Songklanakarin J. Sci. Technol.
35 (3), 275-285, May - Jun. 2013

SJST SONGKLANAKARIN JOURNAL OF SCIENCE AND TECHNOLOGY
http://www.sjst.psu.ac.th

Original Article

Molecular cloning and characterization of the late embryogenesis abundant group 4 (*EgLEA*4) gene from oil palm (*Elaeis guineensis* Jacq)

Watcharasuda Hualkasin[3], Waleerath Thongin[2], Kanokon Petsean[2], Amornrat Phongdara[1,2], and Alisa Nakkaew[1,2]*

[1] *Center for Genomic and Bioinformatics Research, Faculty of Science,*

[2] *Department of Molecular Biotechnology and Bioinformatics, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, 90110 Thailand.*

[3] *Faculty of Science and Industrial Technology, Prince of Songkla University, Mueang, Surat Thani, 84000 Thailand.*

Received 24 October 2012; Accepted 26 March 2013

Abstract

The Late-Embryogenesis Abundant group 4 (*LEA*4) genes is a group of genes that have been reported to be involved in stress and hormone responses. The completed *LEA*4 cDNA sequence was first obtained from a set of EST sequences of oil palm (*Elaeis guineensis* Jacq), named as *EgLEA*4. The open reading frame is 486 bp in length, encoding a deduced amino acid sequence of 161 residues with a molecular weight of 16.5 kDa and a pI value of about 8.0. Five amino acid motif patterns were found in the EgLEA4 (II, I, III, IV and V) and each had a close identity to similar LEA4 patterns of soybean (64%). Comparison of the nucleotide sequences of the cDNA and the genomic DNA demonstrated that the *EgLEA*4 gene is composed of 2 exons and 1 intron. The 52 untranslated region shows a putative promoter sequence involved in the transcription process, drought stress and hormone responsive elements. RT-PCR analysis showed that the *EgLEA*4 gene was only expressed in mesocarp, during the late stages of fruit development. It also had a higher expression in induced drought conditions indicating that the EgLEA4 protein may be involved in plant adaptation and stress (drought) responsive pathway.

Keywords: late-embryogenesis abundant protein, drought, oil palm, *EgLEA*4, *Elaeis guineensis*

1. Introduction

Oil palm is the most productive oil crop in the world, a single hectare of oil palm yields more crude oil than other oil generating crops (the productivities of oil palm: rapeseed: soybean are 4.14: 0.72: 0.40 tonnes/hectare/year) (Oil World, 2012: Wong and Bernardo, 2008). Palm oil is becoming an increasingly important agricultural product for tropical countries owing to its potential to replace fossil fuels. Under selective pressures imposed by adverse environmental conditions, plants have evolved fast stress responses and crop plants have been developed with enhanced tolerance to stress. Developing a basic understanding of the physiological, biochemical and gene regulatory networks associated with stress responses is therefore essential (Dobra *et al.*, 2011; Novikova *et al.*, 2007; Rizhsky *et al.*, 2002; Valliyodan and Nguyen, 2006) and has become a challenging task for the development of crop cultivars with improved adaptation to various environments.

Late embryogenesis abundant (LEA) proteins were initially identified in the late stages of seed maturation in cotton and wheat and constitute a set of proteins widespread in the plant kingdom (Campos *et al.*, 2006; Dure *et al.*, 1989). Five groups of typical LEA-like proteins have been described

* Corresponding author.
Email address: alisa.n@psu.ac.th

on the basis of their similar amino acid sequences and conserved motifs (Ingram and Bartels, 1996). The proteins are highly hydrophilic and contain a high percentage of small amino acid residues such as glycine, alanine and serine. These general characteristics have led to the identification of a wide group of proteins, present in different organisms, termed hydrophilins, in which many LEA proteins are included (Campos *et al.*, 2006; Garay-Arroyo *et al.*, 2000). The expression of these proteins has been reported to be induced by plant stress hormones and is associated with a tolerance to desiccation of their seeds and pollen and with anhydrobiotic plants (Bray, 1993; Dalal *et al.*, 2009; Finkelstein *et al.*, 2002; Wise and Tunnacliffe, 2004).

It has been speculated that LEA proteins may protect cellular and macromolecular structures or are detoxifying molecules that can alleviate the increase in osmotic pressure caused by the increased ion concentrations during stress from low water. Some LEA proteins such as group 4 proteins have been found in cotton, LEA14, D113 and Craterostigma PGC27-45. These can form a supplementary structure that adapts to the conformational changes of other proteins and functions in protecting membrane stability and integration during drying and dehydration (Dure, 1993b; Garay-Arroyo *et al.*, 2000). *In vitro* studies of the LEA4 family of proteins of Arabidopsis showed that their presence, during controlled dehydration experiments, prevented the inactivation of LDH, even after a 99% water loss (Reyes *et al.*, 2005). Group 4 LEA mRNA from Brassica napus was highly up-regulated in flowers under salt stress (Dalal, 2011; Dalal *et al.*, 2009). Furthermore, the overexpression of BnLEA4-1 in transgenic Arabidopsis plants plays a crucial role in its tolerance to abiotic stress and an enhanced tolerance to high salt and drought stress during plant development (reviewed by Dalal *et al.*, 2009; Hong-Bo *et al.*, 2005).

However, the precise function of LEA proteins in plant development and stress responses remains to be clarified (Su *et al.*, 2011). According to the possible important roles of LEA proteins as mentioned above, this work focuses on LEA proteins from oil palm. We report here, for the first time, on the isolation and molecular characterization of the *LEA*4 gene from oil palm and named as *EgLEA*4. We provide important information for a better global understanding of LEA proteins. This includes data on their sequence conservation and variation as well as their expression patterns in seeds at different developmental stages and in various oil palm tissues. A possible role of the EgLEA proteins in drought tolerance has been detected and has opened the way to produce and select oil palm crops with such characteristics

2. Materials and Methods

2.1 Plant materials

Mesocarp tissue from at 10–22 weeks after anthesis (waa) old oil palm fruits (*Elaeis guineensis* Jacq) was used for extraction of chromosomal DNA and mRNA. Young leaves, mature leaves, stems and roots of oil palm from 1-year-old oil palm seedling were used for extraction of RNA.

2.2 Cloning of *EgLEA*4 structural gene

The full length cDNA of *EgLEA*4 was obtained previously from our EST library (Phongdara *et al.*, 2012). The genomic sequence of *EgLEA*4 gene was isolated using the GenomeWalking™ method (Clontech). The genomic DNA was completely digested by restriction enzymes; *Dra*I, *Eco*RV, *Stu*I and *Pvu*II. The products of the digestion were purified by the phenol/chloroform method. Each reaction product was ligated separately to the GenomeWalker Adaptor at the 52-end. The primary PCR used primers AP-1 (GenomeWalker primer) and GSP1 (Table 1), and the ligated products were used as the template. The cycling program was: 94° C for 25 seconds and 72°C for 3 min with 7 cycles, then 94°C for 25 seconds and 67°C for 3 min with 32 cycles and then 67°C for 7 min with 1 cycle. The secondary PCR used nested primers AP-2 (GenomeWalker primer) and GSP2 (Table 1), and the primary PCR product as the template. The cycling program was: 94°C for 25 seconds and 72°C for 3 min with 5 cycles, then 94°C for 25 seconds and 67°C for 3 min with 20 cycles and then 67°C for 7 min with 1 cycle. The PCR products were purified and the sequences were determined.

2.3 Sequencing and bioinformatic analysis of the *EgLEA*4 gene

Protein alignments were then performed using ClustalX and displayed using GENEDOC software (Larkin *et al.*, 2007). The phylogenetic analysis, performed by the protein sequence Parsimony method, was computed by PROPARS and used to construct trees with support for the inferred groups obtained by bootstrap analysis from 1000 replications of the data set using the SEQBOOT and CONSENSE programs. The phylogenetic tree was displayed using the TreeView version 1.0 (Felsenstein, 1997). The protein modeling predictions for the structures were performed with the SWISS-MODEL (Arnold *et al.*, 2006) and the I-TASSER server (Roy *et al.*, 2010) based on multiple template simulations and used for predicting a model structure. Next, each of the predicted models was validated on the best selected model by the PROCHECK software

Table 1. Primers for cloning, chromosome walking and expression of *EgLEA*4

Primer name	Sequences (5' → 3')
EgLEA4F	GCA TGG AGA AGA CCA AGG CCA
EgLEA4R	AAG CTT CTA GTC GTA GCC ACC ACC GG
GSP1	CTT CTC CTC GAG GGT GGC CTT GGT
GSP2	GTC TCC TTG ACG GAC GCC ATG
18SF	CAA AGC AAG CCT ACG CTC TG
18SR	CGC TCC ACC AAC TAA GAA CG

(Laskowski, 2001). The molecular surfaces and protein structure models of the figures were built and visualized with PyMOL (DeLano, 2002). The 52 untranslated region (52-UTR) was analysed by the Genomatix software, that included the data and tools able to detect transcription factor binding sites (MatInspector software, http://www.genomatix.de/online_help/help_matinspector/matinspector_help.html) and promoter modules (composite regulatory elements), and to predict and retrieve promoter sequences.

2.4 Expression analysis of *EgLEA4*

Total RNA was extracted using TRIZOL reagent (Invitrogen) and was used as a template for the synthesis of the of EgLEA4 cDNA. Primers EgLEA4F and EgLEA4R (Table 1) were used in the PCR reaction using the following program: a first step for cDNA synthesis at 50°C for 30 min, 95°C for 15 min, followed by 30 cycles at 95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 1 min with a final extension at 72°C for 10 min. The amplified products were separated by electrophoresis on a 1.5% agarose gel, then visualized by and photographed using the Image Analyzer, Gel Doc 1000 (BIO-RAD). The RT-PCR products with the expected sizes were purified and then inserted into the pGEM®-T Easy Vector (Promega). Subsequently, the recombinant DNA plasmids were sequenced.

2.5 Expression analysis of *EgLEA4* gene in various tissues

For determining the expression in various tissues, RNA was extracted from young leaves, mature leaves, roots, stems tissue of oil palm from 1-year-old oil palm seedling and mesocarp tissues of oil palm fruit at 22 waa as the templates using RT-PCR. For determining the expression at different ages of the oil palm fruit, RNA was extracted from the mesocarp tissues at 10, 12, 14, 16, 18, 20 and 22 waa. For drought stress treatment, RNA was extracted from embryo tissue from normal seed and dehydrated seed (72 h dehydrated in a humidifier chamber at 65% RH, 37°C). The RT-PCR reactions were performed with primers as indicated above. An 18S rDNA primer set (Table 1) of 540 bp fragment size, was used as an internal control. The cycling program used was : one cycle at 94°C for 5 min, 30 cycles at 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 30 seconds and with a final extension at 72°C for 10 min. The amplified products were separated by electrophoresis on a 1.5% agarose gel, then visualized by and photographed using the Image Analyzer, Gel Doc 1000 (BIO-RAD) followed by analysis of the expression levels with the Scion Image program and the SPSS program. The experiments were performed in triple.

3. Result

3.1 Cloning and characterization of the *EgLEA4*

An *EgLEA4* cDNA fragment of 751 bp, corresponding to an ORF of 486 bp (GenBank accession number JF810460), was cloned (Figure 1A). The predicted 161 amino acid polypeptide had a calculated molecular mass of 16.5 kDa and a pI of 8.0. BLAST analysis (http://blast.ncbi.nlm.nih.gov/) and sequence alignment of the encoded polypeptide sequence showed that Eg*LEA4* was 64% homologous at the amino acid level to the soybean LEA4 protein, *Glycine tabacina* (AAG 37441.1), and 61% to that of *G. tomentella* (AAG37451.1). A phylogenetic tree was constructed (Figure 1B). The predicted amino acids of EgLEA showed a preponderance of Ala (alanine), Gly (glycine), Thr (threonine), and Glu (glutamate) that constituted 19.3, 12.4, 10.6 and 9.9%, respectively, of the total. The protein had no phenylalanine, cysteine and tryptophan. Of the total amino acids in EgLEA4 50.33% were hydrophilic residues (S, Q, N, T, R, K, H, D and E) of which 29.2% were charged residues (D, E, R, K, H) and 49.68% were

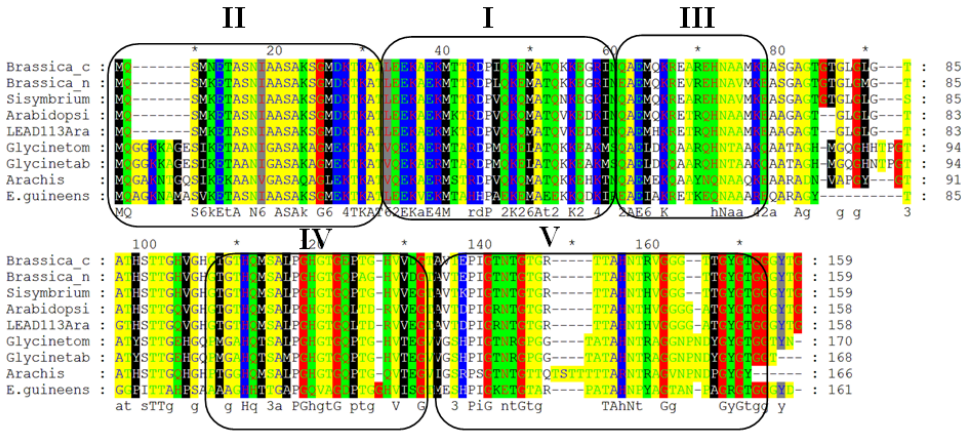

Figure 1A. The multiple amino acid sequences alignment (ClustalX 2.0) of EgLEA4 from oil palm (*Elaeis guineensis* Jacq; JF810461) compared with the LEA4 protein from the other plants; Brassica _c (AAT77224.1), Brassica_n (AAT77223.1), Sisymbrium (AAY26119.1), Arabidopsi (NP_196294.1), LEAD113 Ara (CAA63008.1), Glycine tom (AAG37451.1), Glycine tab (AAG37441.1) and Arachis (ADQ91840.1) and visualized with the Genedoc program showing five conserved motifs (I-V) are indicated by the enclosed black boxes.

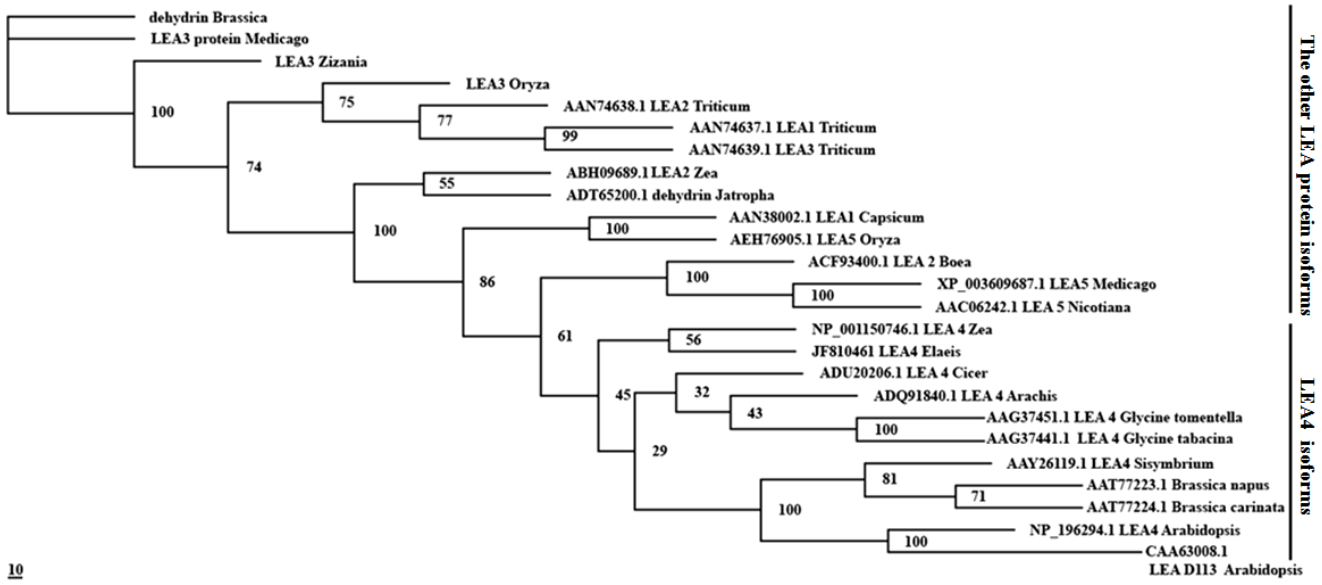

Figure 1B. The phylogenetic tree analysis of EgLEA4 and other LEA protein isoforms was analyzed using PROPARS (Protein Sequence Parsimony Method) programs. Support for the inferred groups was obtained by bootstrap analysis from 1000 replications of the data set using the SEQBOOT in PHYLIP 3.57c package (Felsenstein, 1997). The accession numbers and corresponding species is provided in the supplemental table 1 with abbreviations and accession numbers.

hydrophobic residues (A, V, G, M, L, I, P and Y) (Bies-Etheve *et al.*, 2008; Wise, 2003). Hydropathicity analysis predicted that this protein was a hydrophilic protein and was similar to the Brassica sp. and soybean LEA4 proteins. A pfam03760 domain (E-value, $2.22e^{-09}$) at the amino acids 1 to 73 (motif I-III) of the putative protein sequences was found by a conserved domain search analysis. Because the pfam03760 domain is a classified conserved domain of Group 4 LEA proteins, therefore our EgLEA can be considered as part of the family of LEA 4 isoforms (EgLEA4) (Marchler-Bauer *et al.*, 2011).

The secondary and tertiary structure predictions demonstrated that EgLEA4 contained a heavy α-helix (red) at its N-terminal and a periodically random coil (green) at its C-terminal (Figure 2a). In addition, two binding sites were found in the random coil region. The first binding site was present at positions 87-90 (Gly-Pro-Ile-Thr) (Figure 2b) and the second binding site was at positions 131-132 (Lys-Glu) (Figure 2c). Because of its helix structure both binding sites were in the random coil region of the EgLEA4 and therefore able to play some role in protecting and maintaining the common physiological processes in plant cells under a water deficit (Battaglia *et al.*, 2008; Kyte and Doolittle, 1982; Shih *et al.*, 2004).

3.2 Structural gene analysis of *EgLEA4*

The *EgLEA4* genomic DNA sequence was 799 bp (accession number JF810461) in length. When compared to the cDNA sequence (486 bp) an intron (313 bp) was found between two exons (108 bp and 378 bp) (Figure3). This result is similar to that from the upland cotton LEA4 (Accession no: M73752 *Gossypium hirsutum* Lea4-A gene) that has an intron of 101 bp between two exons. (193 bp and 522 bp).

The EgLEA4 promoter sequence (Figure3; red characters) at the 52-UTR was obtained by the GenomeWalking™ method and analysed by the Genomatix software, which also includes the data and tools to detect transcription factor binding sites (MatInspector database). There were two putative cis-acting regulatory elements perhaps involved in

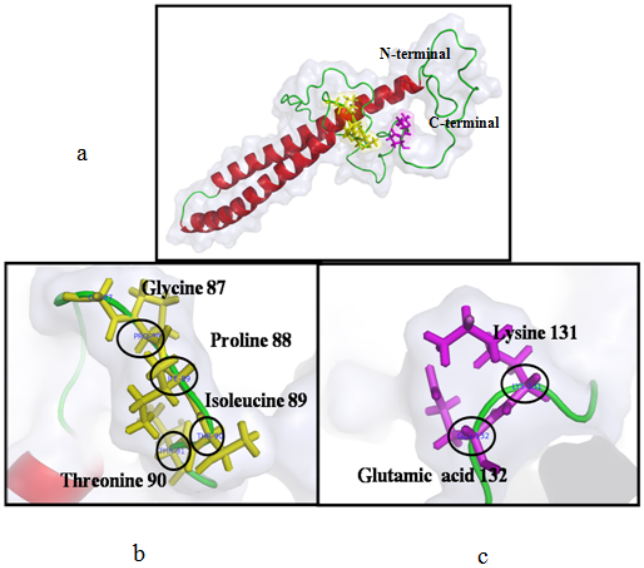

Figure 2. The tertiary structure prediction of the EgLEA4 protein (a).The first binding site (b) and the second binding site (c).

abiotic stress responses, such as an ABA-responsive element (ABRE) and a salt/drought/ dehydration responsive element (DRE) detected. An ACGT containing the ABA-responsive elements was found. Most ACGT-containing ABRE motifs are bound by a basic leucine zipper (bZIP) transcription factor (Yamaguchi-Shinozaki and Shinozaki, 2006). These ABREs and DRE were able to stimulate gene expression in response to ABA under abiotic stress and dehydration (salt, drought and cold stresses) (Shinozaki and Yamaguchi-Shinozaki, 2000; Simpson *et al.*, 2003). The general transcription factor IID and the core promoter motif ten element, promotes transcription by RNA polymerase II (Lim *et al.*, 2004). The *Dc3* promoter binding factors (DPBF) are involved in seed specific and/or ABA-induced expression of the *Dc3* gene, a carrot *LEA* gene (Kim *et al.*, 1997). The Opaque-2 like transcriptional activators contains an ACGT core that is necessary for DNA binding (Schmidt et al., 1992). The promoter of the LEA-like protein genes contains DRE, ABREs and MYB recognition elements. ABRE and MYB/MYC elements control gene expression in response to ABA under abiotic stress (Shinozaki and Yamaguchi-Shinozaki, 2000). However, the MYB/MYC was not found in EgLEA4. The exact positions of the EgLEA4 promoters are listed in Table 2.

3.3 Expression of the *EgLEA*4 gene in oil palm and its relation to drought

A semi-quantitative RT-PCR analysis using the *18S rDNA* gene as an internal control revealed that *EgLEA*4 was not expressed in vegetative tissues but only in the mesocarp and embryo from oil palm fruit under normal conditions (Figure 4A and 4B). The highest expression during the oil palm fruit development was found at the age of between 20 and 22 weeks which is a late stage of oil palm fruit development and then after 22 weeks the oil palm fruit will be harvested (Figure 4B). This result is similar to the expression of the AhLEA4 genes in peanut which is only detected in seeds. This indicates that the EgLEA4 protein accumulates and might endure in some pathway during the late stage of seed/fruit development (Su *et al.*, 2011). When the embryos of oil palm were left at room temperature as a control condition and compared with those left dehydrated in humidifier chamber at 65% RH for 72 h, the expression level of *EgLEA*4 was significantly more highly induced in the dehydrated seed compared to the normal seed (Figure 4C). This indicated that the EgLEA4 protein may play some role in the drought stress tolerance of oil palm seeds.

```
AAATAATAATTTTTTTTAATAATGAATGATATAGTAATATTTTGTTGAATGGTGTAAGAT 60
ATATGTAATAGAAGGATTGTAATCAATAATTTCTCCCTCGAGGAAGGATGACAAAACGAT 120
TAAATGAACCAAGGTGAAACGATAACCTGTGCACGTGCCCTTCTTGCATGCAATCCAGCT 180
TTGCCACGCGTAAACAAATCAAAAGAGTGGAGAAAACCGTTCCCACCACTCGAAGGTCCA 240
AGCGCACCTTCGTCCGCCACGTACCTAACGAGCCCCCAAAGCCGTCCACGTGTCCTCTCT 300
GGCCGAAGATACGTGTCAACCGCACGTAACCGACTCCGCGACTATATCCATGTCTCTGTG 360
CAGGCGATGATCGTACCCACGCCTCCATCCGAAGAGATCAGACGTCAAGCAGGAAAGGTT 420
GGAAGCGAAGAAAAGGAAAGGATGGTGATGCAAGCCGGGAAAAACGCCATGGCGTCCGTC 480
                           M  Q  A  G  K  N  A  M  A  S  V
AAGGAGACGGCCTCCAACATCGCCGCCTCGGCCAAGTCCGGCATGGAGAAGACCAAGGCC 540
 K  E  T  A  S  N  I  A  A  S  A  K  S  G  M  E  K  T  K  A
ACCCTCGAGGAGAAGGTTGGATATATACACGACATGTAGCTTTGCTTAAGAACGTTTCCC 600
 T  L  E  E  K
ATCGGTTAATCTGTTTTCTATTCGATCATGTTTGACGGAATAGCTTATCCACGGAATAAC 660
AATATCACCATTTTCTTAAATATCCACTTCACCTTCAAAACATGAAGGAAATAAAAACAT 720
TTAACGGCTAAAAAGTGGAGTACGTAGATTTGATTCCATTCAACTCTGCTTTAGTCCCAA 780
ACAAAACGGACGTCTGGTTATGGTCTCATCGGGGGCCTTTTTTCGCTTGATTTGTTTCCG 840
GTAATATTTCGTATTTATTGATGAGCAGGTGGAGAAGATGACGGCCCACCACCCGGCGGA 900
                           V  E  M  T  A  H  H  P  A  E
GAAGGAGATGGCAGAGGAGAAGAAGCAGGACAAGATCCGAGAGGCGGAGATCGCAAAGCG 960
  K  E  M  A  E  E  K  K  Q  D  K  I  R  E  A  E  I  A  K  R
CGAGACCAAGGAGCAGAATGCCGCCGCCCGGGAGCAGGCCCGCGCTGGCCACACCGGCGG 1020
  E  T  K  E  Q  N  A  A  A  R  E  Q  A  R  A  G  H  T  G  G
CCCCATCACCACTGCCCACCCTTCCGCAGCCGCCGCCGGTCATCACACCACCGGGGCGCC 1080
  P  I  T  T  A  H  P  S  A  A  A  A  G  H  H  T  T  G  A  P
CGGCCAGGTCGCCGGTGACCCTACCGGGGGTCACGTGATAAGCGGAACAATGGAGTCCCA 1140
  G  Q  V  A  G  D  P  T  G  G  H  V  I  S  G  T  M  E  S  H
TCCCAACGGTAAGGAGACTGGGACTGCCAGGCCTGCCACTGCCCACAACTCCTATGCTGG 1200
  P  N  G  K  E  T  G  T  A  R  P  A  T  A  H  N  S  Y  A  G
CACCGCTAACCCGGCCGGGCGGGGAACCGGTGGTGGCTACGACTAGGAAGGGCTTGGAAT 1260
  T  A  N  P  A  G  R  G  T  G  G  G  Y  D  *
GTGTGAACAAATTTGGAAGGCACCTTGCTGTGTTTTCTATTTTATCCTCTACTACTAGCA 1320
GCAGTAGCAGCGTTCTTAGGAGTGTGGGTTTCTGTTTGTTATTTATTACGACGTGTTTCT 1380
GTTGTTTGTTAAGCTGTGTGCTGCTTGTGTTCCCTGGTATGCTACGTGTGTATGTTTTTA 1440
AGTCAAGAAAAATCACCCCTACTATATTGCAAGGGAAAAAAAAAAAAAAAAAAAAAAAAA 1500
AAAAA                                                          1505
```

Figure 3. The structural gene of *EgLEA*4 in oil palm (*E. guineensis*) genome, consisted of 1505 bp containing an intron (313 bp; blue characters) between two exons (108 bp and 378 bp; black characters) in the genomic sequence with an ORF of 486 bp encoding for a predicted 161 amino acid residues for EgLEA4 (JF810461). Finally, the promoter region of an *EgLEA*4 (447 bp) and the 32 UTR with the poly A tail (259 bp) is shown, respectively.

Table 2. Positions of the various recognition elements from the promoter sequence of the nucleotide of the 5′ untranslated *EgLEA*4 gene in oil palm

Promoter component	Position		score	Sequence
	From	To		
GT-box elements	10	26	0.919	TTCATTATTAAAAAAAA
TEF-box	80	100	0.866	CGAGGGAGAAATTATTGATTA
Plant G-box/C-box bZIP proteins	144	164	0.966	AGAAGGGCACGTGCACAGGTT
	250	270	0.989	CGTTAGGTACGTGGCGGACGA
	279	299	0.992	GAGAGGACACGTGGACGGCTT
	393	413	0.963	CCTGCTTGACGTCTGATCTCT
	303	323	0.968	CCGAAGATACGTGTCAACCGC
ABA response elements (ABRE)	146	162	0.897	CCTGTGCACGTGCCCTT
	281	297	0.917	GCCGTCCACGTGTCCTC
	304	320	0.880	CGAAGATACGTGTCAAC
Plant specific NAC [NAM (no apical meristem), ATAF172, CUC2 (cup-shaped cotyledons 2)] transcription factors	254	266	0.967	CCGCCACGTACCT
General transcription factor IID, GTF2D	371	409	0.712	TCGTACCCACGCCTCCATCCGAA GAGATCAGACGTCAA
Dc3 promoter binding factors	150	160	0.894	TGCACGTGCCC
	285	295	0.934	TCCACGTGTCC
ER stress-response elements	170	188	0.831	GCAATCCAGCTTTGCCACG
Salt/drought responsive elements (DRE)	220	234	0.961	TTCGAGTGGTGGGAA
Core promoter motif ten elements	234	254	0.819	AGGTCCAAGCGCACCTTCGTC
Opaque-2 like transcriptional activators	318	334	0.970	AACCGCACGTAACCGAC

4. Discussion and Conclusion

The *EgLEA*4 gene belongs to the group 4 *LEA* gene family (classified by the presence of the conserved pfam03760 domain) and has been cloned and sequenced from the EST library of Oil palm (*E. guineensis* Jacq). The open reading frame of the *EgLEA*4 gene (GenBank accession number JF810460) is 486 bp, compared to 507 bp for soybean *LEA*4 and 519 bp for peanut *LEA*4 (Shih *et al.*, 2010; Su *et al.*, 2011). The deduced amino acid sequence of the *EgLEA*4 gene in oil palm encrypts for 161 residues. Most of LEA proteins are small in molecular weight ranging from 10-30 kDa (Hundertmark and Hincha, 2008).

The calculated molecular weight and pI of EgLEA4 at 16.5 kDa and 8.0 which is close to the value for soybean and peanut with molecular weight of 17.6 and 17.30 kDa and pI values of 9.17 and 9.68, respectively (Shih *et al.*, 2010; Su *et al.*, 2011). The EgLEA4 protein also contains amino acid residues that can be glycosylated (N, R, S and T) (Spiro, 2002). Its amino acid content is similar to that of other LEA4 proteins, EgLEA4 is rich in glycine and amino acids containing a hydroxyl group (Threonine, Serine) that can bind water molecules and function as a solvation film (Senthil-Kumar *et al.*, 2007), thus stabilizing the surface of membranes.

In the prediction of its secondary structure, the motifs at the N-terminal of the EgLEA4 proteins form a heavy α-helix structure whereas the rest of the protein produces a random coil conformation as reported by others (Hundertmark and Hincha, 2008; Kyte and Doolittle, 1982; Shih *et al.*, 2004). Two possible binding sites for water were found in the random coil region (Dure, 1993a). The circular dichoism (CD) spectroscopy of LEA4 rotifer protein has shown a random coil structure in the hydrated state and a high level of α-helix after drying. An *in vitro* assay, by FTIR and far-UV circular and dichroism (CD) revealed that LEA proteins in solution become unfolded and then refold properly upon desiccation (Goyal *et al.*, 2003).

Earlier studies have indicated that the LEA group of proteins is highly expressed upon desiccation in embryos (Ganesh, 1999; Savitha, 2000; Senthil-Kumar and Udaya-kumar, 2006), so may be implicated in the desiccation tolerance of embryos. The results from our expression profiles of various tissues show that expression of the *EgLEA*4 gene occurs only in the mesocarp tissue and is expressed most strongly in week 20 and 22, during the late stage of oil palm fruit development because at the late stage the water content of fruit's tissue is decreased (Gribaa *et al.*, 2012). This is again similar to other group 4 proteins found to be highly accumu-

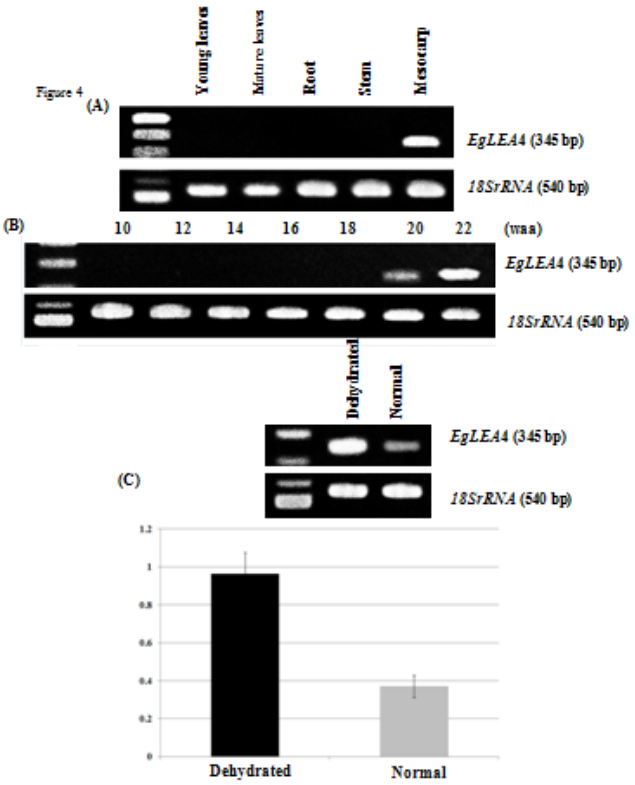

Figure 4. *EgLEA*4 gene expressions of oil palm in various tissues (A), different age of oil palm fruits (B) and embryo tissue in normal and dehydrated seed condition (C).

lated in dry seed embryos or the late stages of seed development (Baker *et al.*, 1988; Delseny *et al.*, 2001; Roberts *et al.*, 1993; Su *et al.*, 2011). Seeds that were dehydrated had a significantly higher expression of *EgLEA*4 gene than found in normal seed. This demonstrates that the EgLEA4 protein may protect the cellular structure by acting perhaps as a chaperone that stabilizes vesicles, proteins and membrane structure, sequestering ions, renaturing unfolded proteins, and protecting other proteins in the same way as other LEAs (Bray, 1993; Kaur and Gupta, 2005) that increase in response to drought (Cohen *et al.*, 1991). A gradual decrease in water accessibility leads to conformation changes in enzymes that are associated with inactivation (Kaur *et al.*, 2007). These prohibited conformational changes do not arise when hydrophillins are present prior to dehydration (Battaglia *et al.*, 2008). However, during severe water limitation, hydrophillins have no protective effect (Reyes *et al.*, 2005). The present study provides evidence to show that the *EgLEA*4 gene may play some role in imparting tolerance to desiccation-stress.

The promoter region of the *EgLEA*4 gene contains the putative cis-acting regulatory elements, ABA-responsive elements (ABRE), salt/drought/ dehydration responsive element (DRE) and an NAC transcription factor element that might be related to drought resistance. The ABREs already described for ABA induction, both in seeds and in vegetative tissue, have been detected (Cao *et al.*, 2007; Shen *et al.*, 1996). The ABRE related to drought stress contains two palindromic motifs CACGTG with the ACGT core element. Our results contain more than one sequence element with the ACGT core (Table 2). The binding studies have revealed that the nucleotides flanking the ACGT core specify a DNA-protein interaction (Narusaka *et al.*, 2003). ACGT elements have been observed in many plant genes regulated by environmental change and physiological factors (Shinozaki and Yamaguchi-Shinozaki, 2000). The existence of the ACGT core sequence in the promoter region indicates that the genes may be mediated by ABA (Busk *et al.*, 1997) and can bind with bZIP-type transcription factors (Uno *et al.*, 2000) also known as auxin binding factors (ABFs). Overexpression of these regions is known to be associated with several ABA-associated phenotypes, including enhanced drought tolerance (Kang *et al.*, 2002). The DRE motif might be a coupling element of ABRE involved in interdependence in the ABA-responsive expression in response to ABA such as in Arabidopsis (Narusaka *et al.*, 2003). When the DRE binding protein was overexpressed in transgenic plants (Arabidopsis), changes in the expression of more than 40 stress-inducible genes were identified, and these changes led to increased freezing, salt, and drought tolerance (Maruyama *et al.*, 2004). Furthermore, NAC transcription factor (TFs) elements were also detected in the promoter region of the *EgLEA*4 gene. Overexpression of the NAC TF proteins resulted in several stress-inducible gene being upregulated and plants showed a significant increase in drought tolerance in plants (Valliyodan and Nguyen, 2006). There are additional gene expression pathways that are regulated through NAC TFs under dehydration stress in Arabidopsis (Nakashima *et al.*, 2009). Three cDNA encoding NAC like proteins in Arabidopsis, NAC domain-containing proteins, have bound to one of the cis-elements in the ERD1 (Early response to dehydrate 1) promoter (Tran *et al.*, 2004). Other transcription factors, zinc-finger homeodomains (ZFHD1) can bind other cis-element in the ERD1 promoter and function as transcriptional activators in response to dehydration stress (Tran *et al.*, 2007). From three elements (ABRE, DRE and NAC TF) in the EgLEA4 promoter, it is possible that the EgLEA4 protein may have a role to enhance drought tolerance.

In conclusion, drought is one of the important factors that limit crop productivity, including that of oil palm as an economic higher plant. In the present study, the molecular data, sequence analysis, gene structure and expression have provided some basic knowledge about the *EgLEA*4 gene which may be of benefit to improvements of drought tolerance in oil palm and other plants.

Acknowledgements

This work was supported by the Research Assistant (RA) fund from the faculty of Science (1-2554-02-007) to Ms. Waleerath Thongin and the National Research University Project of Thailand's Office of the Higher Education Commission (SCI540561S) for financial support to Ms. Kanokon

Petsean. Furthumore, the Program Strategic Scholarships for Frontier Research Network for the Ph.D. Program Thai Doctoral degree from the Office of the Higher Education Commission, Thailand and the Thailand Research Fund; TRF (RDG5420022)for support to Alisa Nakkaew, the Graduated School, Center for Genomics and Bioinformatics Research, Faculty of Science and Prince of Songkla University. We thank Dr. Brian Hodgson for assistance with the English and valuable comments.

References

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. Bioinformatics, 22(2), 195-201.

Baker, J., Steel, C. and Dure, L. 1988. Sequence and characterization of 6 LEA proteins and their genes from cotton. Plant Molecular Biology, 11(1), 277-291.

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F. and Covarrubias, A.A. 2008. The enigmatic LEA proteins and other hydrophilins. Plant Physiology, 148(1), 6-24.

Bies-Etheve, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., Cooke, R. and Delseny, M. 2008. Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. Plant Molecular Biology, 67(1-2), 107-124.

Bray, E.A. 1993. Molecular Responses to Water Deficit. Plant Physiology, 103(4), 1035-1040.

Busk, P.K., Jensen, A.B. and Pages, M. 1997. Regulatory elements in vivo in the promoter of the abscisic acid responsive gene rab17 from maize. Plant Journal, 11 (6), 1285-1295.

Campos, F., Zamudio, F. and Covarrubias, A.A. 2006. Two different late embryogenesis abundant proteins from *Arabidopsis thaliana* contain specific domains that inhibit Escherichia coli growth. Biochemical and Biophysical Research Communications, 342(2), 406-413.

Cao, H.X., Zhang, Z.B., Xu, P., Chu, L.Y., Shao, H.B., Lu, Z.H. and Liu, J.H. 2007. Mutual physiological genetic mechanism of plant high water use efficiency and nutrition use efficiency. Colloids Surf B: Biointerfaces, 57(1), 1-7.

Cohen, A., Plant, A.L., Moses, M.S. and Bray, E.A. 1991. Organ-Specific and Environmentally Regulated Expression of Two Abscisic Acid-Induced Genes of Tomato : Nucleotide Sequence and Analysis of the Corresponding cDNAs. Plant Physiology, 97(4), 1367-1374.

Dalal, I. 2011. Allergy to seeds. Pediatric Allergy Immunology, 22(1 Pt 2), e150-154.

Dalal, M., Tayal, D., Chinnusamy, V. and Bansal, K.C. 2009. Abiotic stress and ABA-inducible Group 4 LEA from *Brassica napus* plays a key role in salt and drought tolerance. Journal of Biotechnology, 139(2), 137-145.

DeLano, W.L. 2002. The PyMOL Molecular Graphics System. http://wwwpymol.org

Delseny, M., Bies-Etheve, N., Carles, C., Hull, G., Vicient, C., Raynal, M., Grellet, F. and Aspart, L. 2001. Late embryogenesis abundant (LEA) protein gene regulation during Arabidopsis seed maturation. Journal of Plant Physiology, 158, 419-427.

Dobra, J., Vankova, R., Havlova, M., Burman, A.J., Libus, J. and Storchova, H. 2011. Tobacco leaves and roots differ in the expression of proline metabolism-related genes in the course of drought stress and subsequent recovery. Journal of Plant Physiology, 168(13), 1588-1597.

Dure, L. 1993a. Structural motifs in LEA proteins In TJ Close, EA Bray, eds, Plant Responses to Cellular Dehydration during Environmental Stress. American Society of Plant Physiologists, Rockville, 91-103.

Dure, L., 3rd 1993b. A repeating 11-mer amino acid motif and plant desiccation. Plant Journal, 3(3), 363-369.

Dure, L., Crouch, M., Harada, J.J., Ho, T., Mundy, J., Quatrano, R.S., Thomas, T.G. and Sung, Z.R. 1989. Common amino acid sequence domains among the LEA proteins of higher plants. Plant Molecular Biology, 12, 475-486.

Felsenstein, J. 1997. An alternating least squares approach to inferring phylogenies from pairwise distances. Systematic Biology, 46(1), 101-111.

Finkelstein, R.R., Gampala, S.S. and Rock, C.D 2002. Abscisic acid signaling in seeds and seedlings. Plant Cell, 14 Supplement, S15-45.

Ganesh, K. 1999. Identification of thermotolerant lines in sunflower (*Helianthus annuus* L) based on temperature induction response (TIR) technique: role of HSPs and LEAs in temperature and osmotic stress. PhD thesis, University of Agricultural Sciences, Gandhi Krishi Vignan Kendra, Bangalore, India.

Garay-Arroyo, A., Colmenero-Flores, J.M., Garciarrubio, A. and Covarrubias, A.A. 2000. Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. The Journal of Biological Chemistry, 275(8), 5668-5674.

Goyal, K., Tisi, L., Basran, A., Browne, J., Burnell, A., Zurdo, J., Tunnacliffe, A. 2003. Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. The Journal of Biological Chemistry, 278(15), 12977-12984.

Hong-Bo, S., Zong-Suo, L. and Ming-An, S. 2005. LEA proteins in higher plants: structure, function, gene expression and regulation. Colloids Surf B: Biointerfaces, 45(3-4), 131-135.

Hundertmark, M. and Hincha, D.K. 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics, 9, 118.

Ingram, J. and Bartels, D. 1996. The Molecular Basis of Dehydration Tolerance in Plants. Annual Review of Plant Physiology and Plant Molecular Biology, 47, 377-403.

Kang, J.Y., Choi, H.I., Im, M.Y. and Kim, S.Y. 2002. Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell, 14(2), 343-357.

Kaur, K., Gupta, A.K. and Kaur, N. 2007. Effect of water deficit on carbohydrate status and enzymes of carbohydrate metabolism in seedlings of wheat cultivars. Indian Journal of Biochemistry and Biophysics, 44(4), 223-230.

Kaur, N. and Gupta, A.K. 2005. Signal transduction pathways under abiotic stresses in plants. Current Science, 88, 1771-1780.

Kim, S.Y., Chung, H.J. and Thomas, T.L. 1997. Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the Dc3 promoter using a modified yeast one-hybrid system. Plant Journal, 11(6), 1237-1251.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology, 157(1), 105-132.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics, 23(21), 2947-2948.

Laskowski, R.A. 2001. PDBsum: summaries and analyses of PDB structures, Nucleic Acids Research, 29(1), 221-222.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U. and Kadonaga, J.T. 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes & Development, 18(13), 1606-1617.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C. and Bryant, S.H. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research, 39 (Database issue), D225-229.

Maruyama, K., Sakuma, Y., Kasuga, M., Ito, Y., Seki, M., Goda, H., Shimada, Y., Yoshida, S., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2004. Identification of cold-inducible downstream genes of the Arabidopsis DREB1A/CBF3 transcriptional factor using two microarray systems. Plant Journal, 38(6), 982-993.

Nakashima, K., Ito, Y. and Yamaguchi-Shinozaki, K. 2009. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiology, 149(1), 88-95.

Narusaka, Y., Nakashima, K., Shinwari, Z.K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2003. Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high-salinity stresses. Plant Journal, 34, 137-148.

Novikova, G.V., Moshkov, I.E. and Los, D.A. 2007. Protein sensors and transducers of cold, hyperosmotic and salt stresses in cyanobacteria and plants. Molecular Biology (Mosk), 41(3), 478-490.

Phongdara, A., Nakkaew, A. and Nualkaew, S. 2012. Isolation of the detoxification enzyme EgP450 from an oil palm EST library. Pharmaceutical Biology, 50(1), 120-127.

Reyes, J.L., Rodrigo, M.J., Colmenero-Flores, J.M., Gil, J.V., Garay-Arroyo, A., Campos, F., Salamini, F., Bartels, D. and Covarrubias, A.A. 2005. Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects *in vitro*. Plant, Cell & Environment, 28, 709-718.

Rizhsky, L., Liang, H., Mittler, R. 2002. The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiology, 130(3), 1143-1151.

Roberts, J.K., DeSimone, N.A., Lingle, W.L. and Dure, L., 3rd 1993. Cellular Concentrations and Uniformity of Cell-Type Accumulation of Two Lea Proteins in Cotton Embryos. Plant Cell, 5(7), 769-780.

Roy, A., Kucukural, A. and Zhang, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocol, 5(4), 725-738.

Savitha, M. 2000. Differential expression of LEA1 and LEA4 group of proteins in embryos and mesophytic tissues and their relevance under osmotic stress. PhD thesis, University of Agricultural Sciences, Gandhi Krishi Vignan Kendra, Bangalore, India.

Schmidt, R.J., Ketudat, M., Aukerman, M.J. and Hoschek, G. 1992. Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes Plant cell 4(6), 689-700.

Senthil-Kumar, M., Govind, G., Kang, L., Mysore, K.S. and Udayakumar, M. 2007. Functional characterization of Nicotiana benthamiana homologs of peanut water deficit-induced genes by virus-induced gene silencing. Planta, 225(3), 523-539.

Senthil-Kumar, M. and Udayakumar, M. 2006. High-throughput virus-induced gene-silencing approach to assess the functional relevance of a moisture stress-induced cDNA homologous to lea4. The Journal of Experimental Botany, 57(10), 2291-2302.

Shen, Q., Zhang, P. and Ho, T.H. 1996. Modular nature of abscisic acid (ABA) response complexes: composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley. Plant Cell, 8(7), 1107-1119.

Shih, M.D., Huang, L.T., Wei, F.J., Wu, M.T., Hoekstra, F.A. and Hsing, Y.I. 2010. OsLEA1a, a new Em-like protein of cereal plants. Plant and Cell Physiology, 51(12), 2132-2144.

Shih, M.D., Lin, S.C., Hsieh, J.S., Tsou, C.H., Chow, T.Y., Lin, T.P. and Hsing, Y.I. 2004. Gene cloning and character-

ization of a soybean (*Glycine max* L) LEA protein, GmPM16. Plant Molecular Biology, 56(5), 689-703.

Shinozaki, K. and Yamaguchi-Shinozaki, K. 2000. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Current Opinion in Plant Biology, 3(3), 217-223.

Simpson, S.D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2003. Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. Plant Journal, 33(2), 259-270.

Spiro, R.G. 2002. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology, 12, 43R-56R

Su, L., Zhao, C.Z., Bi, Y.P., Wan, S.B., Xia, H. and Wang, X.J. 2011. Isolation and expression analysis of LEA genes in peanut (*Arachis hypogaea* L). Journal of Biosciences, 36(2), 223-228.

Tran, L.S., Nakashima, K., Sakuma, Y., Osakabe, Y., Qin, F., Simpson, S.D., Maruyama, K., Fujita, Y., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2007. Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in Arabidopsis. Plant Journal, 49(1), 46-63.

Tran, L.S., Nakashima, K., Sakuma, Y., Simpson, S.D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2004. Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. Plant Cell, 16(9), 2481-2498.

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2000. Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proceedings of the National Academy of Sciences USA, 97(21), 11632-11637.

Valliyodan, B. and Nguyen, H.T. 2006. Understanding regulatory networks and engineering for enhanced drought tolerance in plants. Current Opinion in Plant Biology, 9(2), 189-195.

Wise, M.J. 2003. LEAping to conclusions: a computational reanalysis of late embryogenesis abundant proteins and their possible roles. BMC Bioinformatics, 4, 52.

Wise, M.J. and Tunnacliffe, A. 2004. POPP the question: what do LEA proteins do?. Trends in Plant Science, 9(1), 13-17.

Wong, C.K. and Bernardo, R. 2008. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theoretical and Applied Genetics, 116(6), 815-824.

Yamaguchi-Shinozaki, K. and Shinozaki, K. 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annual Review of Plant Biology, 57, 781-803.

Supplemental Table 1: EgLEA4 and other LEA protein isoforms sequences used in the amino acid alignment and phylogenetic analysis (Felsenstein, 1997).

Identifier/ abbreviations	Species	Accession number
dehydrin Brassica	*Brassica napus*	AAR23753.1
LEA3 protein Medicago	*Medicago sativa*	ADO32575.1
LEA3 Zizania	*Zizania latifolia*	ACV91270.1
LEA3 Oryza	*Oryza sativa*	A2Y720.1
AAN74638.1 LEA2 Triticum	*Triticum aestivum*	AAN74638.1
AAN74637.1 LEA1 Triticum	*Triticum aestivum*	AAN74637.1
AAN74639.1 LEA3 Triticum	*Triticum aestivum*	AAN74639.1
ABH09689.1 LEA2 Zea	*Zea mays*	ABH09689.1
ADT65200.1 dehydrin Jatropha	*Jatropha curcas*	ADT65200.1
ANN38002.1 LEA1 Capsicum	*Capsicum annuum*	AAN38002.1
AEH76905.1 LEA5 Oryza	*Oryza sativa*	AEH76905.1
ACF93400.1 LEA 2 Boea	*Boea hygrometrica*	ACF93400.1
XP_003609687.1 LEA5 Medicago	*Medicago truncatula*	XP_003609687.1
AAC06242.1 LEA 5 Nicotiana	*Nicotiana tabacum*	AAC06242.1
NP_001150746.1 LEA 4 Zea	*Zea mays*	NP_01150746
JF810461 LEA4 Elaeis	*Elaeis guineensis*	JF810461
ADU20206.1 LEA 4 Cicer	*Cicer arietinum*	ADU20206.1
ADQ91840.1 LEA 4 Arachis	*Arachis hypogaea*	ADQ91840.1
AAG37451.1 LEA 4 Glycine tomentella	*Glycine tomentella*	AAG37451.1
AAG37441.1 LEA 4 Glycine tabacina	*Glycine tabacina*	AAG37441.1
AAY26119.1 LEA4 Sisymbrium	*Sisymbrium irio*	AAY26119.1
AAT77223.1 Brassica napus	*Brassica napus*	AAT77223.1
AAT77224.1 Brassica carinata	*Brassica carinata*	AAT77224.1
AAY26119.1 LEA4 Sisymbrium	*Sisymbrium irio*	AAY26119.1
ADQ91840.1 LEA 4 Arachis	*Arachis hypogaea*	ADQ91840.1
NP_196294.1 LEA4 Arabidopsis	*Arabidopsis thaliana*	NP_196294.1
CAA63008.1 LEA D113 Arabidopsis	*Arabidopsis thaliana*	CCA63008.1

Reference

Felsenstein J (1997). An alternating least squares approach to inferring phylogenies from pairwise distances. Systematic Biology, 46(1), 101-111.