Plant Omics Journal POJ

POJ 6(6):433-440 (2013) ISSN:1836-3644

In silico comparative analysis of LEA (Late Embryogenesis Abundant) proteins in *Brachypodium distachyon* L.

Ertugrul Filiz*[1], Ibrahim Ilker Ozyigit[2], Huseyin Tombuloglu[3], Ibrahim Koc[4]

[1]Duzce University, Cilimli Vocational School, Department of Crop and Animal Production, 81750, Duzce, Turkey
[2]Marmara University, Faculty of Science and Arts, Department of Biology, 34722, Goztepe, Istanbul, Turkey
[3]Fatih University, Faculty of Science and Arts, Department of Biology, 34500, Buyukcekmece, Istanbul, Turkey
[4]Gebze Institute of Technology, Faculty of Science, Department of Molecular Biology and Genetics, 41400, Gebze, Kocaeli, Turkey

***Corresponding author: ertugrulfiliz@gmail.com**

Abstract

The Late Embryogenesis Abundant (LEA) proteins in plants are basically related with water deficiency. Recent studies showed that LEA proteins might be molecular chaperones regulating many physiological functions. In this study, LEA proteins were analyzed in model grass *Brachypodium distachyon* L. The data represented here may help to further analyze the *LEA* genes in model grass *Brachypodium* in order to understand their functions especially under conditions of water deficiency and/or other physiological mechanisms. By using the Pfam database, proteins containing at least one LEA conserved repeat (LEA2, LEA3, LEA4, LEA5, and LEA6) were classified as LEA family members. According to these results, 36 LEA proteins were identified in *B. distachyon*. LEA2 repeat was found as the dominant protein among 28 members followed by LEA3 (5 members). Physicochemical analysis showed that *pI* values and GRAVY index ranged from 4.40 to 11.1 and 0.48 to -1.423, respectively. Many LEA proteins were considered as basic character (26 members, 72.2%), while 10 proteins (27.8%) were in acidic form. Moreover, GRAVY index revealed that 19 of the 36 sequences were considered hydrophobic (52.8%) while others were hydrophilic (47.2%). Comparative phylogenetic analysis revealed that BdLEA proteins fall into eight subgroups. They were basically divided into two main groups. Chromosomal distribution of *LEA* genes was determined and segmental and tandem duplications were found in eight genes which may cause expansions of *LEA* genes through the *Brachypodium* genome. These results can be helpful for the further functional analysis of LEA proteins in *Brachypodium*.

Keywords: *Brachypodium distachyon*, LEA proteins, *in silico* analysis, phylogenetics.
Abbreviations: LEA_Late Embryogenesis Abundant; cDNA_complementary DNA; At_*Arabidopsis thaliana*; Mbp_mega base pair; Lys_Lysine; Cys_Cysteine; Trp_Tryptophan.

Introduction

Late Embryogenesis Abundant (LEA) proteins are highly hydrophilic that were first detected in cotton (*Gossypium hirsutum* L.) seeds and observed during late embryogenesis stage (Dure et al. 1981; Galau et al., 1986). It is considered that LEA proteins help plants to withstand water deficiency, which may be stimulated by freezing, drying or saline conditions (Bray, 2004; Tunnacliffe and Wise, 2007). These proteins are ubiquitous in plants and distributed across organisms such as fungi (Eichinger et al., 2005), protozoa (Katinka et al., 2001), nematodes (Gal et al., 2004), and insects (Kikawada et al., 2006).

In general, LEA proteins are composed of charged and uncharged polar amino acid residues, thus they are highly hydrophilic. Dure (1993a) studied characterization of different cDNAs from LEA proteins and supported that these proteins are of hydrophilic nature and they lack or have low levels of Cys and Trp residues. Also, they found vast amounts of Gly, Ala, Glu, Lys/Arg, and Thr residues. These results supported that LEA proteins are subset of hydrophilins. Dehydrins (dehydration induced proteins) are immensely hydrophilic proteins accumulated when water deficiency occurs during embryogenesis (Hara, 2010). The LEA 2 proteins are accepted in dehydrin proteins based on cDNA characterization (Close et al., 1989). According to amino acid sequence and conserved motifs, LEA proteins are classified in five to nine sub-classes in different species. Also, POPP (protein or oligonucleotide probability profile) shows at least four LEA groups (Wise and Tunnacliffe, 2004; Tunnacliffe and Wise, 2007).

Up to now, LEA protein families were identified in several plant and animal species. Among those, *Arabidopsis thaliana* is one of the most studied species. By using some known reference sequences, *Arabidopsis* genome was analyzed and 50 *LEA* genes were discovered. Those genes were defined in nine different groups (At*LEA* 1-9) (Bies Etheve et al., 2008). AtLEA1 proteins include wheat Em protein (Cuming, 1984) and the cotton D19 protein (Baker et al., 1988), divided into two subgroups according to the conserved sequences (Espelund et al., 1995). The AtLEA2 proteins, also called dehydrin or RAB (responsive to ABA) proteins (Close et al., 1989) which were detected in cotton embryos and defined its conserved 15-amino-acid lysine-rich sequence (Galau and Close, 1992). AtLEA3 proteins also known as D7 proteins were characterized by different consensus sequences of 11-mer repeats (Baker et al., 1988). AtLEA4 proteins, also called D113 proteins lacks of

AtLEA1, AtLEA2, and AtLEA3 sequences. AtLEA5 proteins are hydrophobic and cotton LEA D34 and D95 proteins are representatives of this group. Group 6 proteins have not been well defined and are called atypical LEA proteins (Shih et al., 2008).

Brachypodium distachyon L., also called purple false brome has recently emerged as a model organism for grass family. It offers some attractive characteristics such as small genome (~300 Mbp) (International *Brachypodium* Initiative, 2010), short life cycle (Higgins et al., 2010), and a powerful model system for many molecular based researches (Vogel and Hill, 2008). Today, some herbaceous crops (especially grasses) are crucial in obtaining renewable energy and thus model species *B. distachyon* can also be used for bio-energy production (Ozdemir et al., 2008). *Brachypodium* is considered as a bridge between rice and the Triticeae tribe (Hammami et al., 2011). Thus, genes including in *B. distachyon* genome will be helpful to understand major cereal genomes in molecular basis. In this study, *in silico* analysis were conducted to reveal LEA protein family in *B. distachyon*. As a result, 36 BdLEA proteins were analyzed. Physicochemical properties, their chromosomal locations, gene structures of *LEA* genes, and phylogenetic classifications were determined. We assume that findings of this study will be a scientific basis for comparative studies of LEA proteins for both *Brachypodium* and other grass species.

Results and Discussion

Characteristics and sequence analysis of BdLEA proteins

A total of 36 LEA protein sequences from *B. distachyon* were analyzed by using bioinformatics tools. Pfam family domains were searched for the LEA protein sequences and LEA2 (also known as dehydrin proteins) was found to be the most predominant group with 28 members (77.8%). The other LEA groups were found as LEA3 (5 members, 3.9%), LEA4 (1 member, 2.76%), LEA5 (1 member, 2.76%), and LEA6 (1 member, 2.76%), respectively (Table 1). In *A. thaliana*, LEA4 group was the most dominant group followed by LEA2, SMP, and LEA3 (Hundertmark and Hincha, 2008). Based on conserved domain analysis, the distribution of BdLEA groups was not found to be similar to that of *A. thaliana*. Analysis of *LEA* genes revealed that many of *LEA* genes (60.7%) had no introns, whereas 11 *LEA* genes contain introns with varying 1 to 2 (Fig. 1). This finding could support that Bd*LEA* genes were conserved well during *LEA* gene evolution in *Brachypodium*. According to the amino acid composition analysis of *B. distachyon*, the most abundant amino acid residues were found to be alanine (Ala), valine (Val), leucine (Leu), and glycine (Gly), respectively (Supp. Table 1). LEA2 proteins have 15-mer lysine rich conserved sequences (EKKGIMDKIKEKLPG, named K segment) in different numbers (Galau and Close, 1992). While all LEA2 proteins contain K (Lys) residues, some LEA2 proteins have higher percentage (LEA2-21, 9.82%; LEA2-27, 9.47%; LEA2-22, 8.08%) of K residues. Group 2 LEA proteins are rich in glycine or alanine and proline (Bray et al., 2000) and many LEA proteins called hydrophilins have glycine (G) content greater than 6% (Battaliga et al., 2008), thus this data is in agreement with our results (Supp. Table 1). In general, LEA proteins are highly hydrophilic, lack or have low levels of Cys and Trp residues (Dure, 1993a). In the current data set, we found that 8 of the 36 protein sequences had no W (Trp) residues. Similarly, 6 of the 36 protein sequences had no Cys residues. Physicochemical analysis revealed that most of LEA proteins (26 members, 72.2%) were considered as basic character (pI≥7), while 10 protein sequences (27.8%) were considered as acidic (pI≤7) (Table 1 and Fig. 4). Moreover, LEA4, LEA5, and LEA6 proteins were determined as acidic character. The most basic protein was found in LEA2 group (LEA2-20) while the most acidic protein was found in the LEA3 group (LEA3-2). Shih et al. (2008) found that most LEA2 proteins are neutral to basic (mean *pI*: 7.58) but this finding is not in agreement with our data in which most of the observed *pI* values (58.9%) were higher than 9. Therefore, a total of 26 LEA proteins of *B. distachyon* (22 LEA2 and 4 LEA3 members) were considered as basic character. GRAVY index showed that 19 of the 36 sequences were hydrophobic (52.8%) while 17 were hydrophilic proteins (47.2%). All LEA3 proteins (5 sequences), LEA4 protein (1 sequence), LEA5 protein (1 sequence), and LEA6 protein (1 sequence) were found to be of hydrophilic character. Additionally, 9 of the 28 LEA2 sequences (32.1%) were found to be hydrophilic. In other words, most of LEA2 proteins (67.9%) were considered as hydrophobic character. In the LEA2 members, some acidic dehydrins were used as calcium buffers, this was related to protein phosphorylation (Alsheikh et al., 2003; Hara, 2010). In the current analysis, 21.4% of the LEA2 sequences were of acidic character and this can lead to the idea that some members of *Brachypodium* LEA2 proteins may have an ion-binding activity. In general, LEA proteins are relatively small, mostly in the range of 10 to 30 kDa. Also, LEA2 group is the most hydrophobic while LEA4 and LEA5 are the most hydrophilic in *A. thaliana* (Hundertmark and Hincha, 2008). Our results were similar to those values, where molecular weights of LEA proteins were ranging from 9.25 kDa to 37.8 kDa. Also, LEA5 protein was the most hydrophilic followed by LEA6 and LEA4, respectively. Based on prediction of sub-cellular localizations of BdLEAs, 15 of the 36 sequences (38.5%) were related to organelles (mitochondrion and chloroplast) and others were connected with secretory pathway or any other locations in the cells (Table 1). Most of LEA2 proteins accumulate in cytoplasm, while some of them are localized in nucleus or in other cell compartments including mitochondria (Battaliga et al., 2008). These data support our findings in which LEA proteins show diverse sub-cellular localizations. In this study, a total of 3 different motifs were determined by using MEME tool with numbers ranging from 15 to 27 amino acids (Fig. 2 and Table 2). 28 of 36 LEA protein sequences (only LEA2 proteins) were determined as containing conserved protein motifs. Remarkably, LEA3, 4, 5, and 6 proteins had any conserved motifs. These LEA groups may be diverged recently from ancestral LEA2 proteins during LEA protein evolution. According to combined block diagram of LEA proteins, 28 of 36 LEA proteins (77.8%) contain motif I while 13 of 36 and 9 of 36 LEA proteins had motif II and motif III, respectively (Fig. 1). 8 of 36 LEA proteins had all motif types (motif I, motif II, and motif III), whereas 12 of 36 LEA proteins had one motif type (motif I). Motifs including short amino acid residues (5-25 amino acids) are very important for protein evolution and they are related with biological functions or protein structure (Saito et al., 2007). It can be suggested that motif I may be essential component of LEA domain structure, thus it showed demonstrating of stability in many of LEA sequences (77.8%) (Fig. 1).

Genomic organization and gene duplication of BdLEA genes

According to the chromosomal distribution results, *LEA* genes were found to be dispersed over all chromosomes (Fig.

Table 1. Characteristics of LEA proteins in *B. distachyon*, including accession numbers, gene location, physicochemical properties, and subcellular localizations.

LEA domains	Access. no.	Gene	Start	Stop	Seq. length (aa)	Mol. weight (kDa)	*pI*	GRAVY	Subcellular localization
LEA2-1	XP_003573252	BRADI3G13027.1	11647653	11648699	227	24.5	9.28	0.18	Unknown
LEA2-2	XP_003580846	BRADI5G26600.1	27538230	27539252	227	24.6	8.18	0.25	Unknown
LEA2-3	XP_003578883	BRADI4G41990.1	46079830	46080991	215	23.3	8.82	0.17	Unknown
LEA2-4	XP_003575086	BRADI3G45160.1	47094442	47095427	218	23.3	9.45	0.22	Mitochondrion
LEA2-5	XP_003579716	BRADI5G09670.1	12986478	12987932	221	23.4	8.83	0.27	Mitochondrion
LEA2-6	XP_003580847	BRADI5G26610.1	27554849	27555797	206	21.8	8.98	0.39	Secretory pathway
LEA2-7	XP_003573253	BRADI3G13020.1	11645945	11646574	209	23.4	9.21	0.34	Secretory pathway
LEA2-8	XP_003564678	BRADI2G55920.1	54394039	54395221	238	26.5	9.69	-0.27	Unknown
LEA2-9	XP_003571637	BRADI3G20640.1	19656521	19657261	246	26.5	7.94	0.03	Unknown
LEA2-10	XP_003571928	BRADI3G29620.1	31702318	31703626	265	28.3	10.32	0.04	Unknown
LEA2-11	XP_003564453	BRADI2G53027.1	52260517	52261624	299	31.9	9.87	0.00	Unknown
LEA2-12	XP_003577129	BRADI4G44250.1	47736121	47737186	230	24.5	9.96	0.10	Chloroplast
LEA2-13	XP_003576336	BRADI4G25970.1	31285929	31286582	217	22.4	10.04	0.32	Chloroplast
LEA2-14	XP_003559726	BRADI1G15980.1	12919102	12919837	208	22.8	8.66	0.20	Unknown
LEA2-15	XP_003568059	BRADI2G19200.1	16919537	16920388	254	26.7	9.43	0.13	Unknown
LEA2-16	XP_003567090	BRADI2G49140.1	49255066	49255976	261	28.2	5.13	-0.15	Unknown
LEA2-17	XP_003561309	BRADI1G54290.2	52676958	52680396	252	30.0	9.74	0.15	Mitochondrion
LEA2-18	XP_003560997	BRADI1G11810.1	8791659	8795795	225	24.5	9.35	0.21	Unknown
LEA2-19	XP_003565815	BRADI2G07480.1	5821262	5821816	152	16.3	5.35	-0.04	Unknown
LEA2-20	XP_003569392	BRADI2G44020.1	44506348	44507118	177	19.0	4.40	0.01	Unknown
LEA2-21	XP_003562546	BRADI1G02090.2	1419315	1421725	326	35.9	4.95	-0.41	Unknown
LEA2-22	XP_003561266	BRADI1G53000.1	51393475	51395766	390	37.3	4.93	-0.29	Mitochondrion
LEA2-23	XP_003560139	BRADI1G25960.1	21073069	21073632	187	20.1	9.14	0.48	Secretory pathway
LEA2-24	XP_003562991	BRADI1G25800.1	20911051	20911796	201	22.6	9.19	-0.08	Mitochondrion
LEA2-25	XP_003558301	BRADI1G67370.1	66006462	66007719	236	25.1	9.08	0.28	Unknown
LEA2-26	XP_003570065	BRADI3G51270.1	52359488	52363285	346	36.9	10.26	-0.14	Chloroplast
LEA2-27	XP_003567827	BRADI2G15410.1	13732581	13733743	169	18.3	4.85	-0.18	Unknown
LEA2-28	XP_003570072	BRADI3G51360.1	52409629	52412403	202	21.5	9.79	0.05	Mitochondrion
LEA3-1	XP_003566475	BRADI2G09150.1	7435407	7436193	94	92.3	10.13	-0.13	Mitochondrion
LEA3-2	XP_003572634	BRADI3G46190.1	48116389	48117470	86	93.1	11.10	-0.38	Mitochondrion
LEA3-3	XP_003561563	BRADI1G60350.1	59700337	59700988	88	98.4	9.58	-0.34	Mitochondrion
LEA3-4	XP_003566331	BRADI2G27890.1	26817041	26817417	90	97.3	6.23	-0.22	Mitochondrion
LEA3-5	XP_003567608	BRADI2G11940.1	10275789	10276597	89	92.5	9.41	-0.22	Mitochondrion
LEA4	XP_003558017	Bradi1g63816.1	63052837	6305424	357	37.8	6.45	-1.022	Mitochondrion
LEA5	XP_003568642	BRADI2G28480.1	27830766	27831217	112	16.6	6.25	-1.42	Unknown
LEA6	XP_003578310	BRADI4G33400.1	39101863	39102180	105	11.3	5.64	-1.07	Unknown

Table 2. The conserved protein motifs of *Brachypodium* LEA protein sequences.

Motif number	Width	Sequence	Protein sequences	Repeat no.
1	15		LNYTLQVTVRIHNPN	28
2	29		LQVKVDGWVRWKVGAWITGHYHLRVNCPA	9
3	27		QITVPTSLPVMYQGHRDTSVWSPVMSG	7

3). It was determined that each chromosome included at least three *LEA* genes and the number of *LEA* genes distributed to each chromosome (Chr1-5) was as 10, 11, 8, 4, and 3, respectively. The highest *LEA* gene density was determined in Chr2 and the lowest in Chr5. In general, *LEA* genes were not found residing around the centromeric regions of the chromosomes. Gene duplications were also predicted for the *LEA* genes. According to this, two segmental and two tandem duplications were determined among the *Brachypodium LEA* genes (Fig. 3). Gene duplications might be one of the major evolutionary forces for new protein functions (Kondrashov et al., 2002). Also, orthology and paralogy are key concepts in the field of protein and proteome evolution (Makarova et al., 1999). In our study, 8 of 36 *LEA* genes were predicted to be involved in the duplication event. Of those, *LEA*2-23:*LEA*2-24 and *LEA*2-26:*LEA*2-28 genes are arranged in tandem. Considering their sequence similarity, they seem to be paralog genes. At the same time, they reside on different chromosomes representing segmental duplications. Besides, *LEA*2-7:*LEA*2-1 and *LEA*2-2:*LEA*2-6 had the same situation with the previous genes showing segmental and tandem duplications.

Phylogenetic relationship of BdLEA proteins

The phylogenetic tree was constructed using 36 LEA protein sequences of *B. distachyon* and five species (*Arabidopsis thaliana* (9), *Triticum aestivum* (2), *Zea mays* (3), *Hordeum vulgare* (3), and *Oryza sativa* (3) as outgroups (Fig. 5). Two main groups were observed on the phylogenetic tree that the first main group consists of LEA2 and LEA3 groups while the second main group consists of LEA4, 5, and 6 proteins. The highest bootstrap values (62%) were observed between BdLEA2 and HvLEA2 in the subgroup A. The members of LEA3 group were quite diverse compared with other groups of LEA proteins and this diversity was created in the

Fig 1. Representation of intron and exon distribution of *LEA2* genes with their conserved motif distributions. Diagrams of the conserved motifs in grass LEA protein sequences were shown for per genes. Small boxes in different color represent the different conserved motifs. Accordingly, three most conserved motifs, stated in Table 2, were detected on LEA2 proteins. Exon and intron regions were shown as filled boxes and lines, respectively.

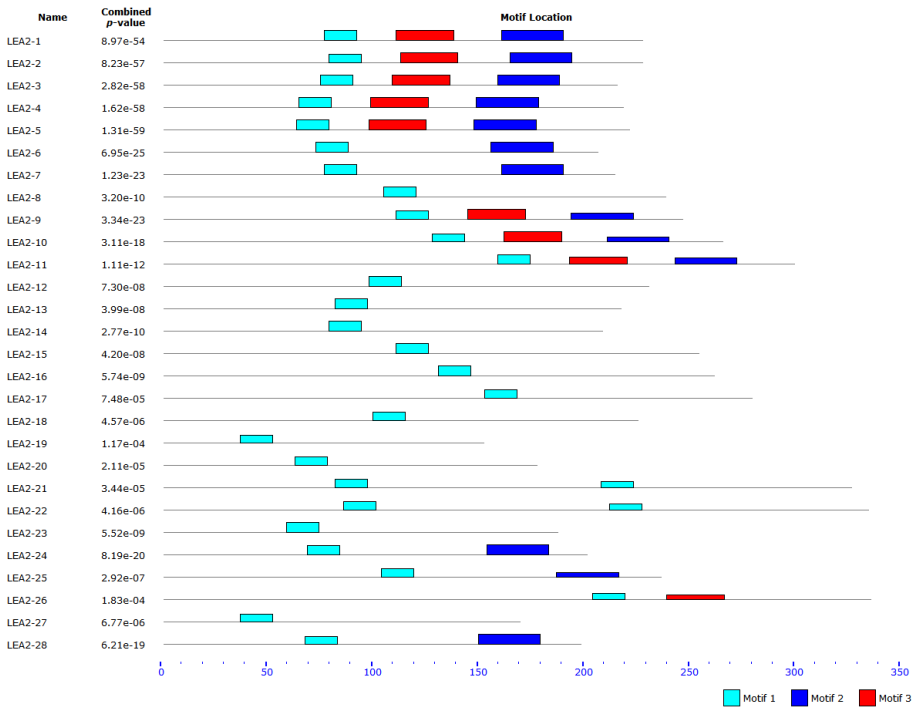

Fig 2. Combined block diagrams of the conserved protein motifs in *Brachypodium* LEA proteins using MEME server. Each motif was represented in boxes with different colors: motif 1, cyan; motif 2, blue; and motif 3, red.

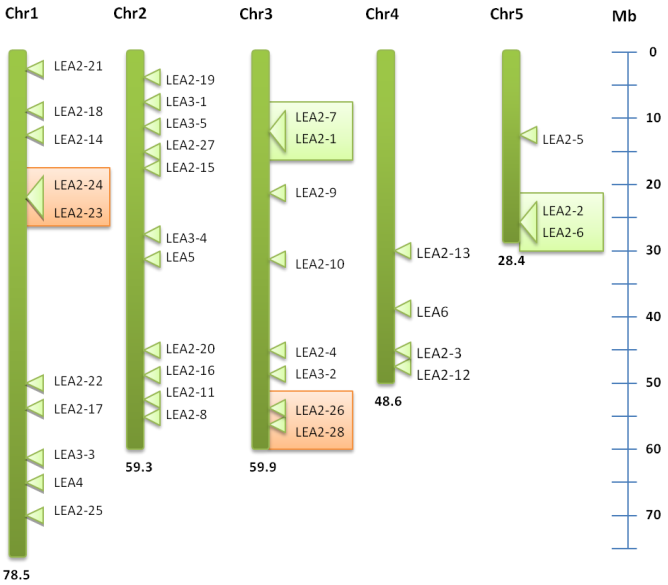

Fig 3. Chromosomal distribution of 36 *LEA* genes in *Brachypodium*. Colored boxes show tandem and segmental duplications. According to the duplication analysis, two tandem and two segmental duplications were observed separately. The size of each chromosome is indicated as Megabase (Mb).

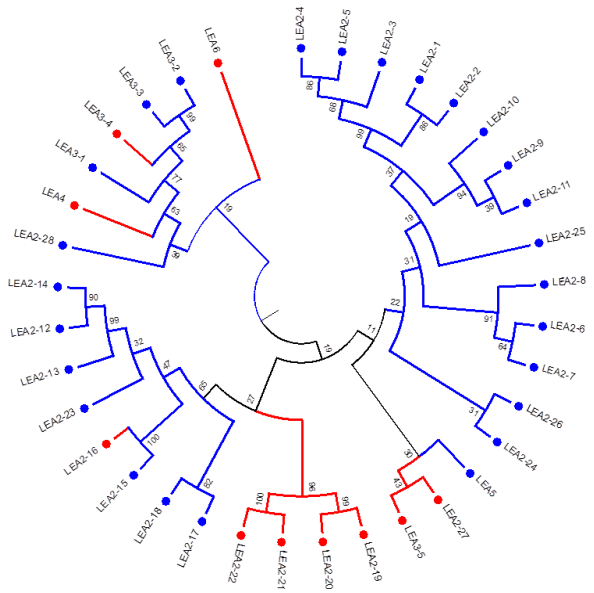

Fig 4. Phylogenetic analysis of 36 LEA proteins in *Brachypodium*. LEA proteins were classified according to their homology analysis by using NJ tree. Circles with blue and red represent the basic and acidic characterized proteins, respectively. Accordingly, 26 of 36 LEA proteins show basic character which constitutes the two main subgroups of the phylogenetic tree.

repeating 11-mer amino acid motif (Dure 1993b; Battaliga et al. 2008). In the phylogenetic tree, we found that LEA3 proteins were seperated from the other LEA groups containing LEA2, 4, 5, and 6. LEA4, LEA5, and LEA6 groups were clustered in the same clade within subgroup H. The convergence of different plant LEA groups (LEA4, LEA5, and LEA6) can be explained by possibly having similar ortholog genes that formed this clade. Subgroups B, C, D, E, F, and G consist of LEA2 proteins, whereas subgroup H contains LEA4, 5, and 6 members. Also, subgroup A includes LEA2 and LEA3 proteins, thus same LEA protein groups (LEA 2-3-4-5-6) were clustered together in joined tree (Fig. 5). The highest bootstrap values were found among LEA2 proteins including 62%, 58%, and 48%, respectivly. It could be proposed that different *LEA* gene clusters were conserved well during the diversification of monocot and dicot lineages. Previous studies indicate that LEA2 proteins accumulate during seed desiccation stimulated by drought, low temperature, or salinity (Nylander et al., 2001). Many LEA1 and LEA2 proteins have significant unstructured (loop) regions resulting four-state predictions (α-helix, β-sheet, loop or no prediction). Hence, absolute function of LEA proteins is unknown (Wise and Tunnacliffe, 2004). In this study, LEA proteins showed different clusters and dominant groups which LEA2 and LEA3 groups had 28 and 5 members, respectively. LEA proteins contain many lysine (K) residues that it may be related with specific physiological role(s) in abiotic stress (Alsheikh et al., 2003).

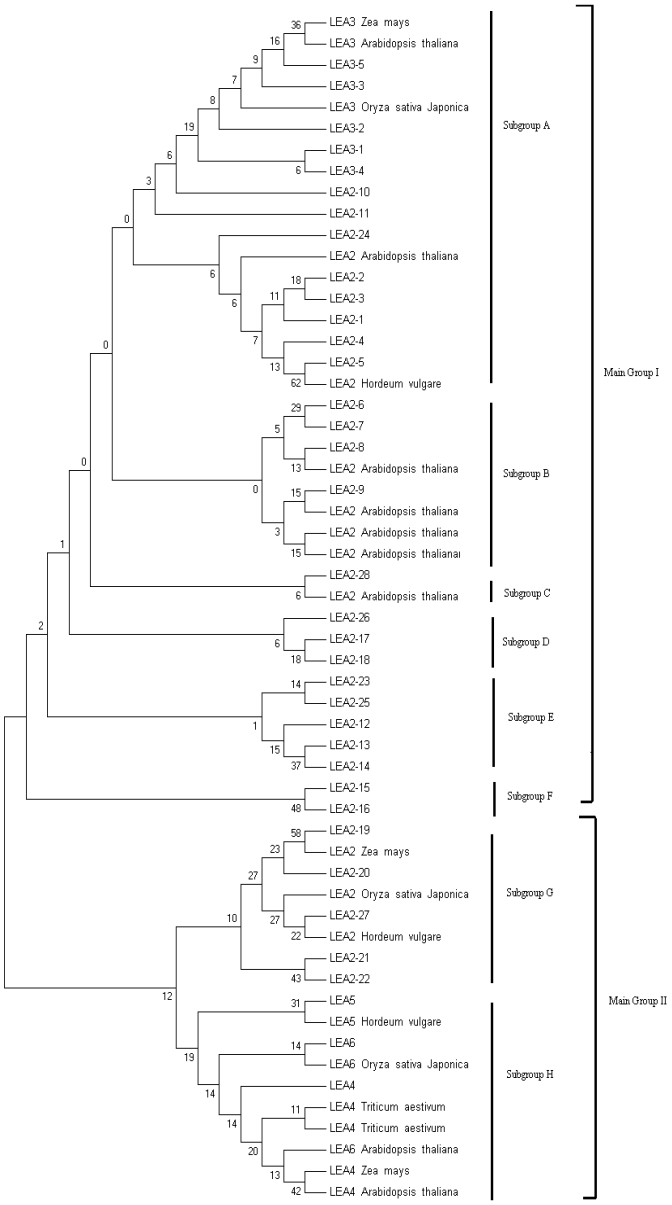

Fig 5. Comparative phylogenetic analysis of *Brachypodium*, *A. thaliana*, *H. vulgare*, *O. sativa*, and *T. aestivum* LEA proteins. Neighbor-joining (NJ) tree method was used to compute the distances of LEA proteins by using the MEGA 5.1 program. According to these results, proteins were fall into eight subgroups classified as subgroup A, B, C, D, E, F, G, and H with two main groups.

Some LEA2 subgroups showed acidic character, while the others were in basic character (Fig. 4). It can be considered that LEA2 proteins reflect various physiochemical properties. It was understood that all plant genomes have mysterious evolutionary history including gene and genome duplication (Flagel and Wendel, 2009). Also, whole genome duplications (WGD) were detected in many sequenced genomes (Semon and Wolfe, 2007). Gene duplications play a considerable role that support adaptive evolution and plant genomic architecture (Flagel and Wendel, 2009). Blanc et al. (2000) showed that *A. thaliana* genome includes long duplicated blocks (megabase-sized) and 45% of the gene pairs of chromosomes have highly similar sequences. In *A. thaliana*, it was found that 57% (653) of the pairs of more recent duplicates and 73% (306) of the pairs of older duplicates were diverged in expression (Blanc and Wolfe, 2004). In both *A. thaliana* and rice, more ancient WGD-duplicated genes showed greater expression divergence than more recent WGD-duplicated genes (Wang et al., 2011). As observed on the LEA protein sequences in *B. distachyon*, five types of LEA proteins were observed (most of them were LEA2 proteins). Also, LEA2 proteins were grouped based on their *pI* values including basic and acidic characters and support to the phylogenetic tree topologies. This result may be related to expression divergence in *LEA* genes because of duplicated genes which can affect expression divergence (Wang et al., 2012). Also, *LEA2* genes may be more ancient than the other ones. So, it may cause greater expression divergence in *Brachypodium* genome evolution.

Materials and Methods

Sequence resources of LEA proteins

A total of 36 LEA protein sequences of *B. distachyon* were retrieved from NCBI (the National Center for Biotechnology Information) protein database (http://www.ncbi.nlm.nih.gov/). LEA2 (access. no: Q03968), LEA3 (access. no: Q03968), LEA4 (access. no: Q03968), LEA5 (access. no: Q03968), and LEA6 (access. no: Q03968) proteins of *Arabidopsis* were used as query sequences to match to the candidate LEA proteins via BLASTP analysis. The sequences were selected as predicted proteins if their E value satisfied E was $\leq e^{-10}$ and redundant sequences were removed. The coding sequences, exon and intron structures of *LEA* genes were retrieved from Gramene *Brachypodium* database (http://www.gramene.org/Brachypodium_distachyon/Info/Index). LEA protein sequences were searched for their conserved domains (Pfam) and their chromosomal locations were determined by using *Brachypodium* Genome Database server (http://www.brachybase.org/gmod/genomic/contigs).

Physicochemical characterization and classification of Brachypodium LEAs

For the physicochemical characterizations, sequence length, molecular weight, theoretical isoelectric point (pI), and grand average hydropathy (GRAVY) index were computed using the Expasy's ProtParam server (Gasteiger, 2005) which calculates various physicochemical properties based on protein sequences. Kyte and Doolittle (1982) have defined the GRAVY as indication of the solubility of proteins: Positive GRAVY value means hydrophobic protein structure while negative GRAVY value means hydrophilic protein structure. Isoelectric point (pI) is a pH value where net charge of protein is zero and shows whether protein character is acidic or basic. The subcellular distribution of the LEA proteins was predicted by using TargetP 1.1 server (http://www.cbs.dtu.dk/services/TargetP/) which predicts the sub-cellular locations of eukaryotic proteins (Emanuelsson et al., 2007). The nomenclature was given according to their similarity analysis by using ClustalW program (Larkin et al., 2007).

Protein identification and phylogenetic analysis of LEA proteins

A multiple sequence alignment was done by using ClustalW (Larkin et al., 2007). Phylogenetic relationships among the LEA protein sequences was established using MEGA 5.1

(Molecular Evolutionary Genetics Analysis) software (Tamura et al., 2011) by the bootstrap analyses with 1000 replications (Felsenstein, 1985). The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT (Jones-Taylor-Thomton) matrix-based model (Jones et al., 1992). Pfam (http://www.sanger.ac.uk/software /pfam/search.html) was used for domain analysis. The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs) (Punta et al., 2012). The conserved protein motifs were deduced by using MEME (Multiple Em for Motif Elicitation) software (Timothy et al., 2009). The following parameter settings were used: distribution of motifs, zero or one per sequence; maximum number of motifs to find 3; minimum width of motif, 6; maximum width of motif, 50.

Conclusion

Computational analyses indicate that the BdLEA proteins have wide sequence diversity, physicochemical properties, distinct phylogenetic tree topology, and subcellular localizations. Also, new experimental and comprehensive analysis can support the discovery of new putative LEA proteins in other plant species. The data represented here can also be important for understanding of physiological properties and roles of LEA proteins in annotated plant genomes.

Acknowledgement

We would like to thank Miss. Berçem Yeman for her contributions to the manuscript proofreading.

References

Alsheikh MK, Heyen BJ, Randall SK (2003) Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. J Biol Chem. 278: 40882–40889

Baker J, Steel C, Dure L (1988) Sequence and characterization of 6 LEA proteins and their genes from cotton. Plant Mol Biol. 11: 277–291

Battaliga M, Olvera-Carrillo Y, Garciarrubio A, Campos F, Covarrubias AA (2008) The enigmatic LEA proteins and other hydrophilins. Plant Physiol. 148: 6–24

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

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. The Plant Cell. 12:1093–1101

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. The Plant Cell. 16: 1679–1691

Bray EA, Bailey-Serres J, Weretilnyk E (2000) Responses to Abiotic Stress 1158-1203. In Biochemistry and Molecular Biology of Plants Edited by: Bob B. Buchanan, Wilhelm Gruissem and Russell L. Jones. American Society of Plant Physiologists

Bray EA (2004) Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. J Exp Bot. 55: 2331-41

Close TJ, Kortt AA, Chandler PM (1989) A cDNA-based comparison of dehydration induced proteins (dehydrins) in barley and corn. Plant Mol Biol. 13: 95–108

Cuming AC (1984) Developmental regulation of gene expression in wheat embryos. Molecular cloning of a DNA sequence encoding the early methionine labelled (Em) polypeptide. Eur J Biochem. 145: 351–357

Dure L, Greenway SC and Galau GA (1981) Developmental biochemistry of cotton seed embryogenesis and germination: Changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. Biochemistry-Us. 20: 4162–4168

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 MD: 91–103

Dure L (1993b) A repeating 11-mer amino acid motif and plant desiccation. Plant J. 3: 363–369

Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M (2005) The genome of the social amoeba *Dictyostelium discoideum*. Nature 435: 43–57

Emanuelsson O, Brunak S, Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. Nat Protoc. 2: 953-971

Espelund M, De Bebout JA, Outlaw WH, Jacobsen KS (1995) Environmental and hormonal regulation of barley late embryogenesis abundant (Lea) mRNAs is via diVerent signal transduction pathways. Plant Cell Environ. 18: 94–949

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39(4): 783–791

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytol. 183: 557–564

Gal TZ, Glazer I, Koltai H (2004) An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress. Febs Lett. 577: 21–26

Galau GA, Hugh DW, Dure L (1986) Abscisic acid induction of cloned cotton late embryogenesis-abundant (lea) mRNAs. Plant Mol Biol. 7: 155–170

Galau GA, Close TJ (1992) Sequences of the cotton group 2 LEA/RAB/dehydrin proteins encoded by Lea3 cDNAs. Plant Physiol. 98: 1523–1525

Gasteiger E (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker ed. The Proteomics Protocols Handbook Humana Press. 571-607

Hammami R, Jouve N, Cuadrado A, Soler C, Gonzalez JM (2011) Prolamin storage proteins and alloploidy in wild populations of the small grass *Brachypodium distachyon* (L.) P. Beauv. Plant Syst Evol. 297: 99–111

Hara M (2010) The multifunctionality of dehydrins. Plant Signal Behav. 5(5): 503-508

Higgins JA, Bailey PC, Laurie DA (2010) Comparative genomics of flowering time pathways using *Brachypodium distachyon* as a model for the temperate grasses. Plos One. 5(4): e10065

Hundertmark M, Hincha DK (2008) LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics 9: 118

International *Brachypodium* Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463: 763–768

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8: 275-282

Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, Alaoui H (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414: 450–453

Kikawada T, Nakahara H, Kanamori Y, Iwata K., Watanabe M., McGee B, Tunnacli A, Okuda T (2006) Dehydration induced expression of LEA proteins in an anhydrobiotic chironomid. Biochem Bioph Res Co. 348: 56–61

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV (2002) Selection in the evolution of gene duplications. Genome Biol. 3: 1–9

Kyte J, RF Doolittle (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol. 157: 105-132

Larkin MA, Blackshields G, Brow NP, Chenna R, McGettigan PA, McWilliam H (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21): 2947–2948

Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV (1999) Comparative genomics of the *archaea* (*euryarchaeota*): evolution of conserved protein families, the stable core, and the variable shell. Genome Res. 9: 608–28

Nylander M, Svensson J, Palva ET, Welin BV (2001) Stress-induced accumulation and tissue-specific localization of dehydrins in *Arabidopsis thaliana*. Plant Mol Biol. 45: 263–279

Ozdemir BS, Hernandez P, Filiz Eü Budak H (2008) *Brachypodium* genomics. Int J Plant Genomics 5:36

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J and Boursnell C (2012) Nucleic Acids Res Database Issue 40: 290-301

Saito H, Kashida S, Inoue T, Shiba K (2007) The role of peptide motifs in the evolution of a protein network. Nucleic Acids Research 19: 6357–6366

Semon M, Wolfe KH (2007) Consequences of genome duplication. Curr Opin Genet Dev. 17:505–512

Shih MD, Hoekstra FA, Hsing YC (2008) Late Embryogenesis Abundant Proteins. Adv Bot Res. 48

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 28: 2731-2739

Timothy L, Bodén MB, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37: 202-208

Tunnacliffe A, Wise MJ (2007) The continuing conundrum of the LEA proteins. Naturwissenschaften. 94: 791–812

Vogel J, Hill T (2008) High-efficiency *Agrobacterium*-mediated transformation of *Brachypodium distachyon* inbred line Bd21–3. Plant Cell Rep. 27: 471–478

Wang Y, Wang X, Tang H, Tan X, Ficklin SP (2011) Modes of Gene Duplication Contribute Differently to Genetic Novelty and Redundancy, but Show Parallels across Divergent Angiosperms. PLoS ONE 6(12): e28150

Wang Y, Wang X, Paterson AH (2012) Genome and gene duplications and gene expression divergence: a view from plants. Ann NY Acad Sci. 1256: 1–14

Wise MJ, Tunnacliffe A (2004) POPP the question: what do LEA proteins do? Trends Plant Sci. 9: 13-17