

# IDENTIFICATION OF GENES PREFERENTIALLY EXPRESSED IN MESOCARP TISSUE OF OIL PALM USING *in silico* ANALYSIS OF TRANSCRIPTS

SITI SURIAWATI BADAI\*; KUANG-LIM CHAN\*; PEK-LAN CHAN\*; GHULAM KADIR AHMAD PARVEEZ\*  
and OMAR ABD RASID\*

## ABSTRACT

Strong demand for palm oil creates a huge challenge for the industry to produce more palm oil on the shrinking cultivable land mass. Genetic manipulation of oil synthesis in the mesocarp tissue of oil palm offers the opportunity to improve yield and quality of palm oil on the limited land resources. However, more than one mesocarp-specific promoters are required to effectively manipulate oil synthesis to avoid epigenetic silencing caused by the sequence homology in their promoter regions. This study was aimed at identifying genes that are specifically expressed in the mesocarp tissue through *in silico* analysis of transcriptome datasets from different oil palm tissues. From this analysis, transcripts XLOC\_10, annotated as pentatricopeptide repeat (PPR) protein was highly expressed in the mesocarp tissue. Gene expression analysis of PPR in 27 tissues of *Elaeis guineensis* indicated that this putative mesocarp-specific transcript was expressed in mesocarp and female inflorescence tissues only. Cis-acting elements that are present in the promoter region of PPR showed that PPR might be involved in the lipid biosynthesis and flowering regulation network in oil palm.

**Keywords:** oil palm, genetic engineering, promoter, transcriptome, RT-qPCR.

**Date received:** 8 August 2018; **Sent for revision:** 9 August 2018; **Received in final form:** 9 December 2018; **Accepted:** 7 May 2019.

## INTRODUCTION

The African oil palm (*Elaeis guineensis*) has the highest yield compared to other cultivated oleaginous crops (USDA, 2017). Palm oil can be processed to form products with different melting point and characteristic for commercial application ranging from food products, detergent, cosmetics and biofuel (Koushki *et al.*, 2015). Global demand for palm oil is expected to rise to over 250 million tonnes by 2050 (Zulkifli *et al.*, 2017). Productivity, sustainability and fulfillment of the high demand of global oils and fats market can be achieved by implementation of

disruptive technologies to oil palm sector (Kushairi *et al.*, 2017). The promising technologies such as plant genetic manipulation for development of new planting materials with desirable agronomic traits is one of the core researches at the Malaysian Palm Oil Board (MPOB). Improved planting materials are important to overcome the challenges in the oil palm industry such as limited land resources, stagnant yields, labour shortage and climate change (Kushairi *et al.*, 2017).

The main focus in oil palm genetic manipulation is to develop transgenic oil palm with high oleic content (Masani *et al.*, 2014; Parveez *et al.*, 2015; Masura *et al.*, 2017). Other targets of oil palm biotechnology research are to increase stearic acid, palmitoleic acid, ricinoleic acid, lycopene (carotenoid) and to produce biodegradable plastics (Parveez *et al.*, 2015). High

\* Malaysian Palm Oil Board, 6 Persiaran Institusi,  
Bandar Baru Bangi, 43000 Kajang,  
Selangor, Malaysia.  
E-mail: suria@mpob.gov.my

oleic acid diet is beneficial for human health which can increase high-density lipoprotein (HDL) and lower low-density lipoprotein (LDL) cholesterol levels (Karacor and Cam, 2015). Food market demands for high oleic oil with high oxidative stability (Montoya *et al.*, 2014). Amongst vegetable oils, palm oil has the highest resistance to oxidative degradation during deep-frying process (Tarmizi *et al.*, 2016). Increased oleic acid content can improve the frying properties of palm oil and results in high-quality and tasty foods (Matthäus, 2007). Moreover, high oleic palm oil is also predominantly used as the feedstock in the oleochemical industry for the production of laundry detergent and personal care items (Salimon *et al.*, 2012).

Biochemical and gene expression studies indicated that oleic acid content in palm oil can be increased through up-regulation of  $\beta$ -ketoacyl-ACP synthase II (KASII) and down-regulation of palmitoyl-ACP thioesterase and oleoyl-CoA desaturase (Sambanthamurthi, 2000; Sambanthamurthi *et al.*, 2002; Parveez *et al.*, 2015). However, construction of transformation vectors using homologous promoters for multigene transformation can induce epigenetic silencing (Matzke *et al.*, 1989; Rajeevkumar *et al.*, 2015). Two types of epigenetic silencing have been proposed, called transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS). TGS acts through repression of transcription by DNA methylation and chromatin remodelling (Matzke *et al.*, 2000). In TGS, genes that share sequence homology in their promoter regions are highly methylated (Matzke *et al.*, 2000; Dalakouras and Wassenegger, 2013). PTGS, also known as RNA silencing, does not affect the transcription of a gene locus, but occurs after the formation of mRNA. PTGS causes degradation of target mRNA through sequence-specific nucleotide interactions induced by double-stranded RNA (Kooter *et al.*, 1999; Matzke *et al.*, 2001; Vaucheret *et al.*, 2001). Epigenetic silencing mechanisms is a huge hindrance to transgenic technologies used in crop improvement (Rajeevkumar *et al.*, 2015).

Since modification of palm oil involves more than one gene, each of the introduced genes must have non-homologous promoters driving them to avoid the homology-dependent gene silencing. Therefore, more mesocarp-specific promoters are essential to effectively modify lipid metabolism and to ensure desired spatial expression of transgenes. For now, two mesocarp-specific promoters derived from type 3 metallothionein-like (MT3-A) (Siti Nor Akmar and Zubaidah, 2008) and lipase class 3 (FLL1) genes (Nurniwalis *et al.*, 2015) have been isolated and characterised (Omidvar *et al.*, 2010; Kamaladini *et al.*, 2011; Zubaidah *et al.*, 2018). The tissue specificity of MT3-A and FLL1 promoters was examined using transient  $\beta$ -glucuronidase

(GUS) expression analysis in the oil palm mesocarp tissue and leaves as control tissue. The expression of GUS driven by both promoters was only detected in a mesocarp tissue revealed that the promoters are specific to mesocarp tissue (Zubaidah *et al.*, 2018). Currently, MT3-A promoter is being used for the modification of the fatty acid content and production of high value-added products in the mesocarp (Yunus and Kadir, 2008; Yunus *et al.*, 2008; Parveez *et al.*, 2015).

Utilisation of transcriptome information (Bourgis *et al.*, 2011; Dussert *et al.*, 2013; Shearman *et al.*, 2013; Tee *et al.*, 2013) and genome sequences of the African oil palm (Singh *et al.*, 2013) together with oil palm gene models annotation (Chan *et al.*, 2017) can facilitate the identification of new mesocarp-specific promoter in oil palm. In this study, we identified oil palm mesocarp-specific genes through *in silico* analysis of multiple transcriptomic datasets. The expression of one of the genes (XLOC\_10) was further validated across 27 different tissues of *E. guineensis* using reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) approach. Promoter region of XLOC\_10 was analysed using PlantPAN 2.0 to identify the presence of relevant *cis*-acting elements that are associated with lipid biosynthesis in the mesocarp tissue of oil palm.

## MATERIALS AND METHODS

### Plant Materials

Various oil palm tissues were obtained from commercial *tenera* planting materials (*dura*  $\times$  *pisifera*). These include mesocarp tissues from five weeks after anthesis (WAA) to 20 WAA, 10 WAA kernel, male and female flowers, spear and mature leaves, primary roots, lateral roots, whitish roots and oil palm tissue culture materials (embryogenic callus, polyembryoid and plantlet) were used for RT-qPCR.

### Transcriptome Data Analysis

A total of 27 transcriptomes of different oil palm tissues from MPOB were used in complementary with mesocarp transcriptome data published by Tranbarger *et al.* (2011). A total of 29 034 contigs from Tranbarger *et al.* (2011) which reflect the biochemical and physiological processes during the maturation and ripening phases of the mesocarp were analysed using Basic Local Alignment Search Tool (BLAST) against MPOB's transcriptome data from 27 tissue types. The E-value cutoff of  $1e^{-5}$  was used in the BLAST analyses and the best 20 hits of each BLAST were retained for the annotation. Putative mesocarp-specific transcripts were identified by filtering the coverage and similarity cutoff.

### In silico Gene Expression Analysis

Ribonucleic acid (RNA) sequencing raw reads from 27 transcriptome datasets (PRJNA201497, PRJNA345530) were mapped to genome P5 build of *E. guineensis* (Singh *et al.*, 2013) with spliced read mapper, TopHat. Cufflinks and Cuffdiff were used to assembly and estimate the abundance of reads mapped to genes body and calculated fragments per kilobase of transcript per million (FPKM) as proxy for gene expression in different tissues. Expression profile of putative mesocarp-specific transcripts was analysed by cummeRbund in R package and a heatmap was generated for visualisation.

### Sequence Analysis

Sequence analysis of putative mesocarp-specific transcripts was performed in order to understand its features, function or structure. Transcript sequences were functionally annotated using BLASTX against NCBI non-redundant protein database (<https://blast.ncbi.nlm.nih.gov/>). Protein sequences of putative mesocarp-specific genes were annotated using BLASTP against the Arabidopsis Information Resource (TAIR) (<https://www.arabidopsis.org/>) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (<https://www.genome.jp/kegg/>). Promoter sequence of putative mesocarp-specific genes was predicted from MPOB genome database. Regulatory motifs in the promoter sequences were identified using Plant Promoter Analysis Navigator 2.0 (PlantPAN <http://PlantPAN2.ips.ncku.edu.tw>) (Chow *et al.*, 2016).

### Total RNA Extraction, Purification and Quality Assessment

A combination of CTAB-based RNA extraction method and a silica column of RNeasy Plant Mini Kit (Qiagen, Germany) was used to isolate total RNA from 27 different tissues of oil palm which consist of mesocarp (16 comparable developing stages from 5 WAA until 20 WAA), kernel (10 WAA), flower (male flower and female flower), leaf (spear leaf and mature leaf), root (primary roots, lateral roots, and whitish roots) and oil palm tissue culture materials (embryogenic callus, polyembryoid and plantlet). Genomic DNA contamination was

removed using RNase-free DNase I according to the manufacturer's instructions (Qiagen USA, Valencia, CA). The purity and quantity of the purified total RNA were determined using a NanoDrop ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc.). RNA integrity was assessed by electrophoretic fractionation on an Agilent 2100 Bioanalyser and a RNA 6000 Nano LabChip (Agilent Technologies, CA).

### Reverse Transcription Quantitative Real-time Polymerase Chain Reaction (RT-qPCR)

The expression of putative mesocarp-specific genes across 27 different tissues of *E. guineensis* was studied using RT-qPCR approach. Gene-specific primers listed in Table 1 were designed using Primer3 software (Untergasser *et al.*, 2012). The RT-qPCR expression test was performed in 96 well-plate using CFX Connect™ Real-Time PCR Detection System (Bio-Rad). Amplification was performed in 20 µl qPCR reaction mixtures containing 16 ng of cDNA template, 500 nM of each primer and 1x iTaq Universal SYBR Green Supermix (Bio-Rad). All reactions were prepared in triplicate (n = 3) for each cDNA sample.

PCR amplification cycle used was as follow: 95°C, 30 s for 1 cycle; 95°C, 3 s and 51°C (depending on the annealing temperature of primer pairs), 30 s for 40 cycles and followed by a melting curve analysis at 65°C to 95°C with 0.5°C increase in temperature at each step. No reverse transcriptase (NRT) and non-template (NTC) reactions were included as negative controls.

PCR amplification efficiencies and R<sup>2</sup> values of reference gene primers were determined across six different pools of cDNA at five different concentration (1, 2, 4, 8, and 16 ng). Six pools of cDNA consist of mesocarp (16 comparable developing stages from 5 WAA until 20 WAA), kernel (10 WAA), flower (male flower and female flower), leaf (spear leaf and mature leaf), root (primary roots, lateral roots, and whitish roots) and oil palm tissue culture materials (embryogenic callus, polyembryoid and plantlet). The quantification cycle (Cq) was automatically captured and the data were analysed using Bio-Rad CFX Manager Software (ΔΔCq analysis mode) based on the method described by Livak and Schmittgen (2001). The expression data were first normalised to

TABLE 1. CANDIDATE MESOCARP-SPECIFIC GENE IN OIL PALM

Gene name	PPR
Forward primer sequence (5' -3')	GGGAAGTGAATGATGTGGC
Reverse primer sequence (5' -3')	AGAAGCACCTCCTTGAGCAT
Amplicon length (bp)	173
Annealing temperature (°C)	51

Note: PPR - pentatrico peptide repeat.

three reference genes, SLU7, CYP2 and GRAS (Yeap *et al.*, 2014). The relative expression of the putative mesocarp-specific transcripts were calculated by using 16WAA mesocarp as the calibrator.

## RESULTS AND DISCUSSION

### Identification of Mesocarp-specific Genes from Oil Palm

The availability of oil palm genome database (Singh *et al.*, 2013) with abundant transcriptome information enables the discovery of genes for important traits in various tissues of interest. MPOB's transcriptome data from 27 tissue types together with mesocarp transcriptome data published by Tranbarger *et al.* (2011) genes in oil palm. A total of 27 932 contigs from mesocarp transcriptome data were shown to have BLAST information from our 27 transcriptome dataset. Systematic identification for putative candidates of mesocarp-specific contigs from the BLAST's results was performed manually by using Microsoft Excel. Tissue ID was assigned to each BLAST result followed by grouping and identification of mesocarp-specific contigs. From 27 932 contigs, we identified 41 contigs that hit specifically to the mesocarp tissue. Afterward, *in silico* gene expression analysis was performed to measure the expression level of these 41 contigs by using BLASTN against 454-based RNA seq mapped transcripts with FPKM-normalised read counts. Based on oil palm gene expression database, only 13 contigs have information on gene expression level. *In silico* gene expression profiling shows that only XLOC\_10 is highly and specifically expressed in the mesocarp tissue (Figure 1). Another putative candidates of mesocarp-specific contigs, which is XLOC\_9 also expressed in the mesocarp tissue but has expression in the kernel tissue. Although the remaining candidates are expressed in the mesocarp tissue but they are also highly expressed in other tissues.

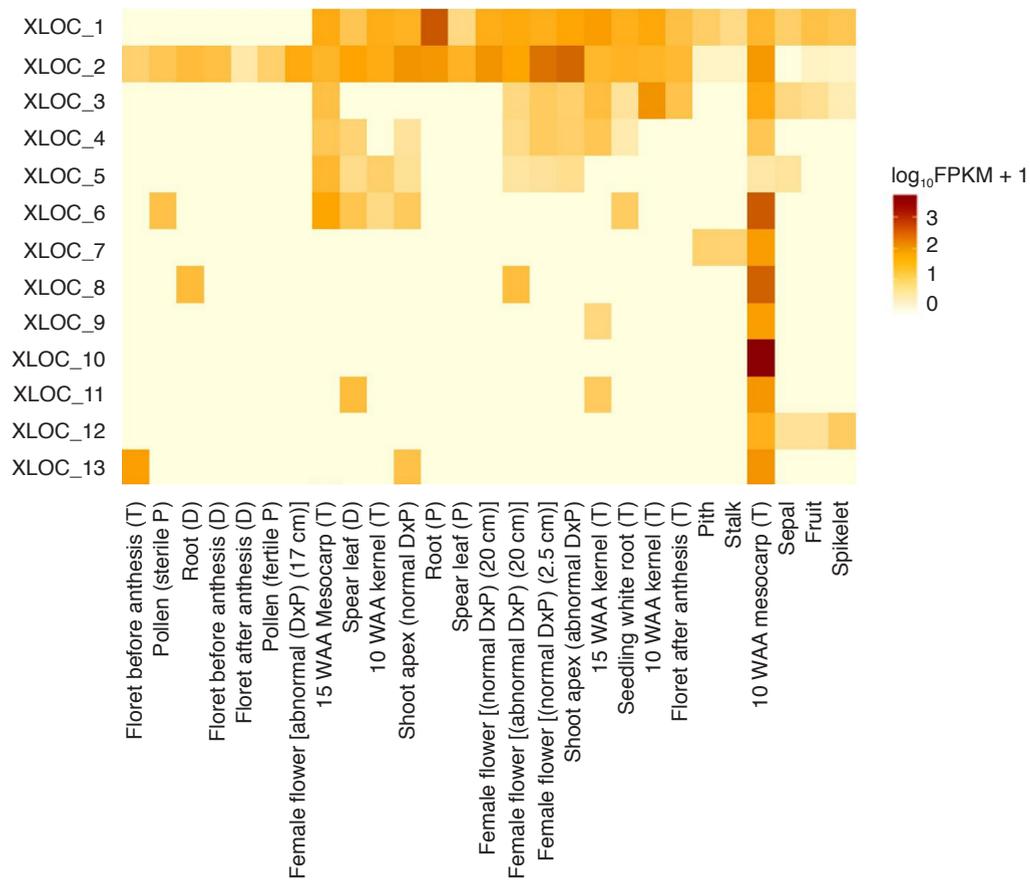
Nucleotide sequences of XLOC\_10 was annotated through different databases including the NCBI non-redundant database, the TAIR, and KEGG. The length of the open reading frame (ORF) of XLOC\_10 is 2016 bp. Based on sequence analysis, XLOC\_10 was annotated as pentatricopeptide repeat (PPR) protein. PPR family is defined by the presence of tandem arrays of a degenerate 35-amino acid motifs (Lurin *et al.*, 2004). PPR protein plays important role in RNA processing and translation in mitochondria and chloroplast (Lurin *et al.*, 2004). For example, *Arabidopsis organelle* transcript processing 43 (*otp43*) PPR mutants are defective in seed development, severely reduced germination rates and poor growth of the plant (de Longevialle *et al.*, 2007).

### Tissue Specificity Analysis of XLOC\_10

Temporal and spatial expression of XLOC\_10 in 27 different tissues of *E. guineensis* was investigated by using RT-qPCR. PCR amplification efficiencies (Ex) and correlation coefficient ( $R^2$ ) values of XLOC\_10 primer were determined across 16 comparable developing stages of mesocarp tissue from 5 WAA until 20 WAA of pooled cDNA. From the standard curve (Figure 2), PCR amplification efficiencies of XLOC\_10 is 99.7% and  $R^2$  value is greater than 0.98, indicating a positive correlation between the amount of cDNA template and the cycle threshold (Ct) values. Gene expression analysis was performed to verify the tissue specificity of XLOC\_10. Based on the expression results, the transcript level of XLOC\_10 was up-regulated to around 2.3-fold in 18 WAA and 8 WAA mesocarp tissues, and 1.6-fold in female flower, 14 WAA and 9 WAA mesocarp as compared to the calibrator (Figure 3). The result of the RT-qPCR appears to slightly contradict to the result of the *in silico* analysis of transcript which indicated that XLOC\_10 was highly and specifically expressed in the 10 WAA mesocarp tissue (Figure 1).

Discrepancy in resulting data from RNA-seq and RT-qPCR is probably due to the different sensitivity and specificity of each platform. RNA-Seq is used to analyse gene spliced transcripts, allele-specific expression and detection of novel transcripts (Maher *et al.*, 2009). The RT-qPCR on the other hand is the technology for sensitive RNA quantification of a targeted set of genes (Zeka *et al.*, 2016). In addition, RNA-seq and RT-qPCR use different methods for data normalisation. Transcript quantification with RNA-seq was normalised using the FPKM value while the relative expression of XLOC\_10 in RT-qPCR was calculated using the comparative method  $2^{-\Delta\Delta Ct}$  and normalised to the corresponding reference gene levels. Different methods of data normalisation can influence the interpretation of the data in RNA-seq and RT-qPCR. In addition, biological materials used for RT-qPCR were different from the biological materials that were used for RNA-seq analysis. Variability of biological replicates may contribute to variation in quantification of gene expression (Sanders *et al.*, 2014). Furthermore, gene expression is a stochastic process, with randomness in transcription and translation leading to significant cell-to-cell variations in mRNA and protein levels (Raj and van Oudenaarden, 2008).

Despite the differences in the results of RNA-seq and RT-qPCR, these methods have been widely used in many gene expression studies. Huang *et al.* (2016) identified differential expressed genes (DEG) that regulate rhizome formation and development in *Atractylodes lancea* through comparative transcriptome analysis. Gene expression analysis of 20 selected genes was verified using RT-qPCR



Note: D - *Dura*, P - *Pisifera*, T - *Tenera*.

Figure 1. Expression patterns of 13 putative mesocarp-specific genes in 27 tissues of *E. guineensis*. The expression levels were normalised based on the fragment per kilobase of transcript per million (FPKM) method (on the right side). The in silico analysis of transcripts indicated that XLOC\_10 was highly expressed in the mesocarp tissue.

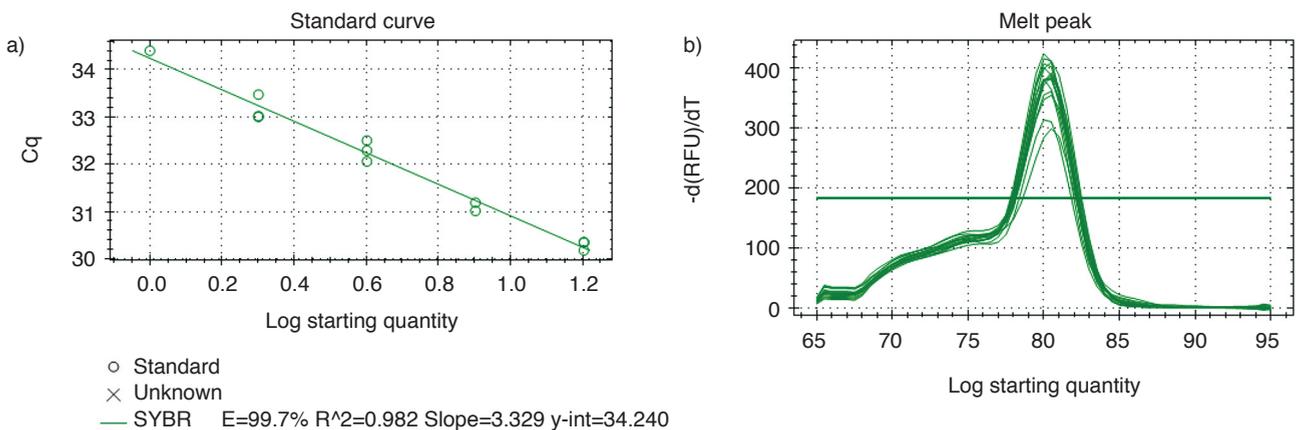


Figure 2. Determination of polymerase chain reaction (PCR) amplification efficiencies ( $E_x$ ) and correlation coefficient ( $R^2$ ) values for XLOC\_10 primer using the slope of standard curve (a). The presence of a single amplicon peak indicated the amplification of a specific PCR product (b).

and all genes show similar trend of the expression patterns in the transcriptome analysis and the RT-qPCR (Huang *et al.*, 2016). Similar method was used by Le Yu and his group, where the transcriptome data of *Populus euphratica* Oliv. seedlings grown with 0 mM, 150 mM or 300 mM NaCl were analysed

to identify tissue-specific DEG under salt stress. Six DEG were selected for RT-qPCR analysis to validate the reliability of RNA-seq data. Fold-changes of these six DEG which detected by sequencing did not exactly match with the fold-changes detected by RT-qPCR, but expression patterns were mostly

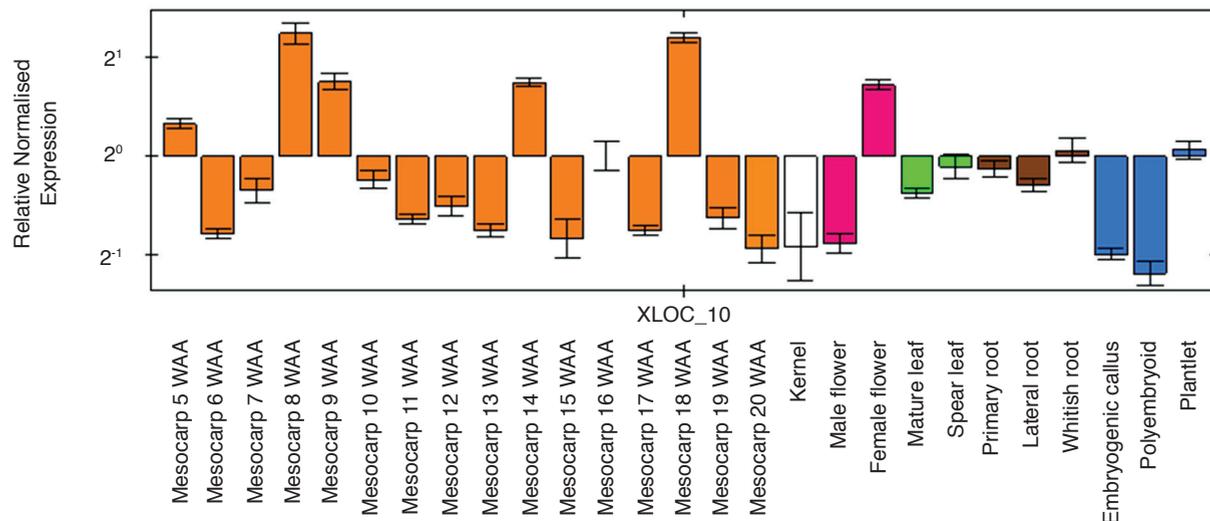


Figure 3. Expression of XLOC\_10 across 27 different oil palm tissues. The expression level is presented in log<sub>2</sub> (relative expression). Standard deviations are shown by vertical bars. All samples were prepared in triplicates.

consistent for all the selected genes (Yu *et al.*, 2017). Analysis of transcriptome dataset consisting of six major tissues dissected from mid-gestational (TS12) mouse embryos reveals 1375 DEG which are involved in the process of controlling mid-gestational development. In this particular study, 10 DEG were selected for gene expression analysis using RT-qPCR and all genes show consistent expression pattern as detected by RNA-seq (Werber *et al.*, 2014).

#### Identification of Regulatory Elements on the Promoter Region of XLOC\_10 Gene

Based on the gene expression analysis, XLOC\_10 was expressed at different stages of mesocarp development and female flower tissues. Temporal and spatial regulation of XLOC\_10 gene depends on the specific *cis*-regulatory elements that are present in its promoter region. Promoter sequence of XLOC\_10 was obtained from the MPOB oil palm genome database and analysed using Plant Promoter Analysis Navigator 2.0 (PlantPAN). The size of XLOC\_10 promoter region is about 1.6 kb and the nucleotide sequence is shown in Figure 4. Analysis of the promoter sequence indicated the presence of several *cis*-regulatory elements including B3 DNA-binding domain, AAAG, homeodomain-leucine zipper (HD-ZIP), and CCTCGTAC motifs. The B3 domain transcription factor is shared in many seed-specific promoters like leafy cotyledon2 (LEC2) transcription factor. LEC2 plays an essential role in seed maturation and controls genes that are involved in oil accumulation in seeds (Stone *et al.*, 2008). The AAAG sequence is the core motif of DNA binding of the DOF (DNA-binding with one finger) domain. This motif is crucial for controlling

the accumulation of lipids (Wang *et al.*, 2007). The HD-ZIP proteins which are characterised by a HD followed by a ZIP motif is one of the largest families of transcription factors unique to plants. Mutation of *Arabidopsis* GLABRA2 (GL2), a member of the HD-ZIP family of transcription factors, has led to the increase of seed oil, implicating its role as a negative regulator (Shi *et al.*, 2012). The *cis*-acting element CCTCGTAC, which is present in the APETALA2 (AP2) ethylene responsive transcription factors such as target of eat1 (TOE1) and TOE2 (Franco-Zorrilla *et al.*, 2014) serves as major integrators in biotic and abiotic stress signaling. In addition to the above motifs, our analysis also indicated the presence of a *cis*-regulatory element, GTAC motif, which is related to the expression of gene XLOC\_10 in the female flower tissue. This motif is important for flowering time regulation. SQUAMOSA promoter binding-like (SPL) proteins are predicted to directly bind to GTAC motif and control the early stage of flower development (Klein *et al.*, 1996; Preston and Hileman, 2013). The presence of these motifs suggests that XLOC\_10 might facilitate an important role in lipid metabolism and hormonal regulation of flower development in oil palm.

#### CONCLUSION

In conclusion, we identified XLOC\_10 as a mesocarp-specific transcript through transcriptome analysis. Gene expression analysis indicated that XLOC\_10 was expressed in the mesocarp and flower tissues. *Cis*-acting elements that are present in the promoter region of XLOC\_10 showed that XLOC\_10 might be involved in lipid biosynthesis and flowering regulation.

ATCTGCTTCAAGATTCGGCAATTAATTTGCACGTCTATATTCTGCTATTAGCTCTTTGC -1501  
 AAGAGGGAGGACCTAGTAAGGCTTCCACCTCAAAGTACGTATGAGATAAAGAAGTCTTA -1441  
 GTGCACTTCTTGAGAGCTGAAAGCCTGAAACTATGTTTCTGAGCAAGATATTAAGATAT -1381  
 GAGAAACATGGATGGAAGCATGTGTGAATGCTTATTCATGACTTTTAAGTTCTGGATCCT -1321  
 CAATTGCAATAGCTTTAGGTGCCATGATTGCTACCAAGGGAAGTCTGCATCTAGGTTACT -1261  
 TGCAAGATTTTCAATTTAGCTTATTTTGAATAAATTTCAACTTTCTTGAAT **SBP**  
**CAAGTACT** -1201  
**TT**GCATTTTCAATTGGCAAAAATAGCCCGTTCTCATCGGCAGAAATAAGCTCTGGAATTTA -1141  
 AGTGCAGGGACTGAATCTTATTTTCTAGTGCCTCTTAAAATCTGCTTCAAGCTTTCTGAC -1081  
 ATTAACCTCCATCTGTCCCATCAAATTTGTGTTTGTAAATATGTTACAGATTAAGTCTCTT -1021  
 CTTTGTGGGCCAATATAGATCTGGCCATTGGTGTGGAGATCATTGAGCTAGATTTCTTAA -961  
 GCTTTGGAACCAGCTTGTGGAAACCTAGCAAGTATGCGTATTTGTTAGTAAATGAGCAG -901  
 TCAAGTAGTAAGGGAGAAGCCTGACATCATTGACCAAAGTATGATGTGATGGGGAGAA -841  
 TTCTGACTTGCTTAGTTGATGAAATCAGCTGACCAAAAAGAAGAGGAAGGATAATTTGT -781  
 GAGAACTAGAGAGTTTATATGGCAC **HD-ZIP**  
**AAATTAATGA**GCATATATTTGCTTACCGTGGAGCC -721  
 AGCATGGTGAAGCTAAGCCAAATCACCCTAAAAGGACTTCAACTAATCATCTGAAAAA -661  
 ACTTCACTTGGACCATGG **AP2** **TGTACGATGG**TTTGAAAATTCAAACCAAAGTACCAC **Dof**  
**TAACTT** -601  
**TAACT**AGCCATCTAAATGAACTTACTTGAACATGGTGAATCATGCCCGTGGTGGAAACAA -541  
 GTGGGTGGTGCAGGATAGCCAAGTAACTTTTAGAATATCTGGATTGCCCCCTTTTGAGTT -481  
 ATCGTGGCCTTAGAAGAAGGAAAAAATATTTTATATGCCCTTAACATATCTTTTAGCA -421  
 TATTGGGACCCTGAAGTGGTTATTTACTAATTAATAGAAATAAATAGATACTAAAGGTTT -361  
 GCCGCAATTGGGTATGCCGGATAACAGGATAATTTAATAATTTTAAAAATTATTTATTT -301  
 TAAAAATAATTATTAATAAAAAAATAAATTCAAACATATTTGATTGATTAAAAATTATT -241  
 TAAAAAATGATATCGATAAAAAAATTGCTATGGTTGATTGAAGATAATTTTATCTAAAA -181  
 ATATTATCAAACCTACGACAAATATGTGTGGACAAATAATTTAAATAATAATATTTTTTTT -121  
 TAAAATCTATTTATTGGTAATCCATG **B3** **GATA BOX**  
**ATCTGCATGA**TAAATGTTGAAATATATTTACATAA -61  
 ATTAATAAATATTTTAAATTAAT **TATA BOX** **TAAATATATATTAGAAAGTATATTTCTCATCCTGGT** **TSS**  
 -1  
 TTCGCCGAGCGGTACTTGAAAACCACGAAAAATTAGATTACCGAGGTAAAAAATTTGAA +60  
 TCTCTCACTCCTCGCGTTCCGTTGGTACGTCGTAGTAGCAGAGACATGAGGCGAGACGGA +120  
**ORF**  
**GGACATG** → +127

Figure 4. The sequence of XLOC\_10 gene promoter from -1501 to +127 non-template relative to the start site of translation is shown, along with identified binding sites for the indicated transcription factors.

## ACKNOWLEDGEMENT

The authors would like to thank the Director-General of MPOB for permission to publish this article. We also would like to acknowledge members of the Transgenic Technology Group, Bioinformatics Unit, Breeding and Tissue Culture Unit and Genomics Unit of Advanced Breeding and Biotechnology Centre, MPOB. Special thanks to Dr Abdul Masani Mat Yunus, Siti Masura Subhi and Lim Fook Hwa for critically reviewing this manuscript.

## REFERENCES

- 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.
- Chan, K L; Rosli, R; Tatarinova, T V; Hogan, M; Firdaus-Raih, M and Low, E T L (2017). Seqping: Gene prediction pipeline for plant genomes using self-training gene models and transcriptomic data. *BMC Bioinformatics*, 18. DOI: 10.1186/s12859-016-1426-6.
- Chow, C N; Zheng, H Q; Wu, N Y; Chien, C H; Huang, D A; Lee, T Y; Chiang-Hsieh, Y F; Hou, P F; Yang, T Y and Chang, W C (2016). PlantPAN 2.0: An update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.*, 44: D1154-D1164.
- Dalakouras, A and Wassenegger, M (2013). Revisiting RNA-directed DNA methylation. *RNA Biol.*, 10: 453-455.
- De Longevialle, A F; Meyer, E H; Andrés, C; Taylor, N L; Lurin, C; Millar, A H and Small, I D (2007). The pentatricopeptide repeat gene OTP43 is required for *trans*-splicing of the mitochondrial nad1 Intron 1 in *Arabidopsis thaliana*. *Plant Cell*, 19: 3256-3265.
- 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.
- Franco-Zorrilla, J M; López-Vidriero, I; Carrasco, J L; Godoy, M; Vera, P and Solano, R (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci.*, 111: 2367-2372.
- Huang, Q; Huang, X; Deng, J; Liu, H; Liu, Y; Yu, K and Huang, B (2016). Differential gene expression between leaf and rhizome in *Atractylodes lancea*: A comparative transcriptome analysis. *Front Plant Sci.*, 7: 348.
- Kamaladini, H; Abdullah, S N A and Aziz, M A (2011). Metal inducible activity of the oil palm metallothionein-like gene promoter (MT3-A) in prokaryotes. *J. Biosci. Bioeng.*, 111: 217-225.
- Karacor, K and Cam, M (2015). Effects of oleic acid. *Med. Sci. Discov.*, 2: 125.
- Klein, J; Saedler, H and Huijser, P (1996). A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene SQUAMOSA. *Mol. Gen. Genet.*, 250: 7-16.
- Kooter, J M; Matzke, M A and Meyer, P (1999). Listening to the silent genes: Transgene silencing, gene regulation and pathogen control. *Trends Plant Sci.*, 4: 340-347.
- Koushki, M; Nahidi, M and Cheraghali, F (2015). Physico-chemical properties, fatty acid profile and nutrition in palm oil. *J. Paramed. Sci. Summer*, 6: 117-134.
- Kushairi, A; Singh, R and Ong-Abdullah, M (2017). The oil palm industry in Malaysia: Thriving with transformative technologies. *J. Oil Palm Res. Vol.* 29(4): 431-439.
- Livak, K J and Schmittgen, T D (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*, 25: 402-408.
- Lurin, C; Andrés, C; Aubourg, S; Bellaoui, M; Bitton, F; Bruyère, C; Caboche, M; Debast, C; Gualberto, J; Hoffmann, B; Lecharny, A; Le Ret, M; Martin-Magniette, M L; Mireau, H; Peeters, N; Renou, J P; Szurek, B; Tacconat, L and Small, I (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, 16: 2089-2103.
- Maher, C A; Palanisamy, N; Brenner, J C; Cao, X; Kalyana-Sundaram, S; Luo, S; Khrebtukova, I; Barrette, T R; Grasso, C; Yu, J; Lonigro, R J; Schroth, G; Kumar-Sinha, C and Chinnaiyan, A M (2009). Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc. Natl. Acad. Sci.*, 106: 12353-12358.
- Masani, M Y A; Noll, G A; Parveez, G K A; Sambanthamurthi, R and Prüfer, D (2014). Efficient transformation of oil palm protoplasts by PEG-

mediated transfection and DNA microinjection. *PLoS ONE*, 9(5): e96831.

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. Vol. 29(4)*: 469-486.

Matthäus, B (2007). Use of palm oil for frying in comparison with other high-stability oils. *Eur. J. Lipid Sci. Technol.*, 109: 400-409.

Matzke, M A; Mette, M F and Matzke, A J (2000). Transgene silencing by the host genome defense: Implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Mol. Biol.*, 43: 401-415.

Matzke, M A; Matzke, A J M; Pruss, G J and Vance, V B (2001). RNA-based silencing strategies in plants. *Curr. Opin. Genet. Dev.*, 11: 221-227.

Matzke, M A; Primig, M; Trnovsky, J and Matzke, A J (1989). Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.*, 8: 643-649.

Montoya, C; Cochard, B; Flori, A; Cros, D; Lopes, R; Cuellar, T; Espeout, S; Syaputra, I; Villeneuve, P; Pina, M; Ritter, E; Leroy, T and Billotte, N (2014). Genetic architecture of palm oil fatty acid composition in cultivated oil palm (*Elaeis guineensis* Jacq.) compared to its wild relative *E. oleifera* (H.B.K) Cortés. *PLoS ONE* 9(5): e95412.

Nurniwalis, A W; Zubaidah, R; Siti Nor Akmar, A; Zulkifli, H; Mohamad Arif, M A; Massawe, F J; Chan, K L and Parveez, G K A (2015). Genomic structure and characterization of a lipase class 3 gene and promoter from oil palm. *Biol. Plant.*, 59: 227-236.

Omidvar, V; Abdullah, S N A; Izadfard, A; Ho, C L and Mahmood, M (2010). The oil palm metallothionein promoter contains a novel AGTTAGG motif conferring its fruit-specific expression and is inducible by abiotic factors. *Planta*, 232: 925-936.

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

Preston, J C and Hileman, L C (2013). Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Front. Plant Sci.*, 4: 80. DOI: 10.3389/fpls.2013.00080.

Raj, A and Van Oudenaarden, A (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, 135: 216-226.

Rajeevkumar, S; Anunanthini, P and Sathishkumar, R (2015). Epigenetic silencing in transgenic plants. *Front. Plant Sci.*, 6: 693.

Salimon, J; Salih, N and Yousif, E (2012). Industrial development and applications of plant oils and their biobased oleochemicals. *Arab. J. Chem.*, 5: 135-145.

Sambanthamurthi, R (2000). Chemistry and biochemistry of palm oil. *Prog. Lipid Res.*, 39: 507-558.

Sambanthamurthi, R; Siti Nor Akmar, A and Parveez, G K A (2002). Genetic manipulation of the oil palm - Challenges and prospects. *The Planter*, 78: 547-562.

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: 6471-6483.

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

Shi, L; Katavic, V; Yu, Y; Kunst, L and Haughn, G (2012). *Arabidopsis* glabra2 mutant seeds deficient in mucilage biosynthesis produce more oil. *Plant J.*, 69: 37-46.

Singh, R; Ong-Abdullah, M; Low, E T L; Manaf, M A A; Rosli, R; Nookiah, R; Ooi, L C 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 interfertile species in Old and New worlds. *Nature*, 500: 335-339.

Siti Nor Akmar, A and Zubaidah, R (2008). Mesocarp-specific metallothionein-like gene promoter for genetic engineering of oil palm. *J. Oil Palm Res. Special Issue (July 2008)*: 1-8.

Stone, S L; Braybrook, S A; Paula, S L; Kwong, L W; Meuser, J; Pelletier, J; Hsieh, T F; Fischer, R L; Goldberg, R B and Harada, J J (2008). *Arabidopsis* leafy cotyledon2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci.*, 105: 3151-3156.

- Tarmizi, A H A; Ismail, R and Kuntom, A (2016). Effect of frying on the palm oil quality attributes - A review. *J. Oil Palm Res. Vol. 28*: 143-153.
- 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 Genet. Genomes*, 9: 377-386.
- Tranbarger, T J; Dussert, S; Joet, 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.
- Untergasser, A; Cutcutache, I; Koressaar, T; Ye, J; Faircloth, B C; Remm, M and Rozen, S G (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res.*, 40 (15): e115.
- USDA (2017). *World Agricultural Production 2017*. Circular Series WAP 05-17. <http://usda.mannlib.cornell.edu/usda/current/worldag-production/worldag-production-08-10-2017.pdf>, accessed on 15 September 2017.
- Vaucheret, H; Beclin, C and Fagard, M (2001). Post-transcriptional gene silencing in plants. *J. Cell Sci.*, 114: 3083-3091.
- Wang, H W; Zhang, B; Hao, Y J; Huang, J; Tian, A G; Liao, Y; Zhang, J S and Chen, S Y (2007). The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic *Arabidopsis* plants. *Plant J.*, 52: 716-729.
- Werber, M; Wittler, L; Timmermann, B; Grote, P and Herrmann, B G (2014). The tissue-specific transcriptomic landscape of the mid-gestational mouse embryo. *Development*, 141: 2325-2330.
- Yeap, W C; Loo, J M; Wong, Y C and Harikrishna, K (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, L; Ma, J; Niu, Z; Bai, X; Lei, W; Shao, X; Chen, N; Zhou, F and Wan, D (2017). Tissue-specific transcriptome analysis reveals multiple responses to salt stress in *Populus euphratica* seedlings. *Genes (Basel)*, 8(12): E372.
- Yunus, A M M; Chai Ling, H and Parveez, G K A (2008). Construction of phb and phbv transformation vectors for bioplastics production in oil palm. *J. Oil Palm Res. Special Issue (July 2008)*: 37-55.
- Yunus, A M M and Kadir, A P G (2008). Development of transformation vectors for the production of potentially high oleate transgenic oil palm. *Electron. J. Biotechnol.*, 11: 1-9.
- Zeka, F; Vanderheyden, K; De Smet, E; Cuvelier, C A; Mestdagh, P and Vandesompele, J (2016). Straightforward and sensitive RT-qPCR based gene expression analysis of FFPE samples. *Sci. Rep.*, 6: 21418.
- Zubaidah, R; Nurniwalis, A W; Chan, P L; Masura, S S; Siti Nor Akmar, A and Parveez, G K A (2018). Tissue-specific promoters: The importance and potential application for genetic engineering in oil palm. *J. Oil Palm Res. Vol. 30(1)*: 1-12.
- Zulkifli, Y; Norziha, A; Naquiuddin, 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. Vol. 29(4)*: 440-455.