Journal of Oil Palm Research Special Issue on Malaysia-MIT Biotechnology Partnership Programme: Volume 2 - Oil Palm Metabolic Engineering (July 2008) p. 87-96

GENE DISCOVERY VIA EXPRESSED SEQUENCE TAGS FROM THE OIL PALM (*Elaeis guineensis* Jacq.) MESOCARP

NURNIWALIS, A W*; SUHAIMI, N; SITI NOR AKMAR, A[+]; AMINAH, S* and MOHAMAD ARIF, M A***

ABSTRACT

Expressed sequence tags (ESTs) have been used for many applications such as to reveal gene expression patterns, gene regulation and sequence diversity. A total of 1011 unique transcripts corresponding to 1463 genes have been identified from the ESTs generated from 17-week oil palm mesocarp cDNA library. This approach was found to be successful in the discovery of new and important genes expressed in the mesocarp tissue which are associated with the various cellular processes of the tissue. It was observed that 12.9% of the total genes expressed in the 17-week mesocarp cDNA library can be categorized under metabolism. This is in agreement with the function of the tissue which is involved in many biochemical processes including amino acid and fatty acid metabolism. Most importantly are the discoveries of genes playing important roles in the fatty acid and wax biosynthesis pathway such as acetyl-CoA carboxylase, stearoyl-ACP desaturase, acyl carrier protein (ACP), lysophosphatidic acid acyltransferase, Δ^6-palmitoyl-ACP desaturase and lipase. These genes can serve as targets for genetic manipulation where such endeavours have been extensively carried out in other plants such as Brassica napus *and* Olea europaea *to help increase the economic value of the oil. Genes and protein associated with ethylene synthesis and signal transduction pathway were also identified from the 17-week mesocarp ESTs. Dot blot analysis was carried out to help in identifying potential tissue-specific genes, which can lead to the isolation of the tissue-specific promoters for manipulation of the mesocarp tissue. This is in particular to direct accumulation of transgenic products, such as new specialty oils and value-added products like pharmaceuticals and nutraceuticals to the mesocarp.*

Keywords: expressed sequence tags (EST), mesocarp, oil palm, *Elaeis guineensis*, tissue-specific genes.

Date received: 31 July 2007; **Sent for revision:** 17 August 2007; **Received in final form:** 25 October 2007; **Accepted:** 7 November 2007.

* Malaysian Palm Oil Board,
P. O. Box 10620,
50720 Kuala Lumpur,
Malaysia.
E-mail: nurni@mpob.gov.my

** Department of Cell and Molecular Biology,
Faculty of Biotechnology and Biomolecular Sciences,
Universiti Putra Malaysia,
43400 UPM Serdang,
Selangor, Malaysia.

[+] Department of Agriculture Technology,
Faculty of Agriculture,
Universiti Putra Malaysia,
43400 UPM Serdang,
Selangor, Malaysia.

INTRODUCTION

The oil palm is the most productive of all oil crops. Being a perennial with a productive life-span of 20-30 years, it has significant advantages over other crops for diverse applications to increase its economic value. Plant genetic engineering offers unique opportunities to improve the versatility and utility of the oil palm and palm oil. It can be used to improve the oil quality and modify the palm to produce high value products, *e.g.*, nutraceuticals and pharmaceuticals (Cheah, 2004).

Palm oil is synthesized in the fruit mesocarp. Thus, this tissue should be targeted for introducing novel characteristics, such as for the production of unique products. The mesocarp can be manipulated

by directing the substrates and intermediates for the production of storage oil or protein to it, changing the levels of existing products or causing novel high value products to be produced without harming the plant (Mohd Basri *et al.*, 2004). Oil palm genetic manipulation can be done by expressing the appropriate transgenes by a tissue-specific or constitutive promoter. A promoter, located upstream of the coding region, is the DNA sequence required to determine the appropriate spatial and temporal expression pattern of a particular gene. Hence, the isolation of genes based on gene expression patterns in specific organs/tissues or at specific developmental stages is an important step towards the isolation of their respective promoters or regulatory sequences.

Expressed sequence tags (ESTs) have been used for many applications, such as to reveal gene expression patterns, gene regulation and sequence diversity on a large scale. ESTs also have potential application for the discovery of important tissue-specific genes for use as targets for genetic manipulation. The use of tissue-specific cDNA libraries not only enables identification of the genes responsible for various biochemical and physiological processes in the tissue but can also lead to discovery of tissue-specific promoters. ESTs generated from pine, specifically from the immature xylem, have proven to be an excellent source of genes with which to study wood formation in conifers, and have resulted in the isolation of cDNA encoding cell wall biosynthetic enzymes (Stersky *et al.*, 1998). The genes encoding the respective enzymes are shown to be up-regulated during xylem formation and have the potential to be used as commercial targets. Xylem-specific promoters of the relevant genes may be useful to reveal the signal transduction pathway that leads to xylem formation (Allona *et al.*, 1998). The EST approach has also resulted in identification of a cDNA encoding a glutelin gene from an oil palm kernel cDNA library (Cheah *et al.*, 2000). The expression of this gene was found to be kernel-specific and the corresponding gene promoter successfully isolated and characterized to target the expression of transgenes for oil palm genetic manipulation to this tissue (Siti Nor Akmar *et al.*, 2007).

The EST approach for gene discovery is very useful for crops of economic importance, such as the oil palm. To date, ESTs have been generated from various oil tissues such as 15-week mesocarp (Siti Nor Akmar *et al.*, 2003), embryo (Halimah, 2005), callus (Tan, 2003), inflorescence (Choi, 2003), kernel as well as young etiolated seedlings. ESTs from various tissues have been generated to accommodate various research such as the rapid discovery of new genes and to understand the complexity of genes expressed in a specific tissue. The ESTs are also incorporated in the DNA Microarray project as well as in the ongoing genome mapping programme with the objective to develop a diagnostic tool for tissue culture amenity and abnormality (Low *et al.*, 2005).

This work is aimed at the identification of genes playing important roles during oil palm fruit development with regards to oil synthesis as well as identification of potential tissue-specific genes from the oil palm mesocarp. Identification of novel or tissue-specific genes provides new candidates for the isolation of strong tissue-specific promoters. These promoters can be use as tools for oil palm genetic manipulation to produce novel and high-value products to increase the economic value for the oil palm.

MATERIALS AND METHODS

Plant Material

Various *E. guineensis* variety *tenera* tissues were collected, frozen in liquid nitrogen and stored at -70°C for future RNA extraction and EST library construction. Fresh fruit bunches were harvested at various weeks after anthesis (WAA) for their mesocarp and kernel tissues. Roots were obtained from 2-year-old seedlings while the germinated seedlings were 1-week-old.

Construction of 17-week Mesocarp cDNA Library

Total RNA and mRNA were extracted as described by Nurniwalis and Siti Nor Akmar (2001). The cDNA was synthesized from 5 μg of 17-week mesocarp mRNA, and the cDNA library constructed using the ZAP-cDNA Gigapack III Gold Cloning Kit (Stratagene, La Jolla, USA) following the protocol of the manufacturer. The cDNA library was constructed using SOLR as host, with pBluescript as vector and cloning sites *Eco*R I/*Xho* I.

Plasmid DNA Isolation and DNA Sequencing

Single and well isolated recombinant clear plaques were randomly selected and subjected to *in vivo* excision. Single *in vivo* excisions were carried out using ExAssist Helper Phage with SOLR strain according to the manufacturer's instructions. High throughput plasmid DNA isolation was carried out by the MPOB Robotics Service using the Montage Plasmid Miniprep$_{96}$ Kit (Millipore Corporation) according to the manufacturer's instructions for automation on the ROBOSMART-384 robotics system (MWG Inc.). DNA sequencing was carried out using the Sequencing Service provided by the Genome Analysis Laboratory for Oil Palm, MPOB.

The sequencing reactions were carried out using ABI PRISM™ Big Dye™Version 2.0 Terminator Ready Reaction Mix (Applied Biosystem, Foster City, USA) according to the manufacturer's instruction. Plasmid/phagemid DNA was used as the template for sequencing and the concentration for the cycle sequencing reactions ranged between 50-500 ng. Sequencing the cDNA clones was carried out from the 5' end using SK or T3 universal primer. Reactions were run and analysed on an ABI Prism 377 automated DNA Analyzer from Applied Biosystem.

EST Sequence Processing and Analysis

The software Phred (Ewing *et al.*, 1998) was used to screen out low quality DNA sequences. A Phred score of 20 was used to indicate the probability of 1 base error for every 100 bases in the DNA sequence. Vector sequences were masked with Cross-Match, the masked sequences then manually edited to remove the vector sequence, poly $(A)^+$ tail and other ambiguous sequences. After screening, the sequences were clustered using the stackPACK™pipeline (http://www.sanbi.ac.za) which had multiple steps, comprising clustering, assembly of clusters and consensus sequences. The sequences were clustered in the same group when the criterion to define nucleotide identity of 96% for more than 100 bp was attained. In addition, the sequences of each contig were aligned by the Phrap program and consensus sequences generated by the Craw program with 90% identity over a minimum of 50 nucleotides. Groups that contained only one sequence were classified as singletons.

Database Searches and Functional Characterization

The edited EST sequences were compared with other known sequences in the GenBank NR database using the BatchBLAST program (Altschul *et al.*, 1990). The program (Blastc13 Version 2.2) allows batches of sequences to be compared to known genes in the database simultaneously. To assign functions, BLASTX searches were performed at a threshold of BLAST scores >50 or E-value $<1.0E^{-05}$. Classification of ESTs into functional categories were according to the Munich Information Centre for Protein Sequences (MIPS) *Arabidopsis thaliana* Database (MAtDB) Functional Catalogue (FunCat version 2.0) (Schoof *et al.*, 2002; Ruepp *et al.*, 2004). The Gene Ontology (GO) system developed by the GO Consortium (Wortman *et al.*, 2003) was also used for further annotation and classification of the ESTs.

Dot Blot Analysis

About 10 μg total RNA were denatured, separated by gel electrophoresis and transferred onto positively charged nylon membranes (Amersham). Preparation and blotting of RNA samples were carried out as described by Nurniwalis (2006). Fixation of RNA to the membrane was by UV irradiation. The blotting manifold was set up using HYBRI DOT Manifold. Several ESTs were selected and used to probe the dot blots containing total RNA from various oil palm tissues. Plasmids were isolated from the selected ESTs using Qiagen Plasmid Isolation Kit and double digested with *Eco*R I and *Xho* I (Fermentas). The fragments were purified using a Qiaquick Gel Extraction Kit and labelled with α^{-32}P dCTP using a Megaprime DNA Labelling System Kit (Amersham) following the manufacturer's instructions. The membranes were pre-hybridized for 4 hr and hybridized overnight at 65°C in hybridization buffer (5X SSC, 5X Denhardt, 0.5% SDS). Washing was carried out twice in 2X SSC, 0.1% SDS for 10 min and finally in 0.5XSSC, 0.1% SDS for 30 min at 65°C prior to overnight exposure to X-ray film at –80°C.

RESULTS AND DISCUSSION

Quality of ESTs, DNA Sequencing and EST Clustering

A 17-week mesocarp cDNA library was constructed to discover potential genes that could be manipulated to increase the economic value of oil palm. Three thousand and five hundred individual and well isolated recombinant plaques were randomly selected for PCR analyses. PCR amplification of the recombinant plaques using T3 and T7 universal primers revealed that the size of the inserts was 0.2 kb to 2.5 kb. Clones that displayed clear, single PCR products with inserts larger than 350 bp were then subjected to *in vivo* excision and plasmid isolation prior to sequencing. The detailed list of the quality control analyses on the ESTs is shown in *Table 1.* The average read length of the edited sequences was 315 bp, comparable to those from EST projects, such as from the autumn leaf of

TABLE 1. GENERAL ANALYSIS ON 17-WEEK MESOCARP EXPRESSED SEQUENCE TAG (ESTS)

Clones subjected to analysis	Number of clones
Primary titre (pfu)	2.15×10^5
Amplified titre (pfu ml^{-1})	3.07×10^{10}
Clones selected at random and PCR amplification	3 500
DNA sequencing	2 114
Phred and Cross_Match analysis	2 049
Sequence passed quality check	1 463
StackPACK analysis	1 463

Populus tremula and young leaf of *Populus tremula* x *tremuloides* cDNA libraries whose average read lengths are 349 bp and 355 bp, respectively (Bhalerao *et al.*, 2003). Clustering of 1463 ESTs sequences resulted in the identification of 1011 unique transcripts. Assembling the EST sequences into contigs revealed at least 57.5% of the clones appeared to be singletons, and the remaining 42.5% grouped into 169 assemblies, most of which contained two ESTs per contig.

Gene Annotation and Functional Classification

Annotation of the 17-week mesocarp sequences was based on sequence similarity to an already identified homologous sequence in the Genbank non-redundant protein database. Identified sequence similarities with top scoring hits of more than 50 or E value of $< 10^{-5}$ were considered statistically significant for sequence compared against BLASTX (Zhu *et al.*, 2001). From the clustering analysis of the EST sequences, 48.0% showed significant similarities to known genes in the public database, 10.2% did not show any significant similarities while the remaining 41.8% were considered to have no match (*Figure 1*).

It would be surreal to imply that all of the 'undiscovered' ESTs encode for novel proteins with no match to those in the public sequence databases. The high number of undiscovered sequences is most likely due to the presence of insufficient or short coding sequences. Thus, an accurate identity assigned by an annotation based on homology may not be established. It is also possible that some of these unidentified sequences may not encode proteins, but instead may function as RNA molecules (Moyle *et al.*, 2005). Nevertheless, among the undiscovered sequences, there may be a subset of genes containing novel coding sequences that have not previously been discovered. Thus, these sequences may contain valuable information for future research.

To obtain an insight into the cellular activities of oil palm fruits at 17 WAA, the functions of the 48% database matched genes representing 684 EST sequences were assigned into 16 different functional role categories based on the *Arabidopsis thaliana* database following the MIPS FunCat schema available at http://mpis.gsf.de. Classifications of the mesocarp ESTs were assigned based on the assumption that functionality between organisms is transferable based on sequence conservation (van der Hoevan *et al.*, 2002). As shown in *Table 2*, the highest percentage (34.6%) of the ESTs with significant hits to known sequences in the public database corresponds to 237 genes and is categorized as 'unable to classify'. These clones were found to have similarities to either unknown proteins, hypothetical proteins or putative proteins with no indication of the function of these genes. Genes which roles were unknown or could not be assigned into suitable functional categories with confidence based on searches in the literature were also grouped in this category. Next, the cell rescue, defence and virulence related genes accounted for 15.4%, followed by 12.9% of the metabolism related genes in the cDNA clones.

The EST sequences were found to match plant genes, including those in rice, *Arabidopsis*, oil palm, maize, tomato, pineapple, tobacco and potato. A majority of the sequences (34.9%) were found to match those from rice. The high degree of sequence homology may reflect conserved gene functions among the monocots. Next, the second highest

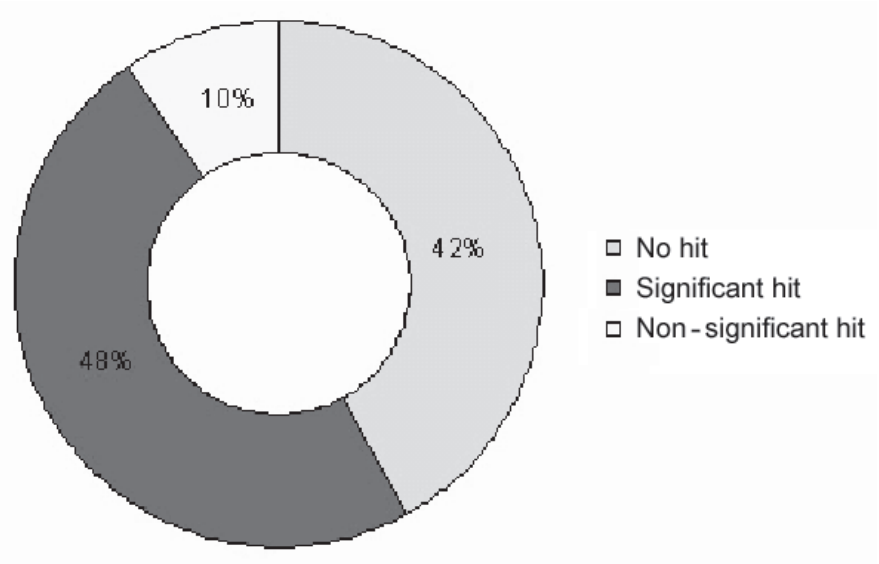

Figure 1. Categorization of the expressed sequence tag (EST) clones based on the hits to known sequences in the public databases. Each category is represented by different coloured legends.

TABLE 2. MIPS FUNCAT ANALYSIS OF 17-WEEK MESOCARP EXPRESSED SEQUENCE TAG (TAG) SEQUENCES

No.	Functional category	% of 17-week mesocarp sequences
01	Metabolism	12.9
02	Energy	3.8
10	Cell cycle and DNA processing	0.2
11	Transcription	5.6
12	Protein synthesis	12.0
14	Protein fate (folding, modification and destination)	5.0
16	Protein with binding function or cofactor requirement	1.0
20	Cellular transport, transport facilitation and route	3.8
30	Cellular communication/signal transduction mechanism	1.7
32	Cell rescue, defence and virulence	15.4
34	Interaction with the cellular environment	0.2
38	Transposable elements, viral and plasmid proteins	2.2
41	Development	1.0
42	Biogenesis of cellular components	0.6
99	Unclassified proteins	34.6

significant EST sequences match were to *Arabidopsis* (20.3%) which may suggest that they are evolutionarily conserved (Van der Hoeven *et al*, 2002). The overall percentages of significant homology match distribution to known genes from different organisms are shown in *Figure 2*.

Metallothionein-like proteins were found to be the most abundant genes expressed in the 17-week mesocarp cDNA library. Similarly, metallothionein-like proteins have also been found to be in abundance in sesame (Suh *et al.*, 2003) and pineapple fruit cDNA libraries (Moyle *et al.*, 2005). Metallothionein gene transcripts were also found to be highly abundant in the EST collection from 15-week oil palm mesocarp (Siti Nor Akmar *et al.*, 2003), somatic embryos (Halimah, 2005) and non-embryogenic calli (Tan, 2003). It is not really clear why metallothioneins are up-regulated in these tissues, but a possible connection between the up-regulation of this transcript in these tissues could be due to various stress responses regulated by a common causal and regulatory mechanisms. The highly abundant metallothionein-like protein was found to be 100% identical at the nucleotide and protein level to a metallothionein-like protein gene (MT3-A) previously isolated from the oil palm via subtractive hybridization (Siti Nor Akmar, 1999). The MT3-A was found to have a strong binding affinity to zinc when expressed as a gluthathione-S-transferase (GST) fusion protein, suggesting that this gene may have a homeostatic role for Zn^{2+} ions in the mesocarp during the second half of the fruit maturing stage.

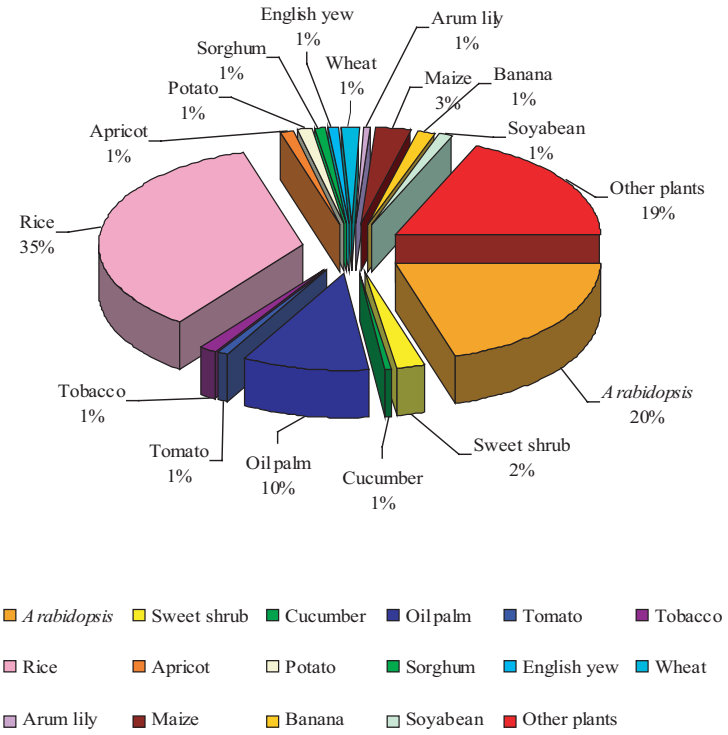

Figure 2. Significant homology to known genes of different organisms. Different organisms are represented by different coloured legends.

Discovery of Potential Genes for Future Manipulation

The mesocarp tissue plays a major role in the productivity of the oil palm. It produces a storage oil (palm oil) which contains equal amounts of saturated and unsaturated fatty acids. In the 17-week mesocarp cDNA library, EST sequences related to metabolism account for approximately 12.9% of the total genes representing the third largest functional category. In agreement with the function of the mesocarp, various genes involved in primary and secondary metabolism were discovered including those involved in lipid, fatty acid, isoprenoid and amino acid metabolism (*Table 3*). Manipulation of several of these genes has been extensively carried

TABLE 3. EXPRESSED SEQUENCE TAG (EST) SEQUENCES ENCODING ENZYMES AND PROTEINS INVOLVED IN LIPID, FATTY ACID AND ISOPRENOID METABOLISM AS WELL AS ETHYLENE SYNTHESIS AND SIGNAL TRANSDUCTION PATHWAY

Putative ID	Accession No.	Celullar function	Reference	Organism
Acetyl-CoA carboxylase (ACCase)	P0C2Y4	Fatty acid biosynthesis	Beisson *et al.*, (2003)	*O. sativa*
Acyl carrier protein 1, chloroplast precursor (ACP 1)	ACP1_CASGL	Fatty acid biosynthesis	Beisson *et al.*, (2003)	*C. glauca*
Acyl-CoA oxidase homolog.	T07901	Degradation of storage lipids	Beisson *et al.*, (2003)	*Cucurbita* sp.
ATP-citrate synthase [ATP-citrate (pro-S-)-lyase]	NP_172414.1	Micellaneous	Beisson *et al.*, (2003)	*A. thaliana*
Epoxide hydrolase	AAM28292.1	Micellaneous	Beisson *et al.*, (2003)	*A. comosus*
Ethylene-forming enzyme, putative	NP_922709.1	Ethylene synthesis	Alexander and Grierson (2002)	*O. sativa*
Ethylene receptor-like protein	AAL86614.1	Ethylene signal transduction	Alexander and Grierson (2002)	*L. esculentum*
Ethylene responsive element binding protein	CAD21849.1	Ethylene signal transduction	Alexander and Grierson (2002)	*F. sylvatica*
Ethylene-responsive factor-like protein 1	AAT75013.1	Ethylene signal transduction	Alexander and Grierson (2002)	*Z. mays*
Lipase class 3 family protein	NP_566484.2	Degradation of storage lipids	Beisson *et al.*, (2003)	*A. thaliana*
Phosphatidylinositol 3- and 4-kinase family protein	NP_179383.1	Lipid signalling	Beisson *et al.*, (2003)	*A. thaliana*
Phosphatidylserine decarboxylase, putative	NP_914239.1	Lipid synthesis	Beisson *et al.*, (2003)	*O. sativa*
Probable 1-acyl-glycerol-3-phosphate acyltransferase (LPAAT)	S52645	Fatty acid biosynthesis	Beisson *et al.*, (2003)	*Z. mays*
Putative beta-ketoacyl-CoA synthase (KCS)	BAD32939.1	Wax biosynthesis	Beisson *et al.*, (2003)	*O. sativa*
Putative lipase homolog	NP922709.1	Degradation of storage lipids	Beisson *et al.*, (2003)	*O. sativa*
Putative 2C-methyl-D-erythritol 2, 4-cyclodiphosphate synthase	BAD29384.1	Lycopene biosyntesis	Bartley and Ishida (2002)	*O. sativa*
Δ^6-palmitoyl-ACP desaturase	AAA82160.1	Fatty acid biosynthesis	Cahoon *et al.*, (1994)	*T. alata*
Δ^9-stearoyl-ACP desaturase (SAD)	TCTP_ORYSA	Fatty acid biosynthesis	Beisson *et al.*, (2003)	*O. sativa*

out in other plants, such as *Brassica napus* and *Olea europaea,* to increase the value of their oils.

Several genes encoding enzymes and proteins involved in ethylene synthesis and the signal transduction pathway were also present in this library. The phytohormone ethylene has been shown to play important roles in many aspects of plant growth and development which includes regulation of fruit ripening (Abeles *et al*., 1992). The possibility that ethylene might be responsible to trigger oil production in the mesocarp and other changes associated with it has been raised, but evidence that ethylene is in fact generated is still lacking (Mohd Haniff, 1982). Thus, it would be wise to explore the possibility of ethylene playing important roles in regulating the ripening process of oil palm fruits.

Dot Blot Analysis

Several genes of interest in fatty acid biosynthesis, degradation of storage lipids and plant signal transduction such as ACP, lipase class 3 family protein and ethylene receptor-like protein were selected and their expression patterns determined by dot blot analysis. Dot blot analysis was carried out to identify more tissue-specific genes from the oil palm mesocarp which can lead to the isolation of its tissue-specific promoters for manipulation of the mesocarp tissue. The inserts of the respective EST clones were then used to probe positively charged nylon membranes containing total RNA from various oil palm tissues. This included tissues from the mesocarp at different stages of development (8, 10, 12, 15, 17, 20 WAA), kernel (10 and 12 WAA), spear leaves, roots, germinated seedlings and the flowers. The expression patterns of the selected genes are shown in *Figure 3*. Two ACP clones, designated Q74EST and V79EST ACP, were shown to differ in their patterns of expression. The Q74EST was expressed constitutively in all the tested tissues with the highest expression in the mesocarp and kernel tissues, especially the stage where oil synthesis starts to occur. The result obtained was similar to that of Rasid *et al*. (1999), and this finding may well coincide with the fact that this two tissues are the sites for storage lipid synthesis. On the other hand, V79EST was expressed at a low level in all of the tested tissues. The ACP exists in multiple isoforms and there are at least three isoforms of ACP-types I, II and III encoded by a multigene family. Some ACPs are constitutively expressed while some are developmentally and tissue-specifically regulated (Voetz *et al*., 1994). Thus, the different expression patterns between two clones might suggest that the clones are most likely different ACP isoforms.

In rapeseed and soyabean, ACP activity has been found to appear just before storage lipid biosynthesis, indicating that the genes are temporally regulated, thus can be used as temporal markers. The ACPs were found to be in abundance in the 17-week mesocarp cDNA library as compared to other related lipid metabolism genes. Manipulation of the different ACP isoforms in combination with different enzymes in the biosynthetic pathway is most likely useful for genetic manipulation of oil palm to help increase its economic value.

Another unique gene encoding for lipase class 3 family protein (O65EST) was also selected to determine its expression pattern in oil palm tissues. As shown in *Figure 3*, O65EST was expressed highly in the mesocarp tissues especially during oil synthesis. A trace level of expression was detected in germinating seedlings and leaves whereas no expression was detected in the rest of the tested

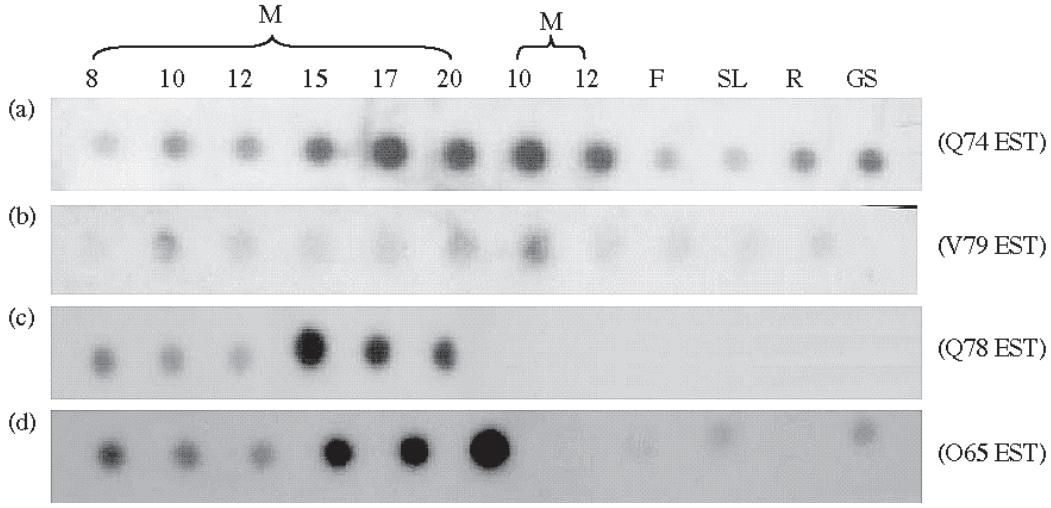

Figure 3. Tissue-specific expressions of four expressed sequence tag (EST) clones determined using dot blot analysis. The respective clones selected from the 17-week mesocarp cDNA library are (a) acyl-carrier protein [Q74 EST],(b) acyl-carrier protein [V79 EST], (c) ethylene receptor-like protein [Q78 EST] and (d) lipase class 3 family protein [O65 EST]. M = mesocarp, K = kernel, F = flower, SL = spear leaf, R = roots and GS = germinated seedlings. The numbers represent the developmental stages of the mesocarp and kernel tissues between 8 to 20 weeks after anthesis.

tissues. In plants, TAG is stored for germination and post-germination growth of seedlings. Upon germination, lipase catalyzes the breakdown of TAGs and the carbon skeletons are used for post-germinative growth (Murphy, 1993). Lipase has also been shown to be involved in the breakdown of membrane lipids in leaves especially during senescence (Wang, 2001). Nevertheless, the mesocarp is a very rich source of TAG especially in ripen fruits. The enzyme lipase in the mesocarp was found to be induced towards the ripening period which also coincides with the period of oil synthesis, *i.e.* between 16 WAA to 21 WAA (Sambanthamurthi *et al.*, 1994).

In oil palm, the mesocarp tissue contains an active endogenous lipase. Hydrolysis of TAG is induced by fruit bruising and poor storage handling. Damage on the tissues allows the endogenous lipase access to the oil droplets, degrading the storage lipids and resulting in the release of FFA. Oxidation of FFA causes undesirable loss of yield as the quality of the palm oil drops drastically due to rancidity and oil mutilation. Thus, to minimize losses of oil quality due to lipase activity, efforts to screen oil palm germplasm for palms with lower lipase activities is being carried out (Wong *et al.*, 2005). Nevertheless, information on the active lipase activity in the mesocarp can also be useful to isolate promoter of this gene. Isolation of a tissue-specific promoter would be useful to target the accumulation of transgenic products, such as pharmaceuticals and nutraceuticals, to the mesocarp tissue via genetic manipulation.

Besides O65EST, an ethylene receptor designated Q78EST was found to show putative tissue-specific expression in the oil palm mesocarp tissue with no mRNA accumulation observed in other tested tissues. The expression in the mesocarp tissues was observed to increase differentially during oil palm fruit development. The mRNA transcript was detected at an early fruit developmental stage (young fruit) with an increase in expression at the onset of oil palm fruit ripening and the expression decreased slightly when the fruit has fully ripen. The pattern of expression was found to be similar to the period when oil starts to be synthesized in this tissue until the stage when the oil synthesis process terminates. The expression of Q78EST in the tested mesocarp tissues was similar to the ethylene receptor gene isolated via RT-PCR (Nurniwalis *et al.*, 2003). In addition, an increase in the transcript level of the ERS1-type ethylene receptors has also been observed during the ripening process in fruits such as tomato (Payton *et al.*, 1996), peach (Rasori *et al.*, 2002). The increase in the transcript level during ripening is most likely due to a tissue-specific response to a heightened need for receptor molecules to facilitate the continued modulation of ethylene responsiveness in ripening fruit tissues (Giovannoni, 2004).

The similar expression pattern of the ethylene receptor ERS1-type in climacteric fruits and O65EST in oil palm indicates that there is a strong co-relation of the gene in regulating the ripening process of the fruits. In climacteric fruits such as tomato (Alexander and Grierson, 2002), the role of ethylene in triggering the ripening process is well established. In these fruits, ripening is associated with a sudden peak in the respiration rate and ethylene production followed by changes in texture, colour and softening of the flesh with an increase in the sweetness, flavour and aroma (Giovannoni, 2001). These changes may differ from oil palm whereby ripening of the oil palm fruit is basically associated with the accumulation of storage oil as oil bodies in both storage tissues, *i.e.*, the mesocarp and the kernel. The role of ethylene in the ripening of oil palm fruits is still unclear; however, ethylene has been detected in air samples taken from the periphery of ripening oil palm fruit bunches (Mohd Haniff, 1982).

Information on the active involvement of an ethylene receptor gene in the mesocarp fruit development from young towards ripening can be manipulated for future isolation of its corresponding promoter. This promoter can be useful for regulating the temporal and spatial expression of genes controlling important economic traits in transgenic plants.

CONCLUSION

We hereby report that single pass partial sequencing of 17-week mesocarp cDNAs was successful in the generation of 1011 unique transcripts corresponding to 1463 genes. Most importantly are the discoveries of new and important genes playing important roles associated with the functions of the mesocarp tissue. These genes can be further characterized for future manipulation of the oil palm for production of novel products. Two clones coding for a lipase class 3 family protein and an ethylene receptor were also found to show very high expression pattern in the mesocarp via dot blot analysis. Further characterization of these genes would be useful for future isolation of their corresponding promoters for manipulation of the mesocarp tissue via genetic engineering.

ACKNOWLEDGEMENT

The authors wish to thank the Director-General of MPOB for permission to publish this article. The authors would also like to thank Dr Ho Chai Ling from Universiti Putra Malaysia, the staff of the Gene Expression Group, particularly Mr Mahadzir Jaafar, and the Genomics Group of MPOB for their technical help and assistance.

REFERENCES

ABELES, F B; MORGAN, P W and SALTVEIT, M E J (1992). *Ethylene in Plant Biology*. 2nd ed. Academic Press, San Diego.

ALEXANDER, L and GRIERSON, D (2002). Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. *J. Exp. Bot., 53*: 2039-2055.

ALLONA, I; QUINN, M; SHOOP, E; SWOPE, K; CYR, S S; CARLIS, J; RIEDL, J; RETZEL, E; CAMPBELL, M M; SEDEROFF, R and WHETTEN, R W (1998). Analysis of xylem formation in pine by cDNA sequencing. *Proc. Natl. Acad. Sci. USA, 95*: 9693–9698.

ALTSCHUL, S F; GISH, W; MILLER, W; MYERS, E W and LIPMAN, D J (1990). Basic local alignment search tool. *J. Mol. Biol., 215*: 403-410.

BARTLEY, G E and ISHIDA, B K (2002). Digital fruit ripening: data mining in the TIGR tomato gene index. *Plant. Mol. Biol. Rep., 20*: 115-130.

BEISSON, F; KOO, A J K; RUUSKA, S; SCHWENDER, J; POLLARD, M; THELEN, J J; PADDOCK, T; SALAS, J J; SAVAGE, L; MILCAMPS, A; MHASKE, V B; CHO, Y and OHLROGGE, J B (2003). *Arabidopsis* genes involved in acyl lipid metabolism: a 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol., 132*: 681-697.

BHALERAO, R; KESKITALO, J; STERKY, F; ERLANDSSON, R; BJÖRKBACKA, H; IRVE, S J; KARLSSON, J; GARDESTRÖM, P; GUSTAFSSON, P; LUNDEBERG, J and JANSSON, S (2003). Gene expression in autumn leaves. *Plant Physiol., 131*: 430–442.

CAHOON, E B; CRANMER, A M; SHANKLIN, J and OHLROGGE, J B (1994). Δ^6 hexadecenoic acid is synthesized by the activity of a soluble Δ^6 palmitoyl-acyl protein desaturase in *Thunbergia alata* endosprem. *J. Bio. Chem., 269*: 27519-27526.

CHEAH, S C; MARIA, M and RAJINDER, S (2000). Oil palm genomics. *Advances in Oil Palm Research* (Yusof, B; Jalani, B S and Chan, K W eds.). Volume 1. MPOB, Bangi. p. 332-370.

CHEAH, S C (2004). Application of biotechnology to improvement in oil palm and its challenges. *Biotechnology Symposium 2004*. Manilla Pavillion, Philippines.

CHOI, M C (2003). *Isolation of Expressed Sequence Tags (ESTs) from an Oil Palm* (Elaeis guineensis *Jacq.*) *Male and Female Floral cDNA Library*. M.Sc. thesis. Universiti Putra Malaysia.

EWING, B; HILLIER, L; WENDL, M and GREEN, P (1998). Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res., 8*: 175-185.

GIOVANNONI, J J (2004). Genetic regulation of fruit development and ripening. *Plant Cell., 16*: S170-S180.

GIOVANNONI, J J (2001). Molecular regulation of fruit ripening. *Annu. Rev. Plant Physiol. Plant Mol. Biol., 52*: 725–749.

HALIMAH, A (2005). *Penjanaan dan Analisis Penanda Jujukan Terungkap (EST) daripada Embrio Somatik Kelapa Sawit* (*Elaeis guineensis* Jacq.). M.Sc. thesis. Universiti Kebangsaan Malaysia.

LOW, E T L; TAN, J S; CHAN, P L; BOON, S H; WONG, Y L; ROZANA, R; OOI, L C-L; ONG-ABDULLAH, M; CHEAH, S C and RAJINDER, S (2005). Application of DNA chip towards developing a diagnostic tool for tissue culture. *Proc. of the PIPOC 2005 International Palm Oil Congress - Agriculture Conferenc*, MPOB, Bangi. p. 562-579.

MOHD BASRI, W; SITI NOR AKMAR, A and HENSON, I E (2004). Oil palm - achievements and potential. *Proc. of the 4th International Crop Science Congress*. p. 1-13.

MOHD HANIFF, H (1982). Effects of growth regulators on oil palm growth and productivity. *PORIM Annual Research Report 1982*. PORIM, Bangi, Selangor.

MOYLE, R; FAIRBAIRN, D; RIPI, J; CROWE, M and BOTELLA, J R (2005). Developing pineapple fruit has a small transcriptome dominated by metallothionein. *J. Exp. Bot., 56*: 101–112.

MURPHY, D J (1993). Plant lipids: their metabolism, function and utilization. *Plant Biochemistry and Molecular Biology* (Lea, P. J. and Leegood, R. C eds.). John Wiley and sons publishers. p. 113-128.

NURNIWALIS, A W (2006). *Isolation of Ethylene Response Sensor Gene and Generation of Expressed Sequence Tags from the Oil Palm* (Elaeis guineensis *Jacq.*) *Mesocarp*. M. Sc. thesis. Universiti Putra Malaysia.

NURNIWALIS, A W; SITI NOR AKMAR, A and TAN, S H (2003). Isolation of mesocarp-specific cDNA clone from oil palm *via* RT-PCR. *Proc. of the 2003 International Palm Oil Congress - Agriculture Conference*. MPOB, Bangi. p. 766-772.

NURNIWALIS, A W and SITI NOR AKMAR, A (2001). Isolation of differentially expressed genes in the oil palm mesocarp by differential display and differential screening. *Proc. of the 2001 PIPOC International Palm Oil Congress - Agriculture Conference*. MPOB, Bangi. p. 450-455.

PAYTON, S; FRAY, R G; BROWN, S and GRIERSON, D (1996). Ethylene receptor expression is regulated during fruit ripening, flower senescence and abscission. *Plant Mol. Biol., 31*: 1227–1231.

RASID, O A; CHEAH, S C and MOHD ARIF, M A (1999). Isolation and sequencing of cDNA clones coding for oil palm (*Elaeis guineensis*) acyl carrier protein (ACP*). J. Oil Palm Research (Special Issue*): 88-95.

RASORI, A; RUPERTI, B; BONGHI, C; TONUTTI, P and RAMINA, A (2002). Characterization of two putative ethylene receptor genes expressed during peach fruit development and abscission. *J. Exp. Bot*., *53*: 2333-2339.

RUEPP, A; ZOLLNER, A; MAIER, D; ALBERMANN, K; HANI, J; MOKREJS, M; TETKO, I; GULDENER, U; MANNHAUPT, G; MUNSTERKOTTER, M and MEWES, H W (2004). The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res*., *32*: 5539-5545.

SAMBANTHAMURTHI, R; OO, K C and SITI HASNAH, P (1994). Lipase activity in the oil palm (*Elaeis guineensis*) mesocarp. *Plant Lipid Metabolism* (J C Kader and P Mazliak eds.). Kluwer Academic Publishers.

SCHOOF, H; ZACCARIA, P; GUNDLACH, H; LEMCKE, K; RUDD, S; KOLESOV, G; ARNOLD, R; MEWES, H W and MAYER, K F (2002). MIPS *Arabidopsis thaliana* database (MAtDB): an integrated biological knowledge resource based on the first complete plant genome. *Nucleic Acids Res*., *30*: 91-93.

SITI NOR AKMAR, A; NURNIWALIS, A W and CHEAH, S C (2007). Expression regulatory elements. Malaysian application No. PI 20070394.

SITI NOR AKMAR, A; NORAZNIZAH, A; MAHADZIR, J; CHAN, P L and NURNIWALIS, A W (2003). Characterization of gene expression in oil palm mesocarp *via* expressed sequence tags. *2nd International Conference of Bioinformatics.*

SITI NOR AKMAR, A (1999). *Structure and Regulation of Stearoyl-ACP Desaturase and Metallothionien-like Genes in Developing Fruits of Oil Palm*. Ph. D thesis. University of East Anglia.

STERKY, F; REGAN, S; KARLSSON, J; HERTZBERG, M; ROHDE, A; HOLMBERG, A; AMINI, B; BHALERAO, R; LARSSON, M; VILLARROEL, R; MONTAGU, M V; SANDBERGDAGGER, G; OLSSONPARALLEL, O; TEERI, T; BOERJAN, W; GUSTAFSSON, P., UHLÉN, M; SUNDBERGDAGGER, B and LUNDEBERG, J (1998). Gene discovery in the wood-forming tissues of poplar: analysis of 5; 692 expressed sequence tags. *Proc. Natl. Acad. Sci. USA, 95*:13330-13335.

SUH, M C; KIM, M J; HUR, C G; BAE, J M; PARK, Y I; CHUNG, C H; KANG, C W and OHLROGGE, J B (2003). Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. *Plant Mol. Biol*., *52*: 1107-1123.

TAN, A (2003). *Generation of Expressed Sequence Tags (ESTs) from Non-embryogenic Callus cDNA Libraries of Oil Palm* (*Elaeis guineensis).* M.Sc. thesis. Universiti Putra Malaysia.

VAN DER HOEVEN, R; RONNING, C; MARTIN, G; GIOVANNONI, J and TANKSLEY, S (2002). Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of large EST collection and selective genomic sequencing. *The Plant Cell., 296*: 343-346.

VOETZ, M; KLEIN, B; SCHELL, J and TÖPFER, R (1994). Three different cDNAs encoding acyl carrier proteins from *Cuphea lanceolata*. *Plant Physiol*., *106*: 785–786.

WANG, X (2001). Plant phospholipases. *Annu. Rev. Plant Physiol. Plant Mol. Biol*., *52*: 211–231.

WONG, Y T; KUSHAIRI, A D; MOHAMAD, O and SAMBANTHAMURTHI, R (2005). Assay of lipase activity in *Elaeis guineensis* in various germplasms. *Proc. of the 2005 Conference on Biotechnology of Plantation Commodities*. p. 441-444.

WORTMAN, J R; HAAS, B J; HANNICK, L I; SMITH, J R K; MAITI, R; RONNING, M C; CHAN, A P; YU, C; AYELE, M; WHITELAW, C A; WHITE, O R and TOWN, C D (2003). Annotation of the *Arabidopsis* genome. *Plant Physiol., 132*: 461–468.

ZHU, H; NOWROUSIANM, M; KUPFER, D; COLOT, H V; BERROCAL-TITO, G; LAI, H; BELL-PEDERSEN, D; ROE, B A; LOROS, J J and DUNLAP, J C (2001). Analysis of expressed sequence tags from two starvation, time-of-day-specific libraries of *Neurospora crassa* reveals novel clock-controlled genes. *Genetics, 157*: 1057–1065.