Journal of Oil Palm Research Vol. 22 December 2010 p. 856-868

CLONING AND *in silico* ANALYSIS OF PROMOTERS OF HIGHLY EXPRESSED GENES IN OIL PALM EMBRYOGENIC CULTURES

OOI, S E*; FESHAH ISHAK* and MEILINA ONG ABDULLAH*

ABSTRACT

Promoters are the regulatory regions of DNA that control the expression of their genes. This study aimed at identifying the 5' upstream regulatory regions of four embryogenically up-regulated genes that may elucidate clues to their regulation. Isolation and sequence analyses of the four putative promoters revealed similarities in the predicted motifs. The putative promoters of OPSC10 and EgHOX1 shared more motifs in common compared to the other two promoters, accounting for similarities in the two former expression profiles. The highly represented motifs in the putative promoters suggested regulation by light, phytohormones and Myb binding proteins. The upstream region of EgPER1 was the most unique of the four, with the possibility of being part of a dicistronic operon.

Keywords: promoter, embryogenic, EgPK1, EgHOX1, EgPER1.

Date received: 10 March 2010; **Sent for revision:** 26 April 2010; **Received in final form:** 8 July 2010; **Accepted:** 6 September 2010.

INTRODUCTION

As different genes are required for the completion of different phases in the development of a plant, promoters are important regulatory parts of DNA that govern the expression of their genes. Whilst some cases of regulation occur post-transcriptionally or post-translationally, most cases of regulation occur at the level of transcription. Promoters are located upstream from the transcription site of the gene of interest. The promoter can roughly be divided into two parts, the distal and core regions (Rombauts *et al.*, 2003). The core region is believed to be responsible for the correct assembling of the RNA polymerase II complex at the right position (Nikolov and Burley, 1997; Berk, 1999). The distal part is believed to contain the elements that regulate the spatio-temporal expression (Tjian and Maniatis, 1994; Fessele *et al.*, 2002). How far upstream or downstream the distal region extends is not defined (Rombauts *et al.*, 2003). Recently, Abeel *et al.* (2009) proposed a technique for identifying and delineating (core) promoters that is based on the properties of long stretches of DNA. The Easy Promoter Prediction Programme (EP3) identifies and delineates promoter regions based on the GC content and large-scale structural features of the DNA.

Analysis of the regulatory regions of genes which showed similar patterns of transcription revealed the presence of short DNA sequences which were held in common by genes with a particular pattern of regulation (review by Latchman, 1997). For example, genes that are transcriptionally induced by elevated temperatures contain a common regulatory element known as the heat shock element or HSE (Davidson *et al.*, 1983). The proof that such sequences are of critical importance in producing the pattern of gene transcription was demonstrated by Pelham (1982), who transferred the HSE from the *hsp70* gene onto the non-heat inducible thymidine kinase gene (Latchman, 1997).

These regulatory elements, known as *cis*-acting elements or CAREs, probably serve as binding sites for transcription factors and, hence, are useful for

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

the characterization of their cognate factors (Katagiri and Chua, 1992). Many enhancers and promoters providing tissue-specific (van der Geest and Hall, 1996; Hong *et al*., 1997), environmental (Fordham-Skelton *et al*., 1997), developmental (Beaudoin and Rothstein, 1997) and hormonal (Aoyama and Chua, 1997; Rouster *et al*., 1997) regulation have been identified, and more are being discovered.

Many of the *cis*-acting elements are recognized by *trans*-acting factors, *e.g.* transcription factors. However, the individual binding of a transcription factor to a regulatory element is insufficient to confer context-specific expression. Thus, the combination and orientation of the transcription factors make up the crucial information, rather than the mere occurrence of several binding sites (Scherf *et al*., 2000). In addition, co-regulated genes may be regulated by common *cis*-acting elements. These elements will be useful for tailoring transgene expression to specific needs (Gallie, 1998).

Briefly, this study aims to identify the motifs or *cis*-acting regulatory elements that may elucidate clues to the regulation and environmental influences of embryogenically up-regulated genes isolated from previous studies, *i.e.* EgHOX1, EgPK1, EgPER1 and OPSC10 (Ong, 2001; See, 2002; Ooi, 2003; Ong-Abdullah and Ooi, 2006; 2007; Ooi *et al*., 2008). Based on their spatial and temporal expression patterns, these promoters may be used in the future for the expression of desirable genes in specific tissues (*e.g*. embryogenic tissues).

MATERIALS AND METHODS

Expression Analysis

Expression studies were conducted to analyze the differential expression profiles exhibited by the four genes of interest using real time quantitative PCR assays. Primers and TaqMan probe sets were designed and synthesized by Assay-By-Design® service (Applied Biosystems) for GAPDH, EgPER1, EgHOX1 and OPSC10, while primers and FAM-labelled fluorogenic probes for EgPK1 were designed and synthesized by Bioneer, Korea (Ooi *et al*., 2008). The primers and probe sequences were designed to the 3′UTR regions of EgPER1 and EgHOX1 (3′UTR), and the 5′ region in the ORFs of EgPK1 and OPSC10, respectively (*Table 1*).

Total RNA was extracted from oil palm tissues using the method by Schultz *et al*. (1994). Contaminating DNA was removed from RNA samples by RNAse-free DNAse1 (Roche) treatment. Subsequently, the purified RNA was analyzed for its integrity and purity by using the Agilent 2100 BioAnalyzer according to the manufacturer's instructions. RNA (2 µg) with a 28S:18S ratio of more than 1 was selected for cDNA synthesis using the High Capacity cDNA Archive kit (Applied Biosystems, USA) according to the manufacturer's instructions. Five microlitres of first strand cDNA (1:150 dilution with 0.1mM EDTA) were used with all primer/probe sets for real-time PCR. PCR

TABLE 1. PRIMERS AND FAM-LABELLED PROBE SEQUENCES FOR EgPER1, EgPK1, EgHOX1 AND OPSC10

Clone ID/gene	Sequence (5′ – 3′)
EgPER1 (forward primer)	CTCCGCTTCACAAAAGTCTAATGTT
EgPER1 (reverse primer)	GGGCCCCAGAAACCACTTATG
EgPER1 (FAM-probe)	CCGTCCGTGATATGTT
EgPK1 (forward primer)	GGAGAGGGAGAGAGATAGAGAGAG
EgPK1 (reverse primer)	CCGTTCACCGCGTTGTAGAG
EgPK1 (FAM-probe)	TGAGCCCTGAGAAGGAGCATCCCATC
EgHOX1 (forward primer)	ACCTCTAGCTTAGATTTCATATATTGATCCCA
EgHOX1 (reverse primer)	CTCCAGCTTCTCTTTTGACCCTATT
EgHOX1 (FAM-probe)	CTTGGACCACCATTCATC
OPSC10 (forward primer)	GCACTACGAGGAGTATTTGAAGCAA
OPSC10 (reverse primer)	TCCGACGACTGGGTCTGA
OPSC10 (FAM-probe)	CCATGGCCAACTCC
GAPDH (forward primer)	ACTGCTACTCAGAAGACTGTTGATG
GAPDH (reverse primer)	TGCTGCTAGGAATGATGTTAAAGCT
GAPDH (FAM-probe)	ACCCCTCCAGTCCTTG

reactions of 25 µl were carried out according to the manufacturer's instructions. Real-time PCR was performed on the ABI Prism 7000 Sequence Detection System (Applied Biosystems, USA) in a 96-well reaction plate according to the manufacturer's recommendations for the PCR programme of 2 min at 50ºC; 10 min at 95ºC; and 40 cycles of 95ºC for 15 s and 60ºC for 1 min. Each PCR reaction was performed in triplicate and no-template controls were also included. The quantification of the relative transcript levels was conducted using the Comparative C_t method (Livak and Schmittgen, 2001). The transcript levels of the target genes were normalized against GAPDH gene levels as the endogenous reference, as described in the *Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR* (P/No. 4371095, Applied Biosystems).

Isolation of 5′-Upstream Genomic Sequences

DNA extraction. DNA was extracted from *Elaeis guineensis dura* x *pisifera* young leaf tissues according to the method described by Dellaporta *et al*. (1983). The DNA was then quantified by a spectrophotometer and checked for integrity by agarose gel electrophoresis.

Database mining. Sequences were mined from the in-house oil palm functional gene database (Malaysian Palm Oil Board). The input sequences used for the mining were the cDNA sequences of the four genes of interest. The 5′-upstream sequences obtained for the four genes ranged from 480 bp to 2 kb. A 2 kb 5′-upstream sequence for EgPER1 was successfully obtained through this method. However, as for the other three genes, genome walking and inverse PCR had to be employed to obtain longer 5′-upstream sequences.

Genome walking. Using primers designed to the 5′-end of the cDNA sequences of the respective genes, genome walking was performed using the DNA Walking SpeedUp kit (Seegene, Korea) according to the manufacturer's instructions and the inverse PCR method.

For the inverse PCR method, 1 µg of genomic DNA was first digested with a variety of restriction enzymes. Selection of the restriction enzymes was based on the absence of the corresponding restriction sites in the cDNA sequence (in the regions located upstream of the priming site).

Re-ligation of the genomic DNA was then carried out at diluted conditions to favour unimolecular ligation. The digested DNA mix of 20 ul was mixed with 1X ligation buffer and 7.5 U T4 DNA ligase (Invitrogen, USA), then topped up with sterile distilled water to 500 ul. The ligation mixture was incubated overnight at 16°C. The DNA was then precipitated with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of absolute ethanol, followed by an overnight incubation at -20°C. The DNA was recovered by centrifugation at 12 000 rpm for 20 min and rinsed with 70% ethanol. The air-dried pellet was then resuspended in 10 ul sterile distilled water.

PCR was then carried out with 1 ul of the DNA template from above, 1X PCR buffer, 2 mM $MgCl_2$, 0.25 mM dNTP mix, 0.25 uM of each primer and 0.05 U Taq polymerase (Invitrogen, USA). The PCR programme comprised: 95°C for 3 min; 35 cycles of 95°C for 30 s, 60°C-65°C for 30 s, 72°C for 2 min, and a final extension step at 72°C for 7 min. At times, PCR was also conducted using an AccuPower mix (Bioneer, Korea), for which 1 ul of DNA template and 0.25 uM of each primer were added to the mix and topped up with sterile distilled water to 20 ul. The PCR cycling was carried out as described above. The primers used in amplification of the different regions of interest for each of the three genes of interest are listed in *Table 2*.

Cloning

After the contig assembly, the entire upstream sequences were obtained for the four genes of interest. Isolation of the entire upstream regions for each of the four genes from genomic DNA was conducted by PCR. The primers used are listed in *Table 3*. Cloning of the PCR products into pCR®2.1-TOPO® was then performed using TOPO TA Cloning (Invitrogen, USA). The screened positive recombinant clones were sent for sequencing with universal primers or designed internal sequencing primers. Sequencing was carried out in-house using a BigDye Terminator kit (Applied Biosystems).

Determination of the Transcription Initiation Sites

As the cDNAs were previously isolated from a cDNA library, there was a possibility that the mRNAs were truncated. Hence, to determine the transcription initiation site, the 5′-cap mRNAs had to be isolated as these would contain the complete 5′-region of the particular mRNA of interest. To isolate the 5′-cap mRNA fragments of the corresponding genes, the CapFishing kit (Seegene, Korea) was utilized according to the manufacturer's instructions. The partial cDNAs corresponding to the 5′-cap region of the respective mRNAs were isolated from 1st-strand cDNA of the embryogenic suspension culture for the four genes above using the primers listed in *Table 4*. This allowed us to determine the location of the transcription initiation site or transcription start site (TSS) on the promoter sequence.

TABLE 2. PRIMER SEQUENCES USED IN INVERSE PCR OR DNA WALKING SPEEDUP (Bioneer, Korea) FOR IDENTIFYING 5′ UPSTREAM SEQUENCES of EgHOX1, OPSC10 AND EgPK1

	Primer sequence (5′ – 3′)
1st Inverse PCR - EgHOX1	TAACACGATTGTCTTAGATGAATGGTGGTCC (forward) CTCTCTCTTCAGCTTCCTGTTCTCG (reverse)
2nd Inverse PCR - EgHOX1	TAACACGATTGTCTTAGATGAATGGTGGTCC (forward) AGGATCTCATATCATCTCTTT (reverse)
3rd Inverse PCR - EgHOX1	CTCCCTCGTATTGCATGAA (forward) CATACAAGATTACCGAGTGTG (reverse)
1st Inverse PCR - EgPK1	GTTCATATTTAATAGAGTTCATGCCAT (forward) CAACCCCCTGCCCTACATTA (reverse)
SpeedUp-EgPK1 (1st primer)	CATCCCCAGTTGGAAACAAC (reverse)
SpeedUp-EgPK1 (nested 2nd primer)	CAACCCCCTGCCCTACATTA (reverse)
1st Inverse PCR (OPSC10)	CTCTACGGAGGACAATCCCA (forward) CAGGTAGCAAGATTCAAGTCCA (reverse)
2nd Inverse PCR (OPSC10)	TCCAAGAACTTTGGCATCCT (forward) TGGAGGCTTATTTTGTGGCT (reverse)

TABLE 3. PRIMERS DESIGNED FOR ISOLATION OF UPSTREAM REGIONS OF THE FOUR GENES OF INTEREST FOR CLONING AND SEQUENCING VERIFICATION

Gene	Primer sequence (5′ – 3′)
EgPER1	TCAAGAGCTGGTCCGCTTTCTAC (forward) AGAAGATGATGGCCCAACC (reverse) ACTCGACCGTGTTCTTCCAC (nested forward)
EgPK1	TACCACGTGTCTCGAAGCTG (forward) ATCGGTTCTCGTTGATCCAG (reverse)
EgHOX1	GCTCTCGCAAACCAAACAAT (forward) GTGGGAACAGGATGTGGAAC (reverse)
OPSC10	GATTCGTTGGATGCCAGACT (forward) AAGAAGCCCCATTTCTCACA (reverse)

TABLE 4. PRIMERS DESIGNED FOR ISOLATION OF 5′-CAP mRNA REGIONS OF THE CORRESPONDING GENES

Gene of interest	Primers (5′ – 3′)
EgHOX1	TTTCTCGTGCCGCCTCTGTTGCTGC
EgPK1	AAGGGTGCATGGACTTCTTGGCGGC
EgPER1	AGCGACTCCAACACCCTGACCACCT
OPSC10	AGCAGCAAGATGATCCCGCCGCCAT

Sequence Analysis

Sequence contigs were generated using BioEdit ver. 7.0.5.3 (Hall, 1999). Primers were designed using the Primer3 online tool (Rozen and Skaletsky, 2000). Promoter analysis was conducted with the online tools PLACE (Higo *et al.*, 1999) and MEME (Bailey and Elkan, 1994).

RESULTS AND DISCUSSION

Expression Analysis of the Four Genes of Interest

Real-time quantitative PCR results suggested that all the four genes were highly expressed in embryogenic calli relative to their respective expression levels in non-embryogenic calli,

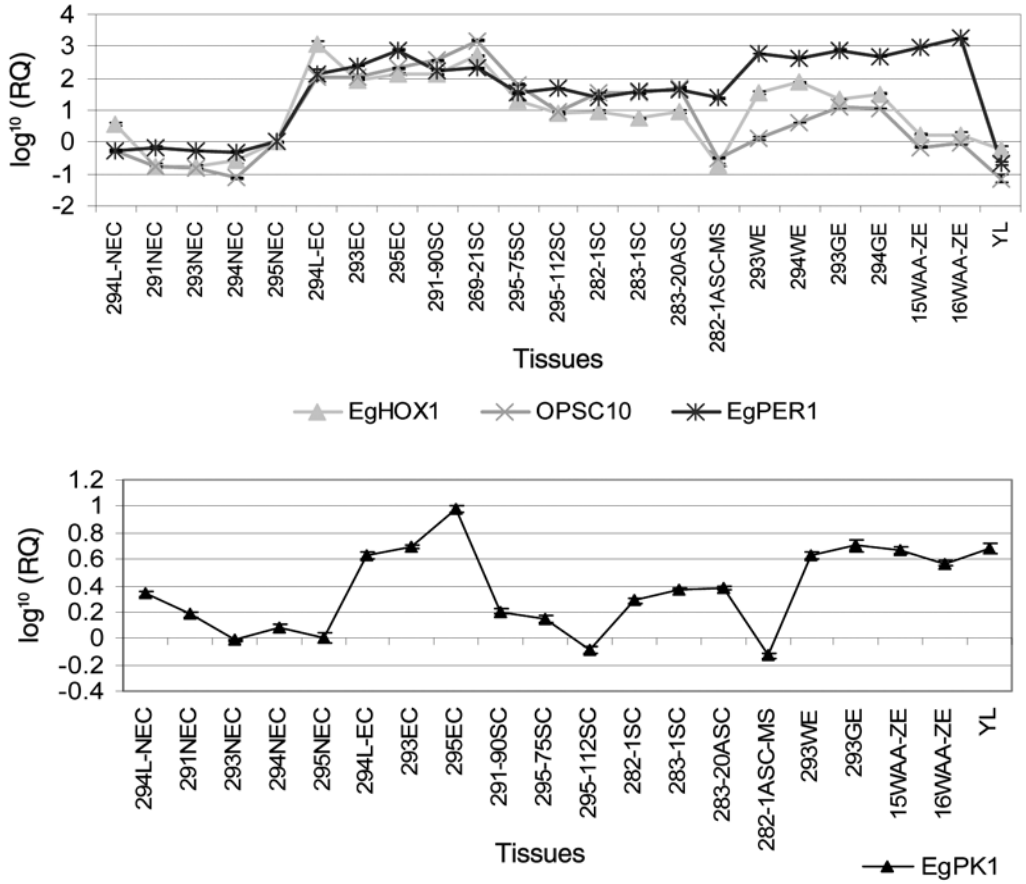

Figure 1. (a) Expression analysis (displayed at log_{10} scale) of EgHOX1, OPSC10 and EgPER1 in various tissues relative to their respective expression values in 295NEC tissue; (b) expression analysis (displayed at log_{10} scale) of EgPK1 relative to its respective expression values in 295NEC tissue (Ooi et al.*, 2008). All the expression levels were previously normalized against the respective GAPDH levels in the tissues.*

Abbreviations: L-NEC = leaf explants detached from non-embryogenic solid cultures; NEC = non-embryogenic calli; L-EC = leaf explants detached from embryogenic solid cultures; EC = embryogenic calli; SC = embryogenic suspension cultures; SC-MS = suspension cultures in MS-basal medium; WE = white embryoids; GE = green embryoids; 15WAA-ZE = zygotic embryos 15 weeks after anthesis (WAA); 16WAA-ZE = zygotic embryos 16WAA; YL = young spear leaves. The numbers preceding the abbreviations are the clone numbers.

regardless of the genotype (*Figure 1*). However, across the different somatic embryo stages, EgPER1 was expressed consistently at high levels in all of the embryo stages tested, while OPSC10 and EgHOX1 were highly expressed only at the early stages, *i.e.* in the embryogenic calli and certain suspension cultures. EgPK1 was highly expressed mainly in the embryogenic calli and maturing embryoid stages (Ooi *et al.*, 2008).

Analysis of the expression profiles of the four genes across the early somatic embryogenesis stages, *i.e.* embryogenic calli till embryogenic suspension cultures, indicated that EgHOX1, OPSC10 and EgPER1 profiles were more similar to each other compared to EgPK1, with the levels in the suspension cultures being close to the levels in the non-embryogenic calli for the latter. However, once embryogenesis was allowed to progress by induction on hormone-free MS medium (SC-MS, WE and GE stages), the EgPER1 profile differed slightly from the profiles of EgHOX1 and OPSC10, particularly in the SC-MS cultures. At this stage of embryo development, EgPK1 appeared to show some similarity in expression profile to EgHOX1 and OPSC10. In zygotic embryos and young leaves, the expression of EgPK1 was the most deviant from the rest, while profiles of EgHOX1 and OPSC10 in these tissues were the most similar.

Although some correlations were seen in the expression profiles of these four genes, it was not confirmed whether these similarities or differences extended further into the vegetative growth or other developmental stages that were not included in this study.

Isolation of Upstream Genomic Regions

EgHOX1. Inverse PCR using the *Cfo*I-digested DNA template successfully generated a 3 kb PCR product. This product was cloned and sequenced to completion using the M13 universal and internal sequencing primers. Sequence analysis verified that this genomic sequence aligned with the partial cDNA sequence of EgHOX1. After locating the *Cfo*I-restriction site in the sequence and thereafter removing the 5′-region located between the forward priming site and *Cfo*I-site, the 5′-upstream region of EgHOX1 obtained from this inverse PCR approach was approximately 300 bp. A second primer set was designed for the next inverse PCR from this 300 bp sequence, and the template DNA used was derived from *Ssp*I-digested DNA. A 1 kb PCR product was generated, cloned and sequenced from this second inverse PCR experiment. Contig analysis was conducted which generated a 950 nt of the 5′-upstream sequence of EgHOX1. Further inverse PCR experiments generated a 640 bp fragment from a recircularized *Nde*I-digested DNA. Sequencing results indicated an extension of 475 nt from the previous 950 nt region.

EgPK1. Mining of the partial genome database extended the 5′-upstream region of EgPK1 by approximately 540 nt. Using this sequence information, primers were designed for inverse PCR. Inverse PCR was carried out using the *Dra*I-digested DNA template. A 700 bp product was obtained, and sequence information verified that it extended the existing 5′-upstream region of EgPK1 by another 600 nt. However, using the DNA Walking SpeedUp kit (Seegene, Korea), a longer fragment was obtained, extending the previous 540 nt 5′-upstream region by another 1000 nt.

OPSC10. Mining of the partial genome database using the existing cDNA sequence information extended the 5′-upstream genomic region of OPSC10 by approximately 580 nt. Inverse PCR with a recircularized *Dra*I-digested DNA template generated a 600 bp product, and subsequent sequence information verified that it extended the existing 5′-upstream region by another 320 nt. Using this region to design the next set of primers, the second inverse PCR generated a fragment of approximately 2 kb. Sequencing results indicated that the existing 5′-upstream genomic region was extended a further 1970 nt.

EgPER1. The 2000 nt long 5′upstream sequence of EgPER1 was solely identified from the partial oil palm genome database. Primers were designed for amplification of the entire length of this upstream region (*Table 3*). The resulting PCR product was cloned and sequenced, and the sequences verified the sequence information obtained from the partial genome database. Further analysis suggested that there was a partial ORF present within this upstream region of EgPER1, which possibly encoded a hypothetical or unknown protein present in other plants like *Vitis vinifera* (Acc. No. CAO40178), *Oryza sativa* (Acc. No. BAC45195) and *Arabidopsis thaliana* (Acc. No. AAM13860). The ORF of the unknown or hypothetical protein spanned from 1 to 1691 nt of the upstream sequence, followed by a non-coding region of 362 nt, and then the coding region of EgPER1.

Isolation of the Entire 5′-Upstream Genomic Regions and Determination of Transcription Initiation Site

With the sequence information obtained from fragments of the 5′-upstream genomic regions of the respective genes, primers were designed to amplify the corresponding whole 5′-upstream regions by PCR. The primer sequences are listed in *Table 4*. Approximately 1.18 kb of the 5′-upstream region of EgHOX1, 2.62 kb of the 5′-upstream region of OPSC10, 2.23 kb of the 5′-upstream region of EgPK1 and 2.03 kb of the 5′-upstream region of EgPER1 were successfully amplified. Each PCR product was cloned into the pCR®2.1-TOPO® vector, and sequenced to completion with the use of internal sequencing primers for verification.

The transcription initiation sites for the corresponding genomic regions were determined by isolating the 5′-cap mRNA fragments of the respective genes. Hence, primers were designed to each of the cDNA sequence of the genes of interest for this purpose (*Table 4*). From the sequencing results, most of the cDNAs isolated previously from the cDNA library were indeed truncated at the 5′-UTR. Thus, with these results, the transcription initiation site could be located on their corresponding genomic sequences, which would also demarcate the boundary of the 5′-UTR (*Figure 2*).

Sequence Analysis

Detection of *cis*-acting elements (CAREs) in the promoter is not self-evident, as such short motifs are statistically expected to occur randomly every few hundred nucleotides (Rombauts *et al.*, 2003). Therefore, the main problem lies in discriminating the 'true' from the 'false' regulatory elements (Blanchette and Sinha, 2001). Successful identification of regulatory DNA patterns therefore depends on the size of the promoter sequence and, to a larger extent, on the quality of the 'related' sequences, which are co-expressed or co-regulated genes (Rombauts *et al.*, 2003).

Due to the various sizes of the upstream regions obtained in this study, sequence analysis was

(a)
```
CCCACAAGCCTACGGACAAATTCAGAACTCAAAGAAAACTACGGAAACTGGAAAAATAACTGAACACTCCATCCATCAA -923
AAGAAGGTGGAGAATTTCCACTACGAAATCCTTAAGCCGTCGAGCTGTAGACTTGGGAACCCCTTCTTTGACAGCTAGT -844
TAGCAGCCTAATTACCTTTCTAAATTGGCAACATTCACTTCCCATAGAATTAATGGAGAAGACTACGTCGGTTGACAAG -765
TGAAGGGCAACTGTCGGATAGAGTACCTTTGTCATGTCATTCCTTCAACATTGAAGTCATGGAAGCTTCCTCTTTCCAT -686
CCAATCATATGGAGCATGACTCTAAAAGACTAAAGTCACTAATTAATTTCCATTTTATTCCCTTTCCAAGTCCTTTCTG -607
TTCTGCCTCAGATCTCCTAGAATCCTTGTAATTAATATAATCCAATTGGCAACCAAACTTTACCCACCATGGTTACACA -528
CAGAATAACAAGCTCCTGCTAAGCTAGATGCAACATGATAAAGAAATCAAACCTTAATGTAGGGCAGGGGGTTGTTTCC -449
AACTGGGGATGCATTTTAGAAATAAAGCAAGCTTGCATAAAGCCTATTGCTCATATTTAATAGAGTTCATGCCATTTTA -370
AAACTAGTGCAGTCTAGTCACTTATTCTATATGTATATATTGATGTGTGTAGATGTAGCACCACTAGTAGTATTATATA -291
TATGCCTCGTAAGTGTGACCATAGACCAGATCCTACTGCTCTAACGAAACATCTTGCACCTTTTATGTATGTATATGTG -212
TGTGTCGATATATATATATATATAATAATTTGATATTGATTAAATAACACGCATTGAAAAGTCAAACCTTGCGGTTGAC -133
TCATCTGTCAAAAAATAATAATAAAAAAATTGTTGTTTTTTCCTCCTCAAATTGTCTATTTTTTAATAAAATTGTAATC -54
                                                                    -1
TCATTATAATTATTCATAAAAATATAAATTATCAAAATTCGAGACTCAGGGCTACGGCGCAGGGAGAGGAAGAGGGGAG +26
AGGGAGAGAGATAGAGAGAGAGAGAGATCTTCTTCTTCTCTTTTGATCGAGGCCGAAGGGGGCAATCGGAGGGAGAA +105
GGGAAGAG
```

(b)
```
ATAAAGGAAGAAAGCTAGTAAGCCACAAAATAAGCCTCCAGAAAATAATGCAAGTGAGGGATTAAATTGTTTCGACTGT -924
CTCAAAAACAATACTGCATATATTTCTATGAAATTCACACATCCATTTAAAGCATGAAAACTCCAATGAGAATTTGCAC -845
AGGAAAAAAAAAAAGATACAAATAGTATATTCCAAGAACTTTGGCATCCTATAAAATGTGTTGCTGCAAGCCCATACACA -766
TGCACTCAGTTGCATCGTCAAAACATCTAAAATGTACATGTTAACATTGTAATAGATCTGAAGGTTTTCATCAACTGTG -687
ATCAGTGTGTATTTGGACAAGTTGGTTAGTTAAAAATAACAAAAATCATCATGGAACCCTAGCAGATTTTCGTTTGCTA -608
GCCTAAGTGATGTTGATTTCAGCAGCTAAATTTAACCAATATGTAAATCTTTTGCGCAGCAATATTGAGGGTGCTAATT -529
TAATAATATATATAAATATACAAATATATATCCTTGGACTTGAATCTTGCTACCTGAATTTTACCTCTCGGGTCGCATT -450
GGTGAACGGAGGACACTCACTACTGTTACTTGCCGTGCCTTTAGCTCTACGGAGGACAATCCCATCTGAACTTGGAGCT -371
CCTTGTCGGAGCTCGAATATTGTGGGTACAATTGTTGATGATGAACTTGGAAGCGCGCCCACCAAAAGGCAAAGGAGGT -292
TCATTGGACTTCTGTTTACTCCAAACCACGGGAAACCAATGGGAGGAGCTACAGTGGATTGTGTTTGGAGGCCAAGTTG -213
TCACTCTAGGGAGAATTTAAAGTTGAAACAGAAGCTTGTCCTTTATTGGGTGCGATCTAATAATAAAAACTATAGCATC -134
CCTCACATGGTGCTGGTGGTGAACTGCATCTACACCAACATGCCATGTTCTATTGTTTGTGTCCTGCTTTCCATGCGGA -55
                                                          -1
CTTGGGGTGCTTGATGGCCACTATAAATGCGGACTGCATGTGGTTTGTCTAATTGTAATGAGCTAGCTTTAGCTAAGCA +25
TATGGAGGAT
```

(c)
```
AAATCATAGAAGGGTTGTTTATCTCACTCTAATCATTCACTTATTCAAAAAAAATAAATCACCTAATATATTAGATTTT -922
TACATGTATATCCTTTCAAATATTTAAATTTACATGAATATCTTCATAAAATTATTATTTAAATATATAATTTTATAAT -843
TTTTTTTCACATTTATCTGTAAAGATATTTTAATCATTTTGATTTTAAATCGTTAATTTTTAACGGTGTTAGATGATAT -764
TGGTATATATATATATATAAAAAGATATTTTATAAGGATATACATGCAGATATCAATTTAGAAAGGATATTTATGTAAA -685
TTTGAATGTTTGAGAAGATATACATATAAAAAATCCTGATATATTTAGATAAATCTGACATCAAGTCTAGCAGCATGCC -606
GTGCTCTGCATTTAAATAAAATTAAATATATTCTCCCGGATGGCACTTAGAATTGTTGGTTCAATTGGATAATTATATA -527
TATTTTGACACCCTATCTCTCCCTCGTATTGCATGAAAAGCCAGGTTTATACAGTTCCTAAGAAATATCATAGAGAAAA -448
TTAGCAGCACTTTTAGCAATAGTATGAAGATCTTAAGAACTTAATGAAATGTTTTATTATATTTCATATGATATTCGTC -369
TAGTATAACTAGAAAGAGACTTCACTAGCAACCATTAAAGTACCTATCAGGGTTATGTCATGAATGCGGATGTTATGAC -290
TACTAATTTCTCACGAGTACACAATTGATTAGGACCCGGAAAGCGCCACGAAATTTTTATGCTGACTTAAAATAGTTGA -211
TAGACTTTGTTAATTTTCCATGGTCCAAAAAGAGATGATATGAGATCCTCTCTCTCTCTCTCTCTCTCTTATTAAAG -132
CTTTGCCTCTACCATTACCTCCACCCATTAACTCACCACCACAAAAAATTTGGAAACTCTCTTCTTCCTACCACCCATG -53
                                                    -1
CACTTGTTCTTCTCTGTATAAAGCCCTCCCACAACCCCCCTTGGCCCTCTTCACCAACAAAGGCCCCCTCTCTGTCTCT +27
CTCATAAAAAGACCCACACATCTAGC
```

Figure 2. The 5'upstream sequences of (a) EgPK1, (b) OPSC10, (c) EgHOX1, (d) EgPER1 of 1 kb in length with the 5'UTR region of the respective genes (in bold). Predicted TATA- and CAAT-boxes are boxed and shaded, respectively. The transcription start sites (TSS) identified by CapFishing (Seegene, Korea) are underlined.

(d)
```
TCCCGGTGGCGGTGAAGATACTCACCCGATTGGACGCCGGCAACAACGCCTTCCCGGCGTGCTCGCCAGACGGGAAGTC -922
GCTGGTCTTTCGGTCGGGGCGGTCCAGCCAGAAGAACCTCTACATCCTCGACGCCGTCAACGGCGAATCCGACGGCGGT -843
GAGGGGATCCGGCAGCTGACTGAAGGGAAGTGGACGGATACGATGCCCACCTGGTCACCGGACGGCGAGCTGATAGCCT -764
TCTCCTCCAACCGCCACGCCCCCTCCAACCCGGACGTCTTCAGCATCTACCTCATCCGCCCGGACGGCACGGGCTTGCG -685
TCGGGTTCACGTGGCTGGGCCCGCGGGGTCCTCCTACGTGGACAGGGAGAGGATCAACCACGTGTGCTTCAGTCCGGAC -606
TCGAAGTGGCTCCTCTTCACGGCCAATCTCGGAAGCGTGACTTCCGAGCCGGTCTCGTGGCCAAACCAATTTCAACCCT -527
ACGGAGACCTGTATGTCTGCCGGCTCGACGGGACTGGACTCAGGAGGCTGACGTGTGGCTCGTACGAGAACGGGACACC -448
GGCGTGGAGCTCGCGTGGTGGGCCCGACGACCTTGGGTCTCTTTCCCTTCGGTCTCCCACCGGAGACAAGCTCCGAGGC -369
CAGTTTGATGAGCCTCTCTGGATAACTTGTGACATATAAGTGATTACTCTTCTCTACCTACTTTGTATCTGGACTGTTC -290
ATCACTGTGCTGCTGTAGCTGATGTTGGAAAGTTGGATTAAAGAAAAGTTATGATGTTGGCATGTTGGAATAAAGAAGA -211
GTTAATGTCCGCCTTATCAAAAAAAAAAAAAAGTTAATTAAGTGTTCCTCCTGCACATCTTAACTGCATGGGACGTGAC -132
ATTGTTGTGAACGATGACACGTGCTTGCAACTTACGGCGAGTGTTGCCACGAGACCACGACGCGGGCCACGTGGCCCAC -53
                                                    -1
ATGCAGCCCCAACCGAGGCCGTTTTAAGCGAGGCCAGGGATGTTGTTCAAGTGGAGTGGGATTCGGTGAATCGGAGCCG +27
TTGAAAG
```

Figure 2. The 5'upstream sequences of (a) EgPK1, (b) OPSC10, (c) EgHOX1, (d) EgPER1 of 1 kb in length with the 5'UTR region of the respective genes (in bold). Predicted TATA- and CAAT-boxes are boxed and shaded, respectively. The transcription start sites (TSS) identified by CapFishing (Seegene, Korea) are underlined (continued).

based on the immediate 1 kb sequence located upstream of the TSS of the respective genes. A systematic deletion analysis of human promoters suggested that a 1 kb promoter can be divided into two regions, the negative regulatory region (-1000 to -500 bp) and the region that positively contributes to the core promoter activity (-300 to -50 bp) (Cooper *et al.*, 2006). Some computational approaches have also limited their analysis to the first 1000 bp upstream of the TSS (Yamamoto *et al.*, 2007; Vandepoele *et al.*, 2009). Motif search to the PLACE database was conducted using the Signal Scan tool (Higo *et al.*, 1999). The motifs were found in one or more sequences, and the number of motifs that were unique or overlapped in two or more sequences are represented in a four-way Venn diagram (*Figure 3*). The motifs present in each of the sequences suggested that the upstream sequences of OPSC10 shared more common motifs

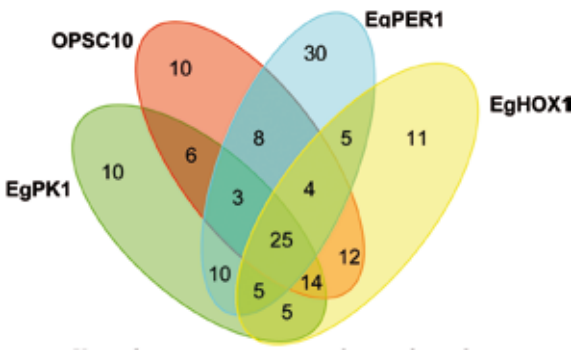

Figure 3. Four-way Venn diagram representing the number of putative motifs present in each of the four upstream sequences (-1 kb) and the number of shared motifs.

with EgHOX1 (12 exclusive motifs) than with the other two sequences. On the other hand, the upstream sequences of EgPER1 and EgPK1 shared 10 motifs in common. Among the four sequences, the EgPER1 upstream sequence appeared to be the most unique, as there were more motifs shared between EgHOX1, EgPK1 and OPSC10. Some of the motifs were present more than once in the sequences. Nevertheless, there were 25 motifs in common among all four sequences.

In the majority of the TATA-containing plant promoters, the TATA box is located 25-30 bp upstream of the TSS (Joshi, 1987; Grace *et al.*, 2004). The TATA boxes were found at the locations at -33, -31 and -36 from the TSS in OPSC10, EgPK1 and EgHOX1, respectively (*Figure 2*). The upstream region of EgPER1 suggested that it is a TATA-less promoter. CAAT-boxes were found at -62 and -82 of EgPK1 and OPSC10 promoters, respectively.

Using Signal Scan, a wide range of motifs was detected in the putative promoter sequences. Some of these motifs are further discussed below (*Table 5)*. One of the motifs shared in common among all four promoters is the BIHD10S, a binding site of OsBIHD1, a rice BELL homeodomain transcription factor. This motif consisting of the sequence TGTCA is a characteristic binding site of some homeodomain transcription factors, including OSH15 (Nagasaki *et al.*, 2001; Luo *et al.*, 2005). The homeodomain transcription factors are a superfamily of genes, regulating various developmental processes in eukaryotes (Gehring *et al.*, 1994). The auxin response element, ARFAT, with a TGTCTC motif, was found in the upstream regions of OPSC10 and EgPER1. It was also found in the 5'UTR region of EgHOX1, suggesting auxin

TABLE 5. LIST OF POSSIBLE MOTIFS PRESENT IN THE 1 kb UPSTREAM REGIONS OF THE TSS OF EgPK1, OPSC10, EgPER1 AND EgHOX1

Name of motif	Motif sequence	Location of the motif upstream of the TSS in the 5′ upstream region			
		EgPK1	OPSC10	EgPER1	EgHOX1
BIHD10S	TGTCA	-127,-730,-735	-214	-339,-135,-117	-313,-521,-629
ARFAT	TGTCTC	-	-926	-385	-
ARR1	NGATT	-53,-175,-479,-564,-582, -679,-892	-161,-229,-530,-550,-579, -611,-687,-781,-937	-255,-328,-581,-856,-974	-264,-633,-653,-795,-803, -811,-928,-944,-970,-999
GATA	GATA	-24,-180,-205,-491,-748	-500,-831	-196,-303,-348,-771,-806, -984	-174,-212,-324,-461,-513, -538,-637,-647,-668,-678, -698,-713,-715,-727,-740, -768,-819,-829,-883,-913, -981
MYB1LEPR	GTTAGTT	-	-662	-	-
MYBATRD22	CTAACCA	-	-664	-	-
MYBGAHV	TAACAAA	-	-650	-	-205
MYBPLANT	MACCWAMC	-556	-	-	-
MYB2CONSENSUSAT	YAACKG	-449,-757,-945	-694,-759	-150,-864	-782
MYBCORE	CNGTTR	-140,-449,-757,-775,-945	-427,-694,-759	-43,-150,-756,-864	-782
MYBST1	GGATA	-749	-500	-349,-807	-539,-699,-728,-913
ABREATCONSENSUS	YACGTGGC	-	-	-64,-66,-677	-
ABRELATERD1	ACGTG	-	-	-63,-64,-113,-114,-138, -476,-625,-626,-649,-676, -677	-
ABREOSRAB21	ACGTSSSC	-	-	-63,-67	-
ABRERATCAL	MACGYGB	-	-	-64,-65,-114,-115,-626, -627,-677	-
ABREZMRAB28	CCACGTGG	-	-	-65	-
ACGTABREMOTIFA2OSEM	ACGTGKC	-	-	-63,-66,-116,-676	-
CATATGGMSAUR	CATATG	-	+24	-	-383
CPBCSPOR	TATTAG	-	-156	-	-932,-938
EVENINGAT	AAAATATCT	-	-	-	-741,-820
LECPLEACS2	TAAAATAT	-	-	-	-739,-818
RYREPEATLEGUMINBOX	CATGCAY	-	-767	-	-56

regulation of these three genes. In addition, the ARR1-binding element was also found in the upstream regions of all four genes, suggesting regulation by the cytokinin-regulated transcription factor, ARR1 (Ross *et al.*, 2004). The GATA motif was also overrepresented in all four putative promoters with as many as 21 motifs present in EgHOX1's putative promoter. GATA-box repeats were found in the promoter of a chlorophyll a/b binding protein, Cab22 gene from petunia (Gidoni *et al.*, 1989). They are also conserved in the promoters of all LHCII type I Cab genes, and are required for high level, light-regulated and tissue-specific expression (Gilmartin *et al.*, 1990; Teakle *et al.*, 2002). The four putative promoters also contained multiple Myb recognition sites.

A number of motifs related to abscisic acid response were overrepresented in the EgPER1 upstream sequence, suggesting that EgPER1 may be regulated by abscisic acid. The upstream region of EgHOX1 contained motifs linked to circadian regulation as it contained two EVENINGAT motifs. It also contained a core element of the *cis*-acting element bound by the tomato cysteine protease (LeCp). LeCp binds on the same *cis*-acting element in *LeAcs2*, an ACC synthase, inducing its expression (Matarasso *et al.*, 2005). The ACC oxidase, which is encoded by OPSC10 (See, 2002), acts downstream of ACC synthase in the ethylene biosynthetic pathway (Yang and Hoffmann, 1984). The presence of this motif in EgHOX1's upstream sequence suggested that EgHOX1 and the oil palm ACC synthase may be induced by the cysteine protease orthologue in oil palm. It is possible then that the ethylene biosynthetic pathway would be up-regulated, leading to higher transcript levels of OPSC10. This may lend support to the observation that both OPSC10 and EgHOX1 displayed similar expression profiles across the tissue culture materials tested. Moreover, the Signal Scan results indicate the putative promoters of EgHOX1 and OPSC10 shared more common motifs compared to the other two promoters, thus supporting the similarities in expression profiles of these two genes.

In the upstream regions of both EgHOX1 and OPSC10, a cytokinin-binding *cis*-acting element,

TABLE 6. SHARED NOVEL PUTATIVE MOTIFS (8 nt or more) IN TWO OR MORE SEQUENCES PREDICTED BY MEME

Predicted motifs by MEME	EgHOX1	EgPER1	EgPK1	OPSC10
	X	X	-	-
	-	-	X	X
	-	-	X	X
	X	-	X	-
	-	X	X	-
	-	-	X	X
	-	X	-	X
	-	-	X	X
	X	X	-	X
	-	-	X	X
	-	X	X	-
	X	-	X	-
	-	-	X	X
	X	-	-	X
	-	X	X	-
	X	-	X	-

Source: Bailey and Elkan (1994).

CPBCSPOR (Fusada *et al.* 2005), was present at two locations in each sequence. In addition, the central element of the gibberellin response complex comprising the TAACAAA box (Gubler *et al.*, 1995) was present as well. They both also contained the legumin box, a *cis*-element found in seed storage protein genes (Fujiwara and Beachy, 1994). The putative promoters of EgHOX1, EgPK1 and OPSC10 also contained a motif found in the NDE element of a SAUR 15A gene promoter from soyabean which is involved in auxin responsiveness (Xu *et al.*, 1997).

Using the MEME tool (Bailey and Elkan, 1994), possible motifs that were present in one or more sequences were predicted (*Table 6*). Some of these motifs consisting of eight nucleotides or more may be novel *cis*-elements, or may just coincidentally be similar sequences shared by the promoters. From the MEME results, the putative promoter of EgPK1 appeared to contain many novel putative motifs in common with the other three putative promoters.

BlastX analysis of EgPER1's upstream sequence resulted in the identification of a possible coding region. This region coded for a truncated fragment of a hypothetical protein found in grapes, rice and *Arabidopsis*. Hence, the region between the coding region of this hypothetical protein and the TSS of EgPER1 was only 329 bp in length. This would mean that the promoter of EgPER1 either lay further upstream of this hypothetical protein, *i.e.* both genes were co-regulated by a single promoter, or the promoter of EgPER1 overlapped with the coding region of this hypothetical protein.

In search for the genome organization of other peroxiredoxins, the Drosophila peroxiredoxin 5, dPrx5, was found to be the second gene in a dicistronic operon (Michalak *et al.*, 2008). Although the dPrx5 coding sequence is located in both dicistronic and monocistronic transcripts, it is translated predominantly from the monocistronic species. The candidate promoter for dPrx-5 spanned the intergenic region between the two ORFs plus a part of the 3'end of the upstream gene. It has not been determined yet whether the promoter for EgPER1 resides in the upstream region of its coding region and possibly overlapping with the coding region of the hypothetical protein, as described in the dicistronic organization of dPrx5. As Michalak *et al.* (2008) suggested that the dicistronic context of the Drosophila dPrx5 appears to be unique, this may be another occurrence of a dicistronic context of a peroxiredoxin in a plant species. Moreover, the transcript size of EgPER1 detected by Northern analysis was approximately 1995 nt (Ong, 2001). This is much larger than the size of the *in silico* analyzed transcript, suggesting that the transcript detected through Northern analysis may be a dicistronic transcript, comprising of the EgPER1 transcript and the transcript of the hypothetical protein.

CONCLUSION

Isolation and analysis of the four putative promoters of embryogenically up-regulated genes from *Elaeis guineensis* revealed that there were similarities in the predicted motifs present in the four putative promoters. The putative promoters of OPSC10 and EgHOX1 shared more motifs in common compared to the other two promoters, which lends support to their similarities in expression profiles. The presence of various predicted *cis*-acting elements in the putative promoters suggests regulation of these genes by light, phytohormones and Myb binding proteins, as these motifs were highly represented in the putative promoters. The upstream region of EgPER1 was found to be the most unique of the four. However, due to the unique nature of EgPER1's upstream region, it is uncertain for now whether this region regulates the expression of EgPER1 until further experiments are conducted.

ACKNOWLEDGEMENT

We thank the Director-General of MPOB for permission to publish this article, and the Genomics Group, MPOB for sequencing services. We are appreciative of Dr Leslie Low Eng Ti for his help in obtaining the upstream sequences from the in-house oil palm functional gene database. We are extremely grateful to our supporting staff for their invaluable technical assistance. This work was financially supported by MPOB's internal funds.

REFERENCES

ABEEL, T; SAEYS, Y; BONNET, E; ROUZE, P and DE PEER, Y V (2009). Generic eukaryotic core promoter prediction using structural features of DNA. *Genome Research*, *18*: 310-323.

AOYAMA, T and CHUA, N-H (1997). A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.*, *11*: 605-612.

BAILEY, T L and ELKAN, C (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. of the Second International Conference on Intelligent Systems for Molecular Biology* (Altman, R; Brutlag, D; Karp, P; Lathrop, R and Searls, D eds.). AAAI Press, Menlo Park, California. p. 28-36.

BEAUDOIN, N and ROTHSTEIN, S J (1997). Developmental regulation of two tomato lipoxygenase promoters in transgenic tobacco and tomato. *Plant Mol. Biol.*, *33*: 835-846.

BERK, A J (1999). Activation of RNA polymerase II transcription. *Curr. Opin. Cell Biol.*, *11*: 330-335.

BLANCHETTE, M and SINHA, S (2001). Separating real motifs from their artifacts. *Bioinformatics*, *17*: 30-38.

COOPER, S J; TRINKLEIN, N D; ANTON, E D; NGUYEN, L and MYERS, R M (2006). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.*, *16*: 1-10.

DAVIDSON, E H; JACOBS, H T and BRITTEN, R J (1983). Very short repeats and coordinate induction of genes. *Nature*, *301*: 468-470.

DELLAPORTA, S L; WOOD, J and HICKS, J B (1983). A plant DNA minipreparation: version II. *Plant Mol. Biol. Rep.*, *1*: 19-21.

FESSELE, S; MAIER, H; ZISCHEK, C; NELSON, P J and WERNER, T (2002). Regulatory context is a crucial part of gene function. *Trends in Genetics*, *18*: 60-63.

FORDHAM-SKELTON, A P; LILLEY, C; URWIN, P E and ROBINSON, N J (1997). GUS expression in *Arabidopsis* directed by 5′ regions of the pea metallothionein-like gene $PsMT_A$. *Plant Mol. Biol.*, *34*: 359-668.

FUJIWARA, T and BEACHY, R N (1994). Tissue-specific and temporal regulation of a beta-conglycinin gene: roles of the RY repeat and other *cis*-acting elements. *Plant Mol. Biol.*, *24*: 261-272.

FUSADA, N; MASUDA, T; KURODA, H; SHIMADA, H; OHTA, H and TAKAMIYA, K (2005). Identification of a novel *cis*-element exhibiting cytokinin-dependent protein binding *in vitro* in the 5′-region of NADPH-protochlorophyllide oxidoreductase gene in cucumber. *Plant Mol. Biol.*, *59*: 631-645.

GALLIE, D R (1998). Controlling gene expression in transgenics. *Curr. Op. Plant Biol.*, *1*: 166-172.

GEHRING, W J; AFFOLTER, M and BURGLIN, T (1994). Homeodomain proteins. *Ann. Rev. Biochem.*, *63*: 487-526.

GIDONI, D; BROSIO, P; BOND-NUTTER, D; BEDBROOK, J and DUNSMUIR, P (1989). Novel *cis*-acting elements in petunia Cab gene promoters. *Mol. Gen. Genet.*, *215(2)*: 337-344.

GILMARTIN, P M; SAROKIN, L; MEMELINK, J and CHUA, N-H (1990). Molecular light switches for plant genes. *Plant Cell*, *2*: 369-378.

GRACE, M L; CHANDRASEKHARAN, M B; HALL, T C and CROWE, A J (2004). Sequence and spacing of TATA box elements are critical for accurate initiation from the beta-phaseolin promoter. *J. Biol. Chem.*, *279*: 8102-8110.

GUBLER, F; KALLA, R; ROBERTS, J K and JACOBSEN, J V (1995). Gibberellin-regulated expression of a myb gene in barley aleurone cells: evidence for Myb transactivation of a high-pI alphase-amylase gene promoter. *Plant Cell*, *7*: 1879-1891.

HALL, T A (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, *41*: 95-98.

HIGO, K; UGAWA, Y; IWAMOTO, M and KORENAGA, T (1999). Plant *cis*-acting regulatory DNA elements (PLACE) database. *Nucl. Acids Res.*, *27(1)*: 297-300.

HONG, H P; ROSS, J H E; GERSTER, J L; RIGAS, S; DATLA, R S S; HATZOPOULOS, P; SCOLES, G; KELLER, W; MURPHY, D J and ROBERT, L S (1997). Promoter sequences from two different *Brassica napus* tapetal oleosin-like genes direct tapetal expression of β-glucuronidase in transgenic *Brassica* plants. *Plant Mol. Biol.*, *34*: 549-588.

JOSHI, C P (1987). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucl. Acids Res.*, *15*: 6643-6653.

KATAGIRI, F and CHUA, N-H (1992). Plant transcription factors: present knowledge and future challenges. *Trends in Genetics*, *8(1)*: 22-26.

LATCHMAN, D S (1997). Transcription factors: an overview. *Intl. J. Biochem. Cell Biol.*, *29*: 1305-1312.

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.

LUO, H; SONG, F; GOODMAN, R M and ZHENG Z (2005). Up-regulation of OsBIHD1, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses. *Plant Biol. (Stuttg)*, *7(5)*: 459-468.

MATARASSO, N; SCHUSTER, S and AVNI, A (2005). A novel plant cysteine protease has a dual function as a regulator of 1-aminocyclopropane-1-carboxylic acid synthase gene expression. *Plant Cell*, *17*: 1205-1216.

MICHALAK, K; ORR, W C and RADYUK, S N (2008). Drosophila peroxiredoxin 5 is the second gene in a dicistronic operon. *Biochem. Biophys. Res. Commun.*, *368(2)*: 273-278.

NAGASAKI, H; SAKAMOTO, T; SATO, Y and MATSUOKA, M (2001). Functional analysis of the conserved domains of a rice KNOX homeodomain protin, OSH15. *Plant Cell*, *13*: 2085-2098.

NIKOLOV, D B and BURLEY, S K (1997). RNA polymerase II transcription initiation: a structural view. *Proc. Natl. Acad. Sci. USA*, *94*: 15-22.

ONG, L M (2001). *An Examination of Embryogenic and Non-embryogenic Cultures of Oil Palm (*Elaeis guineensis *Jacq.)*. Ph.D. dissertation. Universiti Putra Malaysia, Serdang.

ONG-ABDULLAH, M and OOI, S E (2006). Biomarkers: finding a niche in oil palm tissue culture. Part 1 - laying the foundation. *Oil Palm Bulletin No. 53*: 36-48.

ONG-ABDULLAH, M and OOI, S E (2007). Biomarkers: finding a niche in oil palm tissue culture. Part 2 - targeting the transcriptome. *Oil Palm Bulletin No. 54*: 68-74.

OOI, S E (2003). *An Examination of Differentially-expressed Genes from Oil Palm Embryogenic and Non-embryogenic Cultures*. Ph.D. thesis. Universiti Putra Malaysia, Serdang.

OOI, S E; HARIKRISHNA, K and ONG-ABDULLAH, M (2008). Isolation and characterization of a putative serine/threonine kinase expressed during oil palm tissue culture. *JOPR Special Issue on Malaysia-MIT Biotechnology Partnership Programme*, *1*: 14-22.

PELHAM, H R B (1982). A regulatory upstream promoter element in the *Drosophila hsp70* heat shock gene. *Cell*, *30*: 517-528.

ROMBAUTS, S; FLORQUIN, K; LESCOT, M; MARCHAL, K; ROUZE, P and DE PEER, Y V (2003). Computational approaches to identify promoters and *cis*-regulatory elements in plant genomes. *Plant Physiol.*, *132*: 1162-1176.

ROSS, E J; STONE, J M; ELOWSKY, C G; ARREDONDO-PETER, R; KLUCAS, R V and SARATH, G (2004). Activation of the *Oryza sativa* non-symbiotic haemoglobin-2 promoter by the cytokinin-regulated transcription factor, ARR1. *J. Exp. Bot.*, *55*: 1721-1731.

ROUSTER, J; LEAH, R; MUNDY, J and CAMERON-MILLIS, V (1997). Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. *Plant J.*, *11*: 513-523.

ROZEN, S and SKALETSKY, H J (2000). Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz, S and Misener, S eds.). Humana Press, Totowa, NJ. p. 365-386.

SCHERF, M; KLINGENHOFF, A and WERNER, T (2000). Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analyses approach. *J. Mol. Biol.*, *297*: 599-606.

SCHULTZ, D J; CRAIG, R; COX-FOSTER, D L; MUMMA, R O and MEDFORD, J I (1994). RNA isolation from recalcitrant plant tissue. *Plant Mol. Biol. Rep.*, *12*: 310-316.

SEE, P T (2002). *An Examination of Gene Expression (ACC oxidase and receptor-like protein kinases) in Somatic Embryogenesis of Oil Palm* (Elaeis guineensis). M. Sc. thesis. Universiti Putra Malaysia, Serdang.

TEAKLE, G R; MANFIELD, I W; GRAHAM, J F and GILMARTIN, P M (2002). *Arabidopsis thaliana* GATA factors: organization, expression and DNA-binding characteristics. *Plant Mol. Biol.*, *50(1)*: 43-57.

TJIAN, R and MANIATIS, T (1994). Transcriptional activation: a complex puzzle with a few easy pieces. *Cell*, *77*: 5-8.

VAN DER GEEST, A H M and HALL, T C (1996). A 68bp element of the β-phaseolin promoter functions as a seed-specific enhancer. *Plant Mol. Biol.*, *32*: 579-588.

VANDEPOELE, K; QUIMBAYA, M; CASNEUF, Y; DE VEYLDER, L and VAN DE PEER, Y (2009). Unraveling transcriptional control in *Arabidopsis* using *cis*-regulatory elements and coexpression networks. *Plant Physiol.*, *150*: 535-546.

XU, N; HAGEN, G and GUILFOYLE, T (1997). Multiple auxin response modules in the soybean SAUR 15 A promoter. *Plant Sci.*, *126*: 193-201.

YAMAMOTO, Y Y; ICHIDA, H; ABE, T; SUZUKI, Y; SUGANO, S and OBOKATA, J (2007). Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. *Nucl. Acids Res., 35(18)*: 6219-6226.

YANG, S F and HOFFMANN, N E (1984). Ethylene biosynthesis and its regulation in higher plants. *Annu. Rev. Plant Physiol.*, *35*: 155-189.