Journal of Oil Palm Research Vol. 27 (1) March 2015 p. 1-11

IDENTIFICATION OF GENES INVOLVED IN SOMATIC EMBRYOGENESIS DEVELOPMENT IN OIL PALM (*Elaeis guineensis* Jacq.) USING cDNA AFLP

THOSSAPOL PATTARAPIMOL*; MYA THUZAR; APICHART VANAVICHIT**; SOMVONG TRAGOONRUNG‡; SITTIRUK ROYTRAKUL‡ and CHATCHAWAN JANTASURIYARAT***

ABSTRACT

*Oil palm (*Elaeis guineensis *Jacq.) belongs to the Arecaceae family and is the largest source of edible vegetable oil worldwide. Micropropagation of oil palm by somatic embryogenesis takes a long time from callus initiation to the production of plantlets. This research project aims to investigate the genes involved in somatic embryogenesis of oil palm during the tissue culture process. We applied the cDNA-AFLP technique to identify genes associated with somatic embryogenesis in oil palm during the tissue culture process. Sixty-four EcoRI/MseI AFLP primer combinations produced a total of 1449 transcription-derived fragments (TDF), of which 1193 TDF (82.33%) showed polymorphism. The most abundant up-regulated and down-regulated transcripts during somatic embryogenesis were selected for characterisation. Sixty-five TDF were sequenced, and 34 distinct sequences were obtained. Primer pairs were successfully designed for 18 of these sequences. The differential expression of the 18 distinct sequences was analysed by using semi-quantitative RT-PCR. The information obtained from this study provides prior knowledge about the genetic underlying the somatic embryogenesis in oil palm, and will be the starting point that reveals the genetic components and mechanism in the control of this process in the future.*

Keywords: somatic embryogenesis, oil palm, tissue culture, bioenergy, gene identification.

Date received: 21 May 2014; **Sent for revision:** 22 July 2014; **Received in final form:** 3 November 2014; **Accepted:** 5 January 2015.

INTRODUCTION

Oil palm (*Elaeis guineensis* Jacq.) belongs to the Arecaceae family and is the largest source of edible vegetable oil worldwide. Oil palm is the most efficient oil-bearing crop with an average annual yield of 4-5 t of crude oil per hectare per year (Corley, 1998). Palm oil is the most competitive oil for food uses, non-food derivatives, oleochemical and in the biofuel industries. During the last decade, the demand for palm oil has increased rapidly, but the world production of palm oil is far from sufficient to meet the demands (Aratrakorn *et al.*, 2006). In order to meet the global demand for palm oil, the oil palm plantation areas in many oil palm producing countries, including Malaysia, Indonesia and Thailand are expanding (Aratrakorn *et al.*, 2006). To comply with the plantation expansion,

* Department of Genetics, Faculty of Science, Kasetsart University, Bangkhen Campus, Bangkok 10900, Thailand. E-mail: fscicwj@ku.ac.th

** Rice Gene Discovery Unit, Kasetsart University, Kamphaeng Sean Campus, Nakhon Pathom 73140, Thailand.

‡ Genome Institute, BIOTEC, Klong Luang, Pathumthani 12120, Thailand.

hundreds of thousands of high-yielding oil palm seeds will need to be produced to meet the increased demand. Unfortunately, the propagation by seeds of the common commercial variety, *tenera*, which is a hybrid between *dura* and *pisifera* types, will result in segregation and will not meet the true-to-type of an elite *tenera* oil palm tree (Ummi *et al*., 2011). Therefore, tissue culture was used to produce true-to-type *tenera* plantlets in large-scale clonal propagation of elite *tenera* (Thuzar *et al*., 2012).

Oil palm tissue culture is usually started from very small sections of young leaf tissue (explants). The callus, formed from the explants, is nodular in appearance and competent to undergo embryogenesis. Embryogenesis is a serial developmental process that would induce somatic embryo formation and maturation, shoot regeneration, rooting and finally the regeneration of new viable plantlets. The regular oil palm tissue culture takes at least one and a half years from the beginning to the generation of plantlets. The callus formation and somatic embryo formation stages are the longest stages in oil palm tissue culture (Konan *et al*., 2010). The callus formation, or the callogenesis stage of oil palm explants, usually takes two to four months, and the somatic embryo formation, or somatic embryogenesis stage, takes another seven to eight months (Thuzar *et al*., 2011). Comparing the amount of time required for oil palm tissue culture to that of other crops, for example, rice (four to five months) (Hiei *et al*., 1994), tobacco (two to three months) (Svab *et al*., 1975), and cotton (10-12 months) (Sun *et al*., 2006), the oil palm tissue culture process takes much longer. Despite the economic importance of oil palm mass propagation and tissue culture, little is known about the biological mechanisms of the changes associated with callogenesis and embryogenesis in oil palm, especially at the gene expression level. Only a few studies have identified genes which are involved in this process. For example, Aquea and Arce-Johnson (2008) used cDNA-AFLP technique to identify genes expressed during early somatic embryogenesis in *Pinus radiate*. Lai and Lin (2013) used Illumina paired-end sequencing to analyse the transcriptome of longan (*Dimocarpus longan*) embryogenic callus. Identification of genes involved in this process, and an understanding of the molecular mechanisms involved in the reproductive biology of oil palm would provide the tools necessary for the establishment of molecular breeding programmes for oil palm improvement, and will finally shorten the length of the oil palm tissue culture process.

The cDNA amplified fragment length polymorphism (cDNA-AFLP) can be used as one of the genome-wide level expression profiling methods capable of finding genes that have differential expression patterns under the studied condition (Aquea and Arce-Johnson, 2008; Yang *et al*., 2012). Sequence information is absolutely not necessary for cDNA-AFLP experimental design; it is extremely useful for designing the experiment in such a way that the expression of as many genes as possible can be measured with reasonable experimental effort (Kivioja *et al*., 2005). In this study, we applied the cDNA-AFLP technique to identify genes that are associated with somatic embryogenesis in oil palm during tissue culture process and validate the expression patterns for some of the candidate genes through semi-quantitative RT-PCR. The information obtained from this study provides the prior knowledge about the genetics underlying the somatic embryogenesis in oil palm, and it will be the starting point to reveal the genetic components and mechanism in the control of this process in the future.

MATERIALS AND METHODS

Plant Materials

Young leaf tissue of 3-year old *tenera* clone from the Golden Tenera Plantation in Krabi, Thailand was used as explants for recloning. The explants were inoculated on N6 medium supplemented with 120 mg litre^{-1} of 2, 4-D and 3 g litre^{-1} of activated charcoal for callus induction. The callus was then transferred to N6 medium supplemented with 0.16 g litre^{-1} of putrescine, 0.5 g litre^{-1} of casein amino acids, and 0.04 g litre^{-1} of adenine sulphate for somatic embryo maturation. Finally, N6 medium supplemented with 1 g litre^{-1} of activated charcoal was used for plant regeneration (Thuzar *et al*., 2011). Tissues at different developmental stages - callus, somatic embryos at globular, torpedo, and cotyledonary stages, and plantlets were collected for RNA extractions (*Figure 1*).

RNA Extraction and cDNA Synthesis

Total RNA was extracted from oil palm callus, somatic embryos at globular, torpedo, and cotyledonary stages, and oil palm plantlet using Spin Plant RNA (Stratec molecular) as per the manufacturer's instructions. Total RNA was quantified by NANO drop and 1 μg of total RNA was used to check RNA quality with 1% agarose gel. Two microgrammes of total RNA from each sample was used for cDNA synthesis following the manufacturer's instructions (Improm-II Reverse Transcription System, Promega).

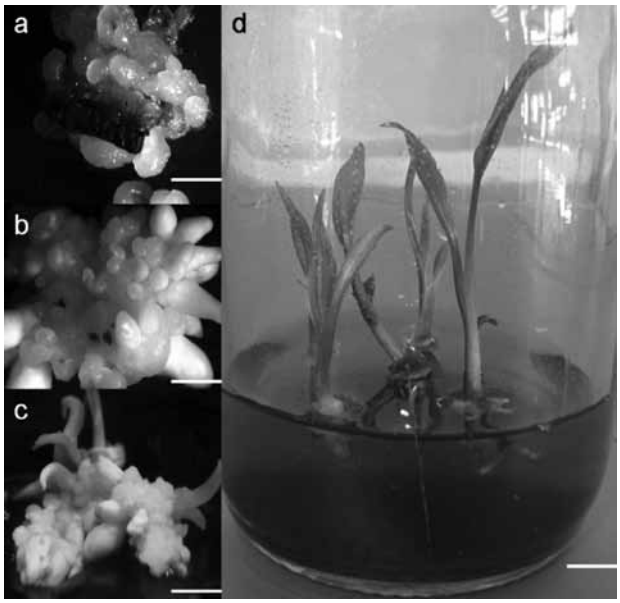


Figure 1. Developmental stages of oil palm somatic embryogenesis from young leaf tissue in modified N6 medium. (a) Callus induction on young leaf tissue (bar 4 mm); (b) differentiation of polyembryoids at three months culture on embryo maturation medium (bar 6 mm) showing globular and torpedo stages; (c) cotyledonary stage somatic embryos (bar 8 mm); (d) regenerated plantlets at 12 months culture (bar 16 mm).

Identification of Differentially Expressed Genes using cDNA AFLP

Fifty nanogrammes of double-stranded cDNA were digested with *Eco*RI and *Mse*I. The digested fragments were ligated with *Eco*RI and *Mse*I adapters by T_4 DNA Ligase at 22°C for 1 hr. The ligated fragments were diluted and pre-amplified with pre-selective AFLP primers. The pre-selective primers were as follows: *Mse*I + N primer, 5´-GATGAGTCCTGAGTAAC-3´ and *Eco*RI + N primer, 5´- GACTGCGTACCAATTCA -3´. The PCR profile for pre-selective PCR was as follows: 25 cycles, consisted of 94°C for 30 s, 56°C for 1 min, and 72°C for 1 min. The fragment products were analysed with 1% agarose gel. The PCR products were diluted 20-fold with sterile water in preparation for the selective amplification reaction. The selective primers for selective amplification were as follows: *Mse*I + CNN primers, 5´-GATGAGTCCTGAGTAACNN-3´ and *Eco*RI + ANN primers, 5´-GACTGCGTACCAATTCANN-3´ in which N could correspond to A, C, G or T. The PCR profile for the selective PCR was as follows: first 12 cycles, the annealing temperature was reduced by 0.7°C per cycle, consisted of 94°C for 30 s, 65°C for 30 s and 72°C for 1 min, and final 25 cycles consisted of 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min. Eight sets of the *Mse*I and *Eco*RI primers were given a total of 64 primer pair combinations. Amplification products were separated on 6% polyacrylamide gels using the Sequigel system (Biorad).

Characterisation of Transcript-derived Fragments (TDF)

TDF corresponding to differentially expressed transcripts were cut out from the polyacrylamide gel by incubation in sterile water at 55°C for 10 min. DNA was eluted and reamplified under the conditions used for selective amplification. The PCR conditions were as follows: 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min for 25 cycles and last extension for 10 min at 72°C. The PCR products were extracted from the gel using purification kit (Favogen), cloned into pGEM-T easy vector (Promega, Wisconsin, USA.) and transformed into *E. coli*. Colony PCR was used to identify the positive clones with DNA insert which checked on 1.5% agarose gel. Then, sequence information was obtained by sequencing individual clones using T7 and/or SP6 at Macrogen, South Korea. The nucleotide sequences were annotated using a BLAST program from the National Centre for Biotechnology Information (NCBI) and non-redundant databases (Altschul *et al.*, 1997). The nucleotide sequences were also classified into their gene function categories.

Validation of TDF using Semi-quantitative RT-PCR

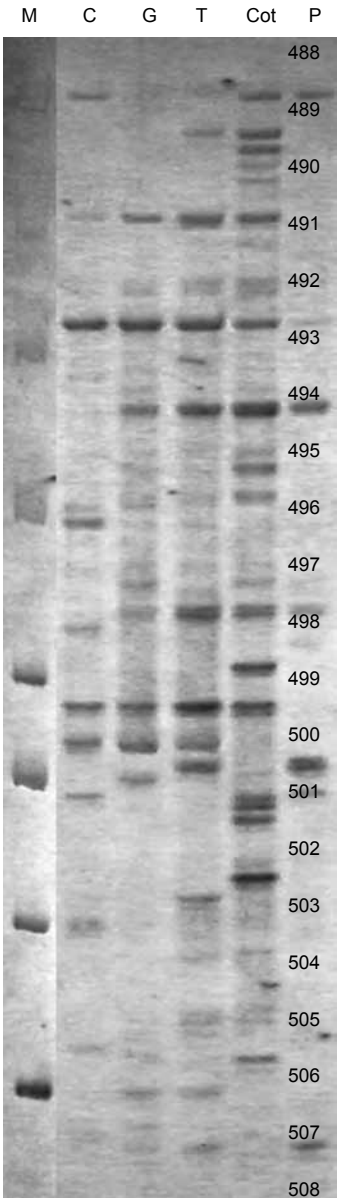
The semi-quantitative RT-PCR was carried out to validate the gene expression. Specific primers for each TDF were designed and used for gene expression confirmation. The cDNA samples from five stages of oil palm somatic embryogenesis were used for amplification with gene specific primers using the same cycling conditions as described previously. The amplified products were analysed with 1.5% agarose gel. Quantification was done using ChemiDoc XRS with Image lab software (Bio-Rad®, USA). Relative expression levels of each gene in each sample were normalised to the expression level of oil palm elongation factor (EF: EU285013.1). Each gene was repeated at least twice with different biological samples for the statistical analysis.

RESULTS

cDNA-AFLP of Oil Palm during Somatic Embryogenesis Development

The cDNA-AFLP analysis was used to identify differentially expressed genes in oil palm during somatic embryogenesis. The expression profile was compared between different developmental stages of somatic embryogenesis consisting of callus (C), globular (G), torpedo (T), cotyledon (Cot), and plantlet (P) (*Figure 1*). Sixty-four *Eco*RI/*Mse*I AFLP primer combinations produced a total of 1449 TDF where 1193 TDF (82.33%) showed polymorphism

(*Figure 2*). Ten primer pairs, namely E-ACA/M-CAC, E-AGC/M-CAC, E-AGT/M-CAC, E-AGG/M-CAC, E-ACC/M-CAG, E-AGT/M-CAG, E-AGG/M-CAG, E-ACA/M-CGC, E-AAG/M-CGG and E-AGG/M-CGG, failed to amplify (*Table 1*). Ten primer pairs, namely E-AGA/M-CAA, E-AGC/M-CGC, E-AGA/M-CGC, E-AGT/M-CGC, E-AGG/M-CGC, E-AGG/M-CGT, E-AAC/M-CGG, E-ACC/M-CGG, E-AGA/M-CGG and E-AAC/M-CGA, gave amplified products which showed 100% polymorphism (*Table 1*). Eighty TDF with the sizes between 100 – 400 bp, which were differentially expressed during somatic embryogenesis, were selected for further analysis. Each TDF was eluted from the gel, and re-amplified with the original primers that had been used for the cDNA-AFLP analysis. After that, the isolated TDF were cloned into pGEM-T easy vector, and eight colonies for each TDF were randomly selected for colony PCR screening and plasmid isolation was performed on those positive clones. Sixty-five out of 80 TDF were successfully amplified, purified and sent for sequencing. These 65 TDF consisted of 17 TDF which were expressed only in the callus stage; five TDF expressed only in the globular stage; two TDF expressed only in the torpedo stage; five TDF expressed only in the cotyledonary stage; 15 TDF expressed only in the plantlet; one TDF expressed in the torpedo and cotyledonary stages; eight TDF expressed in the callus, globular, torpedo, and cotyledonary stages; one TDF expressed in the cotyledonary and plantlet stages; one TDF expressed in the callus and torpedo stages; one TDF expressed in the callus, torpedo, cotyledonary and plantlet stages; one TDF expressed in globular, torpedo and plantlet stages; one TDF expressed in callus, torpedo and cotyledonary stages; one TDF expressed in globular, torpedo, and cotyledonary stages; one TDF expressed in globular and torpedo stages; one TDF expressed in globular, torpedo, cotyledonary and plantlet stages; two TDF expressed in callus, globular, and torpedo stages; one TDF expressed in torpedo, cotyledonary and plantlet stages; and one TDF expressed in callus and globular stages.



Note: M - DNA marker, C - callus, G - globular, T - torpedo, Cot - cotyledonary, P - plantlet.

Figure 2. Expression of transcript associated with somatic embryogenesis development by cDNA-AFLP. An example showing selective amplication with primers EAGA+MCGC.

TABLE 1. NUMBER OF AMPLIFIED TRANSCRIPT-DERIVED FRAGMENTS (TDF) AND DIFFERENTIALLY EXPRESSED TDF DETECTED IN DEVELOPMENTAL STAGE FOR EACH OF THE 64 PRIMER PAIRS

Primer	M-CAA	M-CAC	M-CAG	M-CAT	M-CGC	M-CGT	M-CGG	M-CGA
E-AAC	8[a](6)[b]	6(2)	33(17)	23(15)	28(27)	17(12)	40	40
E-ACC	8(6)	8(7)	ND[c]	36(11)	22(21)	19(17)	43	40(38)
E-AAG	21(16)	6(4)	37(15)	19(10)	10(9)	34(31)	ND	40(39)
E-ACA	9(3)	ND	40(13)	21(16)	ND	19(13)	17(14)	9(6)
E-AGC	13(6)	ND	23(16)	40(39)	16	20(10)	32(27)	42(41)
E-AGA	38	36(34)	38(17)	37(23)	27	9(6)	36	27(24)
E-AGT	44(41)	ND	ND	20(14)	19	13(8)	35(34)	30(29)
E-AGG	49(47)	ND	ND	56(55)	36	45	ND	15(14)

Note: [a] Number of amplified TDF. [b] Number of differentially expressed TDF. [c] ND - no bands detected.

Sequencing and Classification of cDNA Clones

The most abundant up-regulated and down-regulated transcripts during somatic embryogenesis were selected for expression validation. Thirty-four out of 65 TDF were successfully sequenced using the T7 promoter primer. Their derived sequences were used for BLAST search (BLASTN and BLASTX) against NCBI GenBank Databases (*Table 2*). The TDF sequences were also classified into eight functional categories according to gene annotation from the Gene Ontology database. The results showed that eight TDF were classified to a general metabolism category, six TDF were classified to be transposon, four TDF were classified to transcription regulation category, four TDF to cell signalling, three TDF to translation regulation, three TDF to cellular transport, one TDF to mRNA splicing and one TDF to membrane attached protein category respectively. Only four TDF had no similarity (*Figure 3*).

Validation of Differentially Expressed Fragments using Semi-quantitative RT-PCR

Semi-quantitative RT-PCR analysis was used to validate the reliability of cDNA-AFLP results

TABLE 2. SIMILARITY OF TRANSCRIPT-DERIVED FRAGMENTS (TDF) EXPRESSED DURING SOMATIC EMBRYOGENESIS STAGE USING DATABASE FROM GENBANK

TDF#	Length[a] (bp)	Accession number[b]	Homology[C]	BLAST score[d]	Expression pattern[e]
2	250	XM_003567918	Predicted:V-type proton ATPase subunit G-like (*Brachypodium distachyon*)	6e-30	C
6	324	DQ406987	NADH dehydrogenase subunit 5 (nad5), mitochondrial (*Trimenia moorei*)	2e-26	G, T, Cot
8	233	XM_002887309	UDP-glucose: glycoprotein glucosyltransferase (*Arabidopsis lyrata subsp. lyrata*)	4e-26	G, T
16	199	XP_004146504	Predicted: small nuclear ribonucleoprotein Sm D2-like (*Cucumis sativus*)	4e-33	C, T
18	160	ABF74732	Acetyl-CoA carboxylase (*Elaeis guineensis*)	0.30	C
24	402	XM_002270031	Predicted: glyoxysomal fatty acid beta-oxidation multifunctional protein MFP-a-like (*Vitis vinifera*)	1e-16	G, T, P
25	376	AAC79110	Putative polyprotein of LTR transposon (*Arabidopsis thaliana*)	2e-18	G
26	196	XP_002280794	Predicted: Pyrrolidone-carboxylate peptidase isoform 4 (*Vitis vinifera*)	2e-18	C, G, T, Cot
30	209	AM111314*	Myb transcription factor (*Plantago major*)		
39	243	XP_003518260	Predicted: transation initiation factor 4E type 3-like (*Glycine max*)	4e-05	C, G, T, Cot
40	239	XP_003518260	Predicted: transation initiation factor 4E type 3-like (*Glycine max*)	4e-26	C, G, T, Cot
43	187	HQ413157.1*	E2F protein (*Cocos nucifera*)	1e-25	C, G, T, Cot
44	162	XM_003609258*	Receptor like kinase (*Medicago truncatula*)	1e-26	C, G, T, Cot
45	162	XM_003609258*	Receptor like kinase (*Medicago truncatula*)	4e-16	C
47	125	AY014147.1	5.8s Ribosomal RNA (*Laemodonta* sp.)	4e-16	C
48	100	-	No hit found	4e-46	Cot
54	142	EF622021	Cysteine proteinase CPRS1 mRNA, complete cds (*Elaeis guineensis*)	-	T, Cot, P
59	158	-	No hit found	4e-07	T, Cot, P
64	95	DQ002407*	Gypsy retrotransposon grande1, Xilon1 retrotransposon, complete sequence (*Zea mays*)	- 5e-05	Cot P
73	147	-	No hit found	-	C
74	110	AB597036	Eg-ProT1 gene for proline transporter, complete cds (*Elaeis guineensis*)	0.23	T
75	128	AAD24832	Preprotein translocase SECY protein (*Arabidopsis thaliana*)	2e-06	P
76	175	NP_177254	Asparagine-tRNA ligase (*Arabidopsis thaliana*)	1e-09	P
77	116	BAB01972	Copia-like retrotransposable element (*Arabidopsis thaliana*)	3e-05	P
80	159	NP_179454	Geranyl geranyl pyrophosphate synthase 4 (*Arabidopsis thaliana*)	7e-04	P
81	101	AEQ94169	PHD finger transcription factor (*Elaeis guineensis*)	1.00	P
82	220	AFF18862	Peptidase M1 family protein (*Dimocarpus longan*)	7e-35	C, G, T, Cot
84	166	-	No hit found	-	C
85	342	AAD30456	Heat shock protein 90 (*Solanum lycopersicum*)	3e-64	C, G, T, Cot
86	251	AAX96424	Retrotransposon protein, putative, Ty3-gypsy sub-class (*Oryza sativa Japonica* Group)	3e-22	C, G
88	251	AAX96424	Retrotransposon protein, putative, Ty3-gypsy sub-class (*Oryza sativa Japonica* Group)	3e-22	Cot, P
92	248	NP_194350	1-aminocyclopropane-1-carboxylate synthase 7 (*Arabidopsis thaliana*)	1e-04	C
97	255	AAX96424	Retrotransposon protein, putative, Ty3-gypsy sub-class (*Oryza sativa Japonica*)	3e-22	Cot, P
100	136	ABB91774	Methionine aminopeptidase 1 (*Ananas comosus*)	1e-11	C

Note: [a] Length of the TDF. [b] Genebank accession number. [c] cDNA-AFLP fragment generated by 64 primer combinations showing sequence similarity to NCBI Database sequence using BLASTX and BLASTN. All are BLASTX scores except those marked with * which are BLASTN scores. [d] The E-value was used to indicate the significance of sequence similarity. [e] Expression pattern of somatic embryogenesis development stage. #Indicates TDF selected for semi-quantitative RT-PCR.

for detecting differentially expressed genes. From 65 purified TDF which were sequenced, 34 distinct sequences were obtained, and primer pairs were

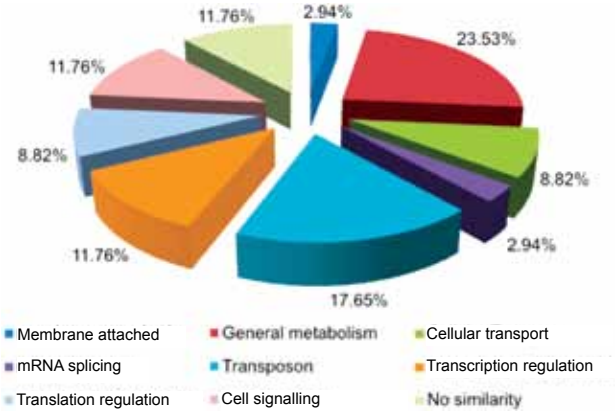


Figure 3. Classification of differentially expressed genes associated with somatic embryogenesis. A total of 34 unique cDNA-AFLP fragments were grouped into eight functional categories. The percentages of gene transcripts in each group are listed.

successfully designed for 18 of these sequences. The differential expression of the 18 distinct sequences was analysed by using semi-quantitative RT-PCR. Seven primer pairs were successfully amplified by the cDNA and confirmed the results of differential gene expression during somatic embryogenesis development of oil palm tissue culture (*Table 3*). The other 11 primer pairs failed to amplify or amplify more than one band, or amplified but did not show polymorphism.

TDF#100, which was highly expressed in callus, showed the most dramatic differential gene expression between callus stage and other stages (*Figure 4g*). It was detected only in the callus and globular stages. The BLAST results revealed that the sequence of TDF#100 encodes a methionine aminopeptidase 1. TDF#77 was specifically expressed in plantlet but not in other stages (*Figure 4c*). The BLAST result showed that TDF#77 encodes a copia-like retrotransposable element. TDF#59, TDF#88, and TDF#97 showed higher levels of expression in the cotyledonary stage than in other stages (*Figures 4b, 4e* and *4f*). TDF#59 did not have significant homology to any known gene. The BLAST result of TDF#88 and TDF#97 showed that they encode for retrotransposon, Ty3-gypsy sub-class. TDF#48 showed high level of expression in torpedo, cotyledonary, and plantlet stages (*Figure 4a*). The sequence of TDF#48 did not have significant homology to any known gene. TDF#85 showed expression in callus, globular, torpedo, and cotyledonaly stages but not in plantlet. The BLAST results of TDF#85 showed that it encodes for heat shock protein 90 (HSP90) (*Table 2*).

DISCUSSION

We identify and characterise differentially expressed genes during somatic embryogenesis development in the oil palm tissue culture process using the cDNA-AFLP technique. This technique uses the standard AFLP protocol on a cDNA template (Breyne *et al.*, 2003). We found that cDNA-AFLP was a rapid and useful tool to identify candidate genes associated with somatic embryogenesis development. Using this technique, it was possible to detect an alteration in gene expression during somatic embryogenesis development in oil palm. The cDNA-AFLP can survey transcriptional changes with no prior assumptions about which genes might be induced or repressed, and it also gives a highly reproducible result (Polesani *et al.*, 2008). Several studies have successfully used cDNA-AFLP technique to identify candidate genes in the past. Teulat-Merah *et al.* (2011)

TABLE 3. SPECIFIC PRIMERS FOR EACH TRANSCRIPT-DERIVED FRAGMENTS (TDF), USED FOR SEMI-QUANTITATIVE RT-PCR, AMPLIFICATION PRODUCT LENGTH, AND ANNEALING TEMPERATURE

Primer ID	Amplification product length (bp)	Annealing temperature (°C)	Forward sequence(5′-3′)	Reverse sequence(5′-3′)
TDF-48	100	54.5	GTAGGCGGAGGAAATAGTA	CGCGGCAAAATCTGGATAA
TDF-59	120	57.3	GCATTCTGAAGACAATGGCG	CGTCACTGATTTTCCCCCTA
TDF-73	106	55.0	AGGAAATATCGGATTCAGGA	GTACTTCGCTCGTGTCACT
TDF-74	114	55.2	GGAGTTCAGCTAGTAAATCG	CGGGCTGAAAATGGAGATAA
TDF-75	424	57.3	CCTTGTACCATTGAGCACCT	GCTGGTATCATGCTGGTTCT
TDF-77	105	56.7	TGGACAAGTATTCGCCTCC	CGACCATTGCAACTCGTGA
TDF-84	144	56.7	GCCCATCTATTTGAGAGCG	GGTCTCTTTTGAGCTCCTC
TDF-86	231	56.0	TCTCGTTCTCTCGGACAG	CTCAGCCGATTCATCTCTC
TDF-88	236	58.8	TCGTTCTCTCGGACGAGCA	CATCATGCTCAGCCGATTC
TDF-97	210	57.3	GTAAGTGCATAGTCAGGTGG	GCATTGTCACCAATCTTCTG
TDF-100	130	57.3	GTTCGATGGACAAAGGTTCG	CAGAAAGCAGCATCTTCTCG

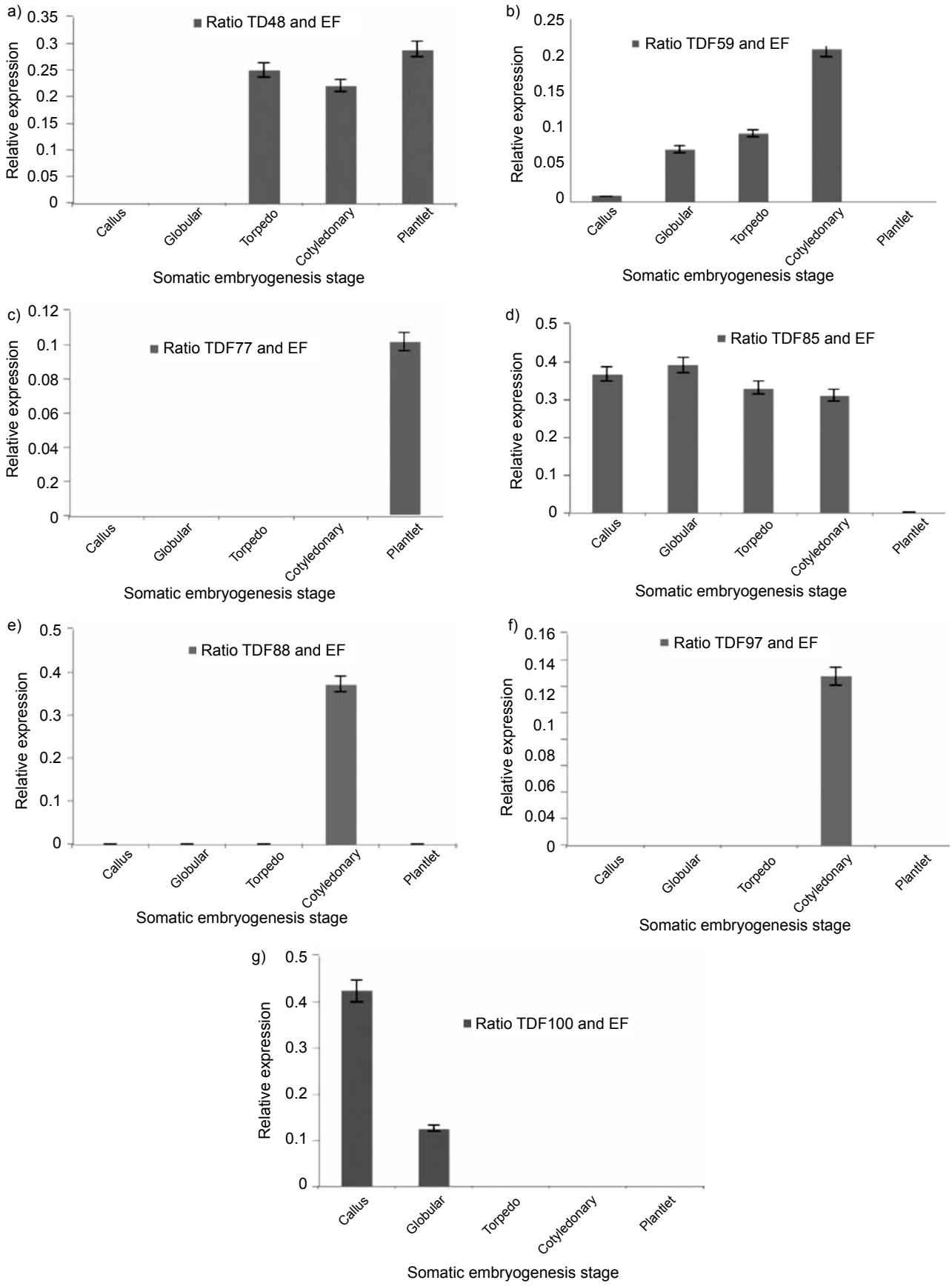


Figure 4. Semi-quantitative analysis of transcript-derived fragments (TDF), (a-g) in somatic embryogenesis development stage expression patterns of (a) unknown gene, (b) unknown gene, (c) copia-like retrotransposable element, (d) heat shock protein 90, (e) retrotransposon protein, putative, Ty3-gypsy sub-class, (f) retrotransposon protein, putative, Ty3-gypsy sub-class, (g) methionine aminopeptidase 1. The relative expression was estimated using the Elongation Factor (EF) gene as constitutive control.

used cDNA-AFLP fingerprinting to identify genes with modulated expression during germination in a common bean (*Phaseolus vulgaris* L.). They identified transcripts encoding proteins belonging to several functional groups, including transcription factors, proteins involved in storage compound hydrolysis, cell elongation and oxidative stress protection. Leymarie *et al.* (2007) used cDNA-AFLP to identify transcripts differentially expressed in dormant and non-dormant embryos. They identified genes belonging to several functional groups, including maintenance of dormancy in barley and probably in other cereals, and signaling elements. Yao *et al.* (2007) used cDNA-AFLP, coupled with bulked segregant analysis (BSA), to screen for differential gene expression between low and high acid content in apple fruits from 'Toko'x'Fuji' (*Malus* x *domestica* Borkn*)* hybrids. They showed that the identified *Mal-DDNA* gene was correlated with apple fruit acidity in malic acid metabolism. Fusco *et al.* (2005) used cDNA-AFLP to identify genes that exhibited a modulated expression following cadmium (Cd) treatment in *Brassica juncea* grown in hydroponic culture. They identified transcripts encoding proteins belonging to several functional groups, including genes involved in cellular metabolism and organisation process and genes involved in the photosynthetic process, suggesting that a multitude of processes are implicated in Cd stress response. Maldonado-Borges *et al*. (2013) successfully used cDNA-AFLP to identify expressed genes during early and late zygotic and somatic embryogenesis of *Musa acuminate* ssp. *malaccensis*, and they helped annotate these genes in the *Musa* genome. In this study, through the combined use of 64 combination cDNA-AFLP primers, we could identify more than 1000 TDF which were differentially expressed during somatic embryogenesis in the oil palm tissue culture process.

The stages of embryogenesis are critical for the subsequent development of a whole plant. The embryo is established at this time through sequential events involving coordinated gene expression. We used cDNA-AFLP analysis to compare expression patterns at different developmental stages during the process of somatic embryogenesis. The differentially expressed TDF analysed were thus developmentally, rather than environmentally regulated, identifying putative gene important to embryogenesis. TDF up- and down-regulated during somatic embryogenesis stage were the subject of this study. Additionally, we determined and validated the expression profile of selected genes, which is important in inferring biological function in the future. The results of the differentially regulated TDF reported here have a direct or indirect relationship with somatic embryogenesis development. Out of the identified TDF, we could confirm the expression by RT-PCR analysis of seven transcripts, which may be associated with the somatic embryo development in oil palm.

In tissue culture, somatic embryogenesis is the process whereby cells redifferentiate and give rise to cells that can form somatic embryos. In several plant systems, cell differentiation and cell fate are affected by cell wall components (McCabe *et al*., 1997). In this study, we found that TDF#2, TDF#18, and TDF#44 are homolog to V-type proton ATPase subunit G-like, acetyl-CoA carboxylase, and receptor-like kinase, respectively. ATPase activity was mainly located on the plasma membrane in the early embryogenic cells. With further development, it was also observed to be distributed in endoplasm, nucleus and vacuoles. With the thickening of embryogenic cell wall, ATPase activity was found in the thickened region and the intercellular space (Li *et al.*, 2003). Acetyl-CoA carboxylase is a biotin-dependent enzyme that catalyses the irreversible carboxylation of acetyl-CoA to produce malonyl-CoA. During carrot somatic embryo development, acetyl-CoA carboxylase activity increased, probably reflecting the greater demands for malonyl-CoA by several metabolic pathways. Malonyl-CoA is required for a cuticular wax-like material to accumulate on the surface of the embryos (Wurtele and Nikolau, 1992). In *A. thaliana*, a receptor-like kinase gene is highly expressed during embryogenic cell formation and globular stage, which can be induced to change fate toward embryogenesis by exogenous auxins (Hecht *et al.*, 2001). The embryogenesis formation in culture depends on the presence of certain arabinogalactan proteins produced by non-embryogenic cells in culture (McCabe *et al.*, 1997). Therefore, some form of signalling between embryogenic and non-embryogenic cells appears to be required for embryo initiation. The identification of these genes supports the idea of embryogenic cells that may be important for callus development (Hecht *et al.*, 2001).

The TDF#85 encodes for a Heat shock protein 90 (HSP90). HSP90 is specifically associated with diverse proteins in many biological processes (Ziemiecki *et al.*, 1986). The selective transcription of Heat shock protein (HSP) involves the interaction between a *tran*-acting 'heat shock transcription factor' and some conserved *cis*-acting elements, including the 'heat shock consensus element', which often occur as overlapping elements, within 100 bp 5' to the transcription initiation site. HSP have been found to be expressed during somatic embryo development in response to hormones such as 2, 4-D (Kitamiya *et al.*, 2000).

The TDF#100 showed similarity with methionine aminopeptidase 1 (MetAP-1) of *A. comosus*. Processing of N-terminal methionines is an essential post-translational process. This process is catalysed by a highly conserved family of enzymes known as methionine aminopeptidases (MetAP) (Bradshaw *et al.*, 1998). MetAP-1 or MetAP-2 function is essential

for embryonic development and survival at the initiation of gastrulation (Yeh *et al.*, 2006).

The TDF#77 showed similarity with copia-like retrotransposable element of *A. thaliana*. The TDF#88 and TDF#97 showed similarity with Ty3-gypsy retrotransposon protein, of *O. sativa*. Retrotransposons are genetic elements that can amplify themselves in genome and ubiquitous components of the plant genome (Sanmiguel and Bennetzen, 1998).

We identified TDF#48 which is highly expressed in torpedo and cotyledonary stages and in plantlet, but it did not have significant homology to any known gene. We also identified TDF#59, which is also highly expressed in cotyledonary stage; it also did not have significant homology to any known gene. However, these genes displayed high expression levels during the embryogenic process; they may play an important role in somatic embryogenesis in oil palm.

In conclusion, the cDNA-AFLP technique was useful to generate information on differences in the gene expression levels during the somatic embryogenesis pathway in oil palm. This information provides a significant contribution to the underlying process of oil palm somatic embryogenesis at the molecular level. We have identified a number of TDF with significant expression levels in different stages; these could encode an interesting pool of candidate proteins involved in regulation of embryogenesis. Identification of these genes provides prior knowledge about the underlying genetics during somatic embryogenesis in oil palm, and will be the starting point to reveal the genetic components and mechanism in the control of this process. Further research on these candidate genes could determine their roles in somatic embryogenesis and in shortening the length of the oil palm tissue culture process. This knowledge may aid in cloning of oil palm commercially. It will also shed light on the way to establish molecular breeding programmes for this economically very important oil bearing crop in the future.

ACKNOWLEDGEMENT

We thank Mr Robert Harman II for critically and grammatically reviewing the manuscript. This research was supported by grants from the Kasetsart University Research and Development Institute (KURDI) and Graduate School, Kasetsart University.

REFERENCES

ALTSCHUL, S F; MADDEN, T L; SCHAFFER, 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 Research, 25*: 3389-3402. http://nar.oxfordjournals.org/content/25/17/3389.full

AQUER, F and ARCE-JOHNSON, P (2008). Identification of genes expressed during early somatic embryogenesis in *Pinus radiata*. *Plant Phyiology and Biochemistry*, *46*: 559-568. http://www.ncbi.nlm.nih.gov/pubmed/18406157.

ARATRAKORN, S; THUNHIKORN, S and DONALD, P F (2006). Changes in bird communities following conversion of lowland forest to oil palm and rubber plantations in southern Thailand. *Bird Conservation International, 16*: 71-82. http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=433868&fileId=S0959270906000062.

BRADSHAW, R A; BRICKEY, W W and WALKER, K W (1998). N-terminal processing: the methionine aminopeptidase and Nα-acetyl transferase families. *Trends in Biochemical Sciences, 23*: 263-267. http://www.ncbi.nlm.nih.gov/pubmed/9697417.

BREYNE, P; DREESEN, R; CANNOOT, B; ROMBAUT, D; VANDEPOELE, K; ROMBAUTS, S; VANDERHAEGHEN, R; INZE, D and ZABEAN, M (2003). Quantitative cDNA-AFLP analysis for genome-wide expression studies. *Molecular Genetics and Genomics, 269*: 173-179. http://www.ncbi.nlm.nih.gov/pubmed/12756529.

CORLEY, R (1998). What is the upper limit to oil extraction ratio? *Proc. of the International Conference on Oil and Kernel Production in Oil Palm - A Global Perspective* (N Rajanaidu; I E Henson and Jalani, B S eds.). PORIM, Bangi. p. 256-269.

FUSCO, N; MICHELETTO, L; DAL CORSO, G; BORGATO, L and FURINI, A (2005). Identification of cadmium-regulated genes by cDNA-AFLP in the heavy metal accumulator *Brassica juncea* L. *J. Experimental Botany, 56*: 3017-3027. http://www.ncbi.nlm.nih.gov/pubmed/16216843.

HECHT, V; VIELLE-CALZADA, J P; HARTOG, M V; SCHMIDT, E D L; BOUTILIER, K; GROSSNIKLAUS, U and DE VRIES, S C (2001). The *Arabidopsis* somatic embryogenesis receptor kinase 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiology, 127*: 803-816. http://www.ncbi.nlm.nih.gov/pubmed/11706164

HIEI, Y; OHTA, S; KOMARI, T and KUMASHIRO, T (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant*

Journal, *6*: 271-282. http://www.ncbi.nlm.nih.gov/pubmed/7920717

KITAMIYA, E; SUZUKI, S; SANO, T and NAGATA, T (2000). Isolation of two genes that were induced upon the initiation of somatic embryogenesis on carrot hypocotyls by high concentrations of 2,4-D. *Plant Cell Rep*, *19*: 551-557. http://link.springer.com/article/10.1007%2Fs002990050772

KIVIOJA, T; ARVAS, M; SALOHEIMO, M; PENTTILA, M and UKKONEN, E (2005). Optimization of cDNA-AFLP experiments using genomic sequence data. *Bioinformatics, 21*: 2573-2579. http://www.ncbi.nlm.nih.gov/pubmed/15774551

KONAN, K E; DURAND-GASSELIN, T; KOUADIO, Y J; FLORI, A; RIVAL, A; DUVAL, Y and PANNETIER, C (2010). *In vitro* conservation of oil palm somatic embryos for 20 years on a hormone-free culture medium: characteristics of the embryogenic cultures, derived plantlets and adult palms. *Plant Cell Reports, 29*: 1-13. http://www.ncbi.nlm.nih.gov/pubmed/19898823

LEYMARIE, J; BRUNEAUX, E; GIBOT-LECLERC, S and CORBINEAU, F (2007). Identification of transcripts potentially involved in barley seed germination and dormancy using cDNA-AFLP. *J. Experimental Botany, 58*: 425-437. http://www.ncbi.nlm.nih.gov/pubmed/17175551

LI, S; XING, G M; CUI, K R; YU, C H; ZHANG, X; XU, H X and WANG, Y F (2003). Ultracytochemical localization of calcium and ATPase activity on the 2,4-D induced somatic embryogenesis of *Lycium barbarum* L. *Shi Yan Sheng Wu Xue Bao*, *36*: 414-420. http://www.ncbi.nlm.nih.gov/pubmed/14724931

LAI, Z and LIN Y (2013). Analysis of the global transcriptome of longan (*Dimocarpus longan* Lour.) embryogenic callus using Illumina paired-end sequencing. *BMC Genomics*, *14*: 561. http://www.biomedcentral.com/1471-2164/14/561

MALDONADO-BORGES, J I; KU-CAUICH, J R and ESCOBEDO-GRACIAMEDRANO R M (2013). Annotation of differentially expressed genes in the somatic embryogenesis of *Musa* and their location in the banana genome. *The Scientific World Journal*. Article ID 535737, Doi:10.1155/2013/535737. http://www.hindawi.com/journals/tswj/2013/535737/

MCCABE, P F; VALENTINE, T A; FORSBERG, L S and PENNELL, R I (1997). Soluble signals from cells identified at the cell wall establish a developmental pathway in carrot. *The Plant Cell Online, 9*: 2225-2241. http://www.plantcell.org/content/9/12/2225.abstract

POLESANI, M; DESARIO, F; FERRARINI, A; ZAMBONI, A; PEZZOTTI, M; KORTEKAMP, A and POLVERARI, A (2008). cDNA-AFLP analysis of plant and pathogen genes expressed in grapevine infected with *Plasmopara viticola*. *BMC Genomics, 9*: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2292706/

SANMIGUEL, P and BENNETZEN, J L (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany, 82*: 37-44. http://aob.oxfordjournals.org/content/82/suppl_1/37.abstract

SUN, Y; ZHANG, X; HUANG, C; GUO, X and NIE, Y (2006). Somatic embryogenesis and plant regeneration from different wild diploid cotton (*Gossypium*) species. *Plant Cell Rep, 35*: 289-296. http://www.ncbi.nlm.nih.gov/pubmed/16315034

SVAB, Z; HAJDUKIEWICZ, P and MALIGA, P (1975). Transgenic tobacco plants by co-cultivation of leaf disks with pPZP Agrobacterium binary vectors. *Methods in Plant Molecular Biology-A Laboratory Manual* (Maliga, P; Klessig, D; Cashmore, A; Gruissem, W and Varner, J eds.). Cold Spring Harbor Press. p. 55-77.

TEULAT-MERAH, B; MORERE-LE PAVEN, M C; RICOULT, C; AUBRY, C and PELTIER, D (2011). cDNA-AFLP profiling in the embryo axes during common bean germination. *Biologia Plantarum, 55*: 437-447. http://link.springer.com/article/10.1007%2Fs10535-011-0108-5

THUZAR, M; VANAVICHIT, A; TRAGOONRUNG, S and JANTASURIYARAT, C (2011). Efficient and rapid plant regeneration of oil palm zygotic embryos cv. 'Tenera' through somatic embryogenesis. *Acta Physiol Plant, 33*: 123-128. http://link.springer.com/article/10.1007%2Fs11738-010-0526-6

THUZAR, M; VANAVICHIT, A; TRAGOONRUNG, S and JANTASURIYARAT, C (2012). Recloning of regenerated plantlets from elite oil palm (*Elaeis guineensis* Jacq.) cv. Tenera. *African Journal of Biotechnology, 11*: 14761-14770. http://www.academicjournals.org/journal/AJB/article-abstract/349957627499

UMMI, K H; NUR SHIELA, M N; TENGKU NUR DIYANA, T M and GUNASELAN, S (2011). *In vitro* germination in encapsulated and non-encapsulated *Elaeis guineensis* Jacquin (1763) embryo. *Empowering Science, 121*: 764-769.

WURTELE, E S and NIKOLAU, B J (1992). Differential accumulation of biotin enzymes during

carrot somatic embryogenesis. *Plant Physiology, 99*: 1699-1703. http://www.ncbi.nlm.nih.gov/pubmed/16669096

YANG, D; MA, P; LIANG, X; LIANG, Z; ZHANG, M; SHEN, S; LIU, H and LIU, Y (2012). Metabolic profiles and cDNA-AFLP analysis of *Salvia miltiorrhiza* and *Salvia castanea* Diel f. tomentosa Stib. *PLoS ONE, 7*: e29678. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0029678

YAO, Y X; LI, M; LIU, Z; HAO, Y J and ZHAI, H (2007). A novel gene, screened by cDNA-AFLP approach, contributes to lowering the acidity of fruit in apple. *Plant Physiology and Biochemistry, 45*: 139-145. http://www.ncbi.nlm.nih.gov/pubmed/17344054

YEH, J R J; JU, R; BRDLIK, C M; ZHANG, W; ZHANG, Y; MATYSKIELA, M E; SHOTWELL, J D and CREWS, C M (2006). Targeted gene disruption of methionine aminopeptidase 2 results in an embryonic gastrulation defect and endothelial cell growth arrest. *Proc. of the National Academy of Sciences, 103:* 10379-10384. http://www.ncbi.nlm.nih.gov/pubmed/16790550

ZIEMIECKI, A; CATELLI, M G; JOAB, I and MONCHARMONT, B (1986). Association of the heat shock protein HSP90 with steroid hormone receptors and tyrosine kinase oncogene products. *Biochemical and Biophysical Research Communications, 138*: 1298-1307. http://www.sciencedirect.com/science/article/pii/S0006291X86804247