez-Garcia, C M and Finer, J J (2014). Identification and validation of promoters and *cis*-acting regulatory elements. *Plant Sci., 217-218*: 109-119.

Ho, C-L; Tan, Y-C; Yeoh, K-A; Ghazali, A-K; Yee, W-Y and Hoh, C-C (2016). De novo transcriptome analyses of host fungal interactions in oil palm (*Elaeis guineensis* Jacq.). *BMC Genomics*, 17: 66.

Hsieh, P-H; Oyang, Y-J and Chen, C-Y (2019). Effect of *de novo* transcriptome assembly on transcript quantification. *Sci. Rep., 9(1)*: 8304. DOI: 10.1038/s41598-019-44499-3.

Jeong, J S; Kim, Y S; Baek, K H; Jung, H; Ha, S H; Choi, Y D; Kim, M; Reuzeau, C and Kim, J K (2010). Root-specific expression of *OsNAC10* improves drought tolerance and grain yield in rice under field drought conditions. *Plant Physiol. Biochem., 153*: 185-197.

Josè-Estanyol, M; Ruiz-Avila, L and Puigdomenech, P (1992). A maize embryo-specific gene encodes a proline-rich and hydrophobic protein. *Plant Cell, 4*: 413-423.

Joshi, R; Pareek, A and Singla-Pareek, S L (2016). Chapter 9 - Plant Metallothioneins: Classification, distribution, function, and regulation. *Plant Metal Interaction* (Ahmad, P ed.). Elsevier. p. 239-261.

Keller, B and Lamb, C J (1989). Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. *Genes Dev., 3:* 1639-1646.

Kushairi, A; Ong-Abdullah, M; Nambiappan, B; Hishamuddin, E; Izzudin, Z B; Razmah, G; Subramaniam, V; Sundram, S and Parveez, G K A (2018). Oil palm economic performance in Malaysia and R&D progress in 2018. *J. Oil Palm Res., 31(2)*: 165-194.

Kushairi, A; Singh, R and Ong-Abdullah, M (2017). The oil palm industry in Malaysia: Thriving with transformative technologies. *J. Oil Palm Res., 29(4)*: 431-439.

Lalonde, E; Ha, K C H; Wang, Z; Bemmo, A; Kleinman, C L; Kwan, T; Pastinen, T and Majewski, J (2011). RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res., 21(4)*: 545-554.

Li, Y; Li, C; Cheng, L; Yu, S; Shen, C and Pan, Y (2019). Over-expression of *OsPT2* under a rice root specific promoter *Os03g01700. Plant Physiol. Biochem., 136*: 52-57.

Lister, R; O'Malley, R C; Tonti-Filippini, J; Gregory, B D; Berry, C C; Millar, A H and Ecker, J R (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell, 133*: 523-536.

Liu, J J and Ekramoddoullah, A K (2003). Root-specific expression of a western white pine *PR10* gene is mediated by different promoter regions in transgenic tobacco. *Plant Mol. Biol., 52:* 103-120. DOI: 10.1023/A:1023930326839.

Liu, Q; Ding, C; Lang, X; Guo, G; Chen, J and Su, X (2019). Small non-coding RNA discovery and profiling with sRNA tools based on high-throughput sequencing. *Brief. Bioinform., 22(1)*: 463-473. DOI: 10.1093/bib/bbz151.

Livak, K J and Schmittgen, T D (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2 (-Delta Delta C(T)) method. *Methods, 25(4)*: 402-408. DOI: 10.1006/meth.2001.1262.

Low, E T L; Jayanthi, N; Chan, K-L; Sanusi, N S N M; Halim, M A A; Rosli, R; Azizi, N; Amiruddin, N; Angel, L P L; Ong-Abdullah, M; Singh, R; Manaf, M A A; Sambanthamurthi, R; Parveez, G K A and Kushairi, A (2017). The oil palm genome revolution. *J. Oil Palm Res., 29(4)*: 456-468.

Lu, G and Moriyama, E N (2004). Vector NTI, a balanced all-in-one sequence analysis suite. *Brief. Bioinform., 5(4):* 378-388. DOI: 10.1093/bib/5.4.378.

Masani, M Y A; Izawati, A M D; Omar, O A and Parveez, G K A (2018). Biotechnology of oil palm: Current status of oil palm genetic transformation. *Biocatal. Agric. Biotechnol., 15*: 335-347.

Masura, S S; Tahir, N I; Rasid, O A; Ramli, U S; Othman, A; Masani, M Y A; Parveez, G K A and Kushairi, A (2017). Post-genomic technologies for the advancement of oil palm research. *J. Oil Palm Res., 29(4)*: 469-486.

Meister, R; Rajani, M S; Ruzicka, D and Schachtman, D P (2014). Challenges of modifying root traits in crops for agriculture. *Trends Plant Sci., 19*: 779-788.

Morris, E P; Chan, P K; Ooi, C L L; Ong, P W; Kamaruddin, K; Marjuni, M; Amiruddin, M A; Chan, K L; Rosli, R; Low, E T L; Singh, R; Shaharuddin, N A; Murphy, D J; Manaf M A A and Siti Nor Akmar, A (2020). Transcriptome analysis across various mesocarp developmental stages of MPOB-Angola *Dura*. *J. Oil Palm Res., 32(3)*: 406-419.

Naher, L; Yusuf, U K; Ismail, A; Tan, S G and Mondal, M M A (2013). Ecological status of *Ganoderma* and basal stem rot disease of oil palms (*Elaeis guineensis* Jacq.). *Aust. J. Crop Sci., 7(11)*: 1723-1727.

Nitz, I; Berkefeld, H; Puzio, P S and Grundler, F M V (2001). *Pyk10*, a seedling and root specific gene and promoter from *Arabidopsis thaliana. Plant Sci., 161*: 337-346.

Oikonomopoulos, S; Bayega, A; Fahiminiya, S; Djambazian, H; Berube, P and Ragoussis, J (2020). Methodologies for transcript profiling using long-read technologies. *Front. Genet., 11*: 606. DOI: 10.3389/fgene.2020.00606.

O'Leary, N A; Wright, M W; Brister, J R; Ciufo, S; Haddad, D; McVeigh, R; Rajput, B; Robbertse, B; Smith-White, B; Ako-Adjei, D; Astashyn, A; Badretdin, A; Bao, Y; Blinkova, O; Brover, V; Chetvernin, V; Choi, J; Cox, E; Ermolaeva, O; Farrell, C M; Goldfarb, T; Gupta, T; Haft, D; Hatcher, E; Hlavina, W; Joardar, V S; Kodali, V K; Li, W; Maglott, D; Masterson, P; McGarvey, K M; Murphy, M R; O'Neill, K; Pujar, S; Rangwala, S H; Rausch, D; Riddick, L D; Schoch, C; Shkeda, A; Storz, S S; Sun, H; Thibaud-Nissen, F; Tolstoy, I; Tully, R E; Vatsan, A R; Wallin, C; Webb, D; Wu, W; Landrum, M J; Kimchi, A; Tatusova, T; DiCuccio, M; Kitts, P; Murphy, T D and Pruitt, K D (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res., 44(D1)*: D733-745. DOI: 10.1093/nar/gkv1189.

Parveez, G K A; Rasid, O A; Masani, M Y A and Sambanthamurhi, R (2015). Biotechnology of oil palm: Strategies towards manipulation of lipid content and composition. *Plant Cell Rep., 34*: 533-543.

Potenza, C; Aleman, L and Sengupta-Gopalan, C (2004). Targeting transgene expression in research, agricultural and environmental applications: Promoters used in plant transformation. *In Vitro Cell. Dev. Biol. Plant, 40*: 1-22.

Pruitt, K D; Tatusova, T and Maglott, D R (2005). NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res., 33 (Database issue)*: D501-D504. DOI: 10.1093/nar/gki025.

Rasid, O A; Masura, S S; Hanin, A N; Izawati, A M D; Bohari, B; Masani, M Y A and Parveez, G K A (2020). Oil palm transgenic research: Challenges, update and future outlook. *The oil palm genome. Compendium of Plant Genome* (Ithnin, M and Kushairi, A eds.). Springer, Cham. p. 69-81. DOI: 10.1007/978-3-030-22549-0_6.

Reid, S J and Ross, G S (1997). Upregulation of two cDNA clones encoding metallothionein-like proteins in apple fruit during cool storage. *Physiol. Plant., 100*: 183-189. DOI: 110.1111/j.1399-3054.1997.tb03471.x.

Rival, A (2017). Breeding the oil palm *(Elaeis guineensis* Jacq.) for climate change. *OCL-Oilseeds Fats Crops Lipids, 24*: D107.

Rivas, M A; Beaudoin, M; Gardet, A; Stevens, C; Sharma, Y; Zhang, C K; Boucher, G; Ripke, S; Ellinghaus, D; Burtt, N; Fennell, T; Kirby, A; Latiano, A; Goyette, P; Green T; Halfvarson, J; Haritunians, T; Korn, J M; Kuruvilla, F; Lagacé, C; Neale, B; Lo, K S; Schumm, P and Törkvist, L (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet., 43*: 1066-1073.

Sanders, R; Mason, D J; Foy, C A and Huggett, J F (2014). Considerations for accurate gene expression measurement by reverse transcription quantitative PCR when analysing clinical samples. *Anal. Bioanal. Chem., 406(26)*: 6471-6483. DOI: 10.1007/s00216-014-7857-x.

Shabb, J B; Muhonen, W W and Mehus, A A (2017). Quantitation of human metallothionein isoforms in cells, tissues and cerebrospinal fluid by mass spectrometry. *Methods Enzymol., 586*: 413-431.

Showalter, A M and Rumeau, D (1990). Molecular biology of the plant cell wall hydroxyproline-rich glycoproteins. *Organization and Assembly of Plant and Animal Extracellular Matrix* (Adair, W S and Mecham, R P eds.). Academic Press, New York. p. 247-281.

Singh, R; Meilina, O A; Low, E T L; Arif, M A A; Rozana, R; Rajanaidu, N; Ooi, C L L; Ooi, S E; Chan, K L; Mohd Amin, H; Norazah, A; Jayanthi, N; 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 interfertile species in old and new worlds. *Nature, 500(7462)*: 335-339.

Srinath, T; Reddy, V D and Rao, K V (2017). Isolation and functional characterization of a novel stress inducible promoter from pigeon pea (*Cajanus cajan* L). *Plant Cell Tiss. Organ Cult., 128*: 457-468.

Suzuki, H; Wagner, T and Tierney, M L (1993). Differential expression of two soybean *(Glycine max* L.) proline-rich protein genes after wounding. *Plant Physiol., 101*: 1283-1287.

Tikhonovich, I A and Provorov, N A (2007). Chapter 14 - Beneficial plant-microbe interactions. *Comprehensive and Molecular Phytopathology* (Dyakov, Y T; Dzhavakhiya, V G and Korpela, T eds.). Elsevier, Amsterdam. p. 365-420.

Trapnell, C; Pachter, L and Salzberg, S L (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics, 25:* 1105-1111.

Trapnell, C; Williams, B A; Pertea, G; Mortazavi, A; Kwan, G; van Baren, M J; Salzberg, S T; Wold, B J and Pachter, L (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol., 28(5)*: 511-515. DOI: 10.1038/nbt.1621.

Wasson, A P; Richards, R A; Chatrath, R; Misra, S C; Prasad, S V and Rebetzke, G J (2012). Traits and selection strategies to improve root systems and water uptake in water-limited wheat crops. *J. Exp. Bot., 63*: 3485-3498. DOI: 3410.1093/jxb/ers3111.

Wilson, R C; Long, F; Maruoka, E M and Cooper, J B (1994). A new proline-rich early nodulin from *Medicago truncatula* is highly expressed in nodule meristematic cells. *Plant Cell, 6*: 1265-1275.

Winicov, I; Valliyodan, B; Xue, L and Hoober, J K (2004). The *MsPRP2* promoter enables strong heterologous gene expression in a root-specific manner and is enhanced by overexpression of Alfin 1. *Planta, 219*: 925-935.

Xu, Y; Buchholz, W G; DeRose, R T and Hall, T C (1995). Characterization of a rice gene family encoding root-specific proteins. *Plant Mol. Biol., 27*: 237-248.

Xue, G-P; Rae, A L; White, R G; Drenth, J; Richardson, T and McIntyre, C L (2016). A strong root-specific expression system for stable transgene expression in bread wheat. *Plant Cell Rep., 35(2)*: 469-481.

Yamamoto, Y T; Taylor, C G; Acedo, G N; Cheng, C L and Conkling, M A (1991). Characterization of *cis*-acting sequences regulating root-specific gene expression in tobacco. *Plant Cell, 3*: 371-382.

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 Tissue Organ Cult., 116(1)*: 55-66. DOI: 10.1007/s11240-013-0382-3.

Zeng, Y and Yang, T (2002). RNA isolation from highly viscous samples rich in polyphenols and polysaccharides. *Plant Mol. Biol. Rep., 20*: 417a-417e.

Zubaidah, R and Siti Nor Akmar, A (2005). The effects of metal ions on root-specific expression of the oil palm *MT3-B* gene promoter. *Proc. of the PIPOC 2005 International Palm Oil Congress-Agriculture Conference.* MPOB, Bangi. p. 1104-1110.

Zulkifli, Y; Norziha, A; Naqiuddin, M H; Fadila, A M; Nor Azwani, A B; Suzana, M; Samsul, K R; Ong-Abdullah, M; Singh, R; Parveez, G K A and Kushairi, A (2017). Designing the oil palm of the future. *J. Oil Palm Res., 29*: 440-455.