J Genet Resour2021;7(2): 174-187 *Homepage: http://sc.journals.umz.ac.ir/*

RESEARCH ARTICLE *DOI: 10.22080/jgr.2021.20336.1229*

J|G|R

Phylogenetic and Evolutionary Analysis of the Late Embryogenesis Abundant (LEA) Gene Product in Poaceae

Darush Choobineh[1], Nafiseh Mahdinezhad[*1], Ali Niazi[2], Baratali Fakheri[1] and Abbasali Emamjomeh[1]

[1] Department of Plant Breeding and Biotechnology, College of Agriculture, University of Zabol, Zabol, Iran
[2] Institute of Biotechnology, College of Agriculture, Shiraz University, Shiraz, Iran

ARTICLE INFO

Article history:
Received 25 November 2020
Accepted 07 March 2021
Available online 01 June 2021

Keywords:
Abundant (LEA) protein family
Evolutionary analysis
Late Embryogenesis
Phylogenetic analysis
Poaceae

Supplementary information:
Supplementary information for this article are available at *http://sc.journals.umz.ac.ir/*

Corresponding authors:
✉ N. Mahdinezhad
Nmahdinezhad@uoz.ac.ir

p-ISSN 2423-4257
e-ISSN 2588-2589

ABSTRACT

The late Embryogenesis Abundant (LEA) protein family obtains a group of stress-induced hyper-hydrophilic proteins that accumulate in response to cellular dehydration. They are generally unstructured polypeptides without a well-defined three-dimensional structure and have been identified in a wide range of organisms from bacteria to higher plants. Herein, we made a phylogenetic and evolutionary analysis for LEA proteins in Poaceae. The full-length LEA protein sequences were acquired by performing the sequence search of sequenced *hva1* against Poaceae species in the non-redundant protein database by a BlastX search tool. The sequences were aligned with the Clustal Omega tool. The MEME suite searched for conserved blocks among each LEA protein sequence. Also, the evolutionary relationship among the LEA protein sequences evaluates using the MEGA tool. The results display close sequence similarity not only into the species but also between species. The results demonstrated that LEA proteins cluster into two large subgroups. The overall average evolutionary difference in LEA protein sequence pairs estimated as 0.4022 amino acid substitutions per site from averaging over all sequence pairs. The LEA protein sequences contain a significant percentage of glycine residues but lack cysteine and tryptophan residues. The results indicate the occurrence of homologs in the subgroup before the divergence of the species. However, the expansion of the gene number in the Poaceae was approved by the duplication events in the preexisting genes rather than by the appearance of the altered LEA gene. Our data will provide novel insights for further studies of the Late Embryogenesis Abundant protein family in Poaceae.

Please cite this paper as: Choobineh D, Mahdinezhad N, Niazi A, Fakheri B, Emamjomeh A. 2021. Phylogenetic and evolutionary analysis of the late embryogenesis abundant (lEA) gene product in Poaceae. *J Genet Resour* 7(2): 174-187. doi: 10.22080/jgr.2021.20336.1229.

Introduction

Late Embryogenesis Abundant (LEA) protein is a hyper-hydrophilic protein family. It characterizes by hydrophilic and glycine-rich domains and lacks or has reduced possible content of tryptophan and cysteine residues (Stacy and Aalen, 1998). As its name suggests, for the first, LEA proteins are known as accumulating in high levels in the late stages of embryo development in cotton (*Gossypium hirsutum*) seeds (Dure *et al.,* 1981). Then, it realized that similar gene product(s) found in other plant species (Altunoglu *et al.,* 2016; Gao *et al.,* 2016; Liang *et al.,* 2016; Altunoglu *et al.,* 2017). They were found not only in the reproductive tissues but also in the vegetative ones under abiotic stresses. They lead to dehydration conditions, including cold, drought, and high salinity (Ingram *et al.,* 1996; Bray, 1997; Thomashow, 1998). LEA proteins have low molecular weight (10-30 kDa) and are generally unstructured polypeptides without a well-defined three-dimensional structure. They contain a high percentage of charged amino acid

residues (Hong-Bo *et al.,* 2005; Cao and Li, 2015) with a propensity for alpha-helix formation (Rorat, 2006) in the native state. The structure-function relationship of LEA protein remains challenging. In higher plants, LEA proteins involve in tolerance to freezing and desiccation of environmental induced stresses. However, the LEA protein's precise function(s) is unknown (Wise, 2003).

LEA proteins have been identified in a wide range of organisms from microorganisms such as bacteria (Garay-Arroyo *et al.,* 2000), to nematodes (Solomon *et al.,* 2000; Browne *et al.,* 2004), archaea (Campos *et al.,* 2013), and fungi (Abba *et al.,* 2006) and even to metazoans. In plants, the LEA protein family appears in ubiquitous members. They are found not only in angiosperms and gymnosperms (Bray, 1997; Cuming, 1999; Ibrahim *et al.,* 2019) but also in seedless vascular plants (Oliver *et al.,* 2000; Alpert, 2005; Iturriaga *et al.,* 2006) and bryophytes (Alpert *et al.,* 2002; Oliver *et al.,* 2004; Saavedra *et al.,* 2006; Proctor *et al.,* 2007), pteridophytes (Reynolds *et al.,* 1993), and algae (Tanaka *et al.,* 2004). LEA proteins have been classified into eight groups based on the amino acid sequence similarity and by the presence of sequence motifs (Bray, 1997; Wise, 2003). Recently, 12 LEA protein groups were characterized based on their motifs and a unique set of physicochemical properties (Jaspard *et al.,* 2012). However, the Pfam database defined LEA proteins into eight diverse sub-families (Hunault *et al.,* 2010). It speculates that each LEA protein group has specific functions during dehydration and response to stresses (Dure *et al.,* 1989; Dure, 1993; Ibrahim *et al.,* 2019).

Poaceae contains many pivotal monocotyledonous flowering plants. Its species are the most important and economically family for humans agriculturally, economically, and ecologically. In this study, in the Poaceae species, LEA protein sequence grouping and similarities were analyzed. Next, a sequence composition, specific signature motifs, substitution rate estimation, and divergence analysis were analyzed using the MEGA tool. Additionally, the gene evolutionarily history, its duplication event was studied by using the LEA protein sequences in the Poaceae. We suggested a model of the evolutionary history for LEA proteins in the Poaceae species.

Materials and Methods

Plant material and growing condition

In this study, *Hordeum vulgare* var. Zehak was used. The seeds were obtained from the Wheat and Barely Research Department of Agricultural research center, Zabol, Iran. Five sterile seeds were sown in a complex mixture of sandy and clay culture (1:1) in the pot separately under field conditions. Then before seedling emergence, they watered by tap twice a week. For gene cloning, the barely seedlings 20-days old leaves have been collected carefully for successful DNA extraction. Appropriate samples from numerous cultivars were carefully collected, frozen immediately in liquid nitrogen, and stored at -80°C for further analysis.

DNA extraction, polymerase chain reaction, and sequencing

Total genomic DNA was extracted from the barely samples based on Dellaporta *et al.* (1993) extraction method graciously according to published instructions with some necessary modification. The extracted genomic DNA was eluted in 20 µl of ddH_2O at 70 °C and then subjected to a polymerase chain reaction (PCR) or stored at -20 °C until PCR amplification. For the amplification of *hva1* total DNA was used. The *hva1* genomic DNA coding sequence amplified using specifically designed primer pairs by Abdul Kader *et al*. (2012). The designed primer pairs (Forward: 5′-TGGCCTCCAACCAGAACCAG-3′ and Reverse: 5′-ACGACTAAAGGAACGGAAAT-3′) were checked by OligoAnalyser 3.1 program. The PCR conditions for the gene amplification set according to the primer pair properties and the PCR product quality.

The PCR reaction was conducted in a concluding volume of 50 µl reaction mixture containing precisely 10 mM, Tris-HCL, pH 8.4; 50 mM KCl; 1.5 mM $MgCl_2$, 200 µM, dNTPs; 20 pmol of each forward and reverse primer, 2U *Taq*-polymerase, 2.5 µl of DNA extracted as standard template and ddH_2O. In the negative control, 2.5 µl ddH_2O was added instead of the DNA template to the PCR reaction mixture.

The tubes were placed in a palm thermal cycler model GP001, Correbett research, Australia. The thermal cycler is programmed as follows: setting an initial denaturation at 95°C for 5 min, 30 amplification cycles (denaturalizing at 94°C for 45 s, annealing at 55 °C for the 30 sec and extension at 72 °C for 20 sec) followed by a final extension of 5 min at 72 °C. For assessing the excellent quality of the PCR amplification, eight µl of PCR products were visualized with direct Ultra Violet illumination after electrophoresis and possible separation in 1.5% TAE-buffered agarose gels and staining with ethidium bromide. Then excised from agarose gels traditionally using the Qiaquick Gel Extraction Kit (Qiagen Benelux B.V., the Netherlands), and carefully stored at -20 °C for further analysis. The experiments were replicated three considerable times. For sequencing the specific products in both forward and reverse directions, three samples from each purified product sent to TAKAPOOZIST Co. Ltd. The same nucleotide sequences (independent replications) were subsequently examined and compared by a global sequence alignment tool; Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo) for merely ensuring the sample sequencing accuracy.

Sequence retrieval, alignment, and analysis

The protein sequence entries for all completely LEA protein sequences available to date for Poaceae species in FASTA format from National Center for Biotechnology Information protein database (http://www.ncbi.nlm.nih.gov/guide/proteins/) were acquired by BlastX performance of the *hva1* products against Poaceae species in the non-redundant protein database. Then, to eliminate the redundant peptides and repeated sequences in each species from Poaceae, the Clustal Omega (available at http://www.ebi.ac.uk/Tools/msa/clustalo/; Madeira *et al*., 2019) and CodonCode aligner V.8.0.2 (http://www.codoncode.com/index.htm) were applied. The Clustal Omega tool constructs multiple sequence alignments of full-length LEA protein sequences. The generated alignment file was visualized and analyzed by Jalview 2.10.3 software (Waterhouse *et al*., 2009). In the aligned LEA protein sequences, the MEGA 7.0.26 tool applied to calculating the amino acid frequencies, the best amino acid substitution pattern, and the model with the lowest Bayesian Information Criterion (BIC score) toward the amino acid substitution matrix were considered. We also used the MEGA to estimate the evolutionary difference between each pair of the sequence using the best fitting amino acid substitution model. Additionally, the MEGA was applied to conduct the number of amino acid substitutions per site from each pair of sequences and overall sequence pairs by the selected substitution model.

Motif prediction

We also used the MEME (Multiple Em for Motif Elicitation)-suite (http://meme-suite.org/; Bailey *et al.,* 2009) to search relatively short sequences of conserved amino acids (called blocks) among each LEA protein sequence under the following criteria.1). Each motif site occurs zero or one. 2) Optimum motif length is set from six to 50 amino acids, and 3) the maximum number of motifs in each is five.

Phylogenetic and evolutionary analysis

For the determination of the evolutionary relationship among the LEA protein sequences, Neighbor-joining (NJ) phylogenetic analysis was constructed with high supporting bootstrap values (500 replications) with the help of the MEGA 7.0.26 tool using the best-fitting amino acid substitution model into the LEA protein sequences. The maximum likelihood of the amino acid substitution pattern of the LEA protein sequences for 56 amino acid substitution models evaluated using the MEGA 7.0.26 tool (Kumar *et al.,* 2016). Also, to identify the gene duplications, all branching points in the species tree were searched using the MEGA 7.0.26 tool with at least one species that present in both subtrees of the branching tree in such a way that the minimum number of duplication events met. Besides, for each amino acid residue alongside the rest in the aligned protein sequences, the mutation data matrix (relative mean evolution rate) was calculated. For understanding the origin of the protein, the ancestral protein sequence at each ancestral node was reconstructed by the maximum likelihood method and the utmost parsimony method

separately under the best-fitting amino acid substitution model by the MEGA 7.0.26 tool. Also, the LEA protein divergence time was estimated by the molecular clock test by helping the selected substitution model using the MEGA 7.0.26 tool. At a 5% significant level, the null hypothesis of equal evolutionary rates through the tree was rejected. Moreover, the LEA protein divergence time was estimated by the molecular clock test by helping the selected substitution model using the MEGA 7.0.26 tool. At a 5% significant level, the null hypothesis of equal evolutionary rates through the tree was rejected.

Results

The PCR products of the barely genome using the specific *hva1* primers to amplify the *Hva1* gene, visualized on electrophoresis gel showed the target band of about 800 bp (Fig. 1).

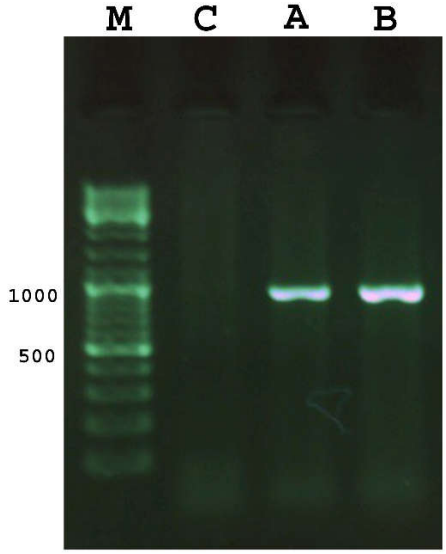

Fig. 1. PCR amplification of the barely genome using the specific *hva1* primers to amplify the *Hva1* gene: Lane A, B corresponds with the *hva1* products isolated from barely. Lane C corresponds to the negative control. Lane M corresponds with a 100 bp ladder DNA size marker.

The sequencing result shows 678 nucleotides in its length. This amplicon shows similarity to the sequence of accession number: GU108379.1 registered as *hva1* gene (Fig. 2).

By performing the sequence search of *hva1* designed primer products against Poaceae species in the non-redundant protein database by BlastX, 100 full-lengths or near full-length LEA protein sequences with annotated Late Embryogenesis Abundant Protein from 26 species were considered after elimination of the redundant peptides, repeated sequence in each species, and used for further analysis (Table 1; Supplementary 1).

>Amplified sequence 754 bp
GCCTCCAACCAGAACCAGGGGAGCTACCACGCCGGCGAGACCAAGGCCCGCAC
CGAGGTGACCGTCGTCTCCTTGGTGTCTGTCTATACTCTGCCTGCCGCGCGCAT
GCGGCGTTGCTCCGGCGGTGATCTGATATGTTCTTTTGTATCTGTTGGATGAGT
TGCAGGAGAAGACCGGGCAGATGATGGGCGCCACCAAGGACAAGGCCGGGGCA
GACCACCGAGGCCACCAAGCAGAAGGCCGGCGAGACGGCCGAGGCCGCCAAG
CAGAAGGCCGCCGAGGCCAAGGACAAGACGGCGCAGACGGCGCAGGCGGCCA
AGGACAAGACGTACGAGACGGCGCAGGCGGCCAAGGAGCGCGCCGCCCAGGG
CAAGGACCAGACCGGCAGCACCCTCGGCGAGAAGACGGAGGCGGCCAAGCAG
AAGGCCGCCGAGACAACGGAGGCGGCCAAGCAGAAGGCCGCCGAGGCAACCG
AGGCGGCCAAGCAGAAGGCGTCCGACACGGCGCAGTACACCAAGGAGTCCGC
GGTGGCCGGCAAGGACAAGACCGGCAGCGTCCTCCAGCAGGCCGGCGAGACG
GTGGTGAACGCCGTGGTGGGCGCCAAGGACGCCGTGGCAAACACGCTGGGCA

Fig. 2. The nucleotide sequence of amplified regions using the designed *hva1* forward and reverse primers shows similarity with the sequences registered in NCBI for the *hva1* gene.

Table 1. The number of LEA protein sequences from the species were analyzed in this study

Species	# of LEA protein sequences
Aegilops tauschii	9
Agropyron cristatum	1
Agropyron mongolicum	1
Ampeloca lamuscalcareus	1
Brachypodium distachyon	7
Bromus inermis	3
Deschampsia antarctica	1
Dichanthe liumoligosanthes	1
Eremopyrum triticeum	1
Festuca arundinacea	3
Hordeum vulgare	11
Oryza brachyantha	2
Oryza nivara	1
Oryza sativa	10
Panicum hallii	2
Pogona therumpaniceum	6
Setaria italica	2
Sorghum arundinaceum	1
Sorghum bicolor	6
Sorghum halepense,	1
Sporobolus stapfianus	1
Triticum aestivum	17
Triticum turgidum	2
Triticum urartu	7
Zea mays	2
Zizani alatifolia	1

The relationships between the Poaceae species, their divergence times, and their evolutionary timescale are available in the timetree (freely

available at http://www.timetree.org/) database (Supplementary 2). LEA protein sequences were searched for matching Pfam families (https://pfam.xfam.org/search#tabview=tab1), and found that 78 out of 100 LEA proteins belong to the LEA-4 family as defined by the Pfam database. The HMMER web server (www.ebi.ac.uk/Tools/hmmer/search/phmmer) of Finn *et al*., (2015) used for sensitive searches of the other orthologues to confirm their groups. Consequently, we found that 15 LEA proteins are similar to the LEA-3 group, 5 LEA proteins are analogous to the LEA-14 group, one LEA protein sequence is equivalent to the LEA-19 group, and one homogenous to LEA proteins, but its group is remained unknown (Table 2).

Table 2: The characteristics of the LEA proteins from Poaceae species.

Protein/ species	LEA protein grouping		*E*- valuate
XP 020173182.1/*Aegilops tauschii tauschii*	LEA-4	defined by Pfam database	-
Q03968.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
ALD18913.1/*Triticum turgidum durum*	LEA-4	defined by Pfam database	-
AAN74637.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
ACH89913.1/*Hordeum vulgare vulgar*	LEA-4	defined by Pfam database	-
AEJ88291.1/*Agropyron cristatum*	LEA-4	defined by Pfam database	-
ALD18912.1/*Triticum turgidum durum*	LEA-4	defined by Pfam database	-
AKC92683.1/*Hordeum vulgare*	LEA-4	defined by Pfam database	-
ACH89911.1/*Hordeum vulgare vulgare*	LEA-4	defined by Pfam database	-
AAN74639.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
P14928.1/*Hordeum vulgare vulgare*	LEA-4	defined by Pfam database	-
ACH89910.1/*Hordeum vulgare vulgare*	LEA-4	defined by Pfam database	-
ANB44749.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
APJ36638.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
AIZ11400.1/*Eremopyrum triticeum*	LEA-4	defined by Pfam database	-
BAF79928.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
AQM73730.1/*Festuca arundinacea*	LEA-4	defined by Pfam database	-
AAD33850.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
AEY78063.1/*Deschampsia Antarctica*	LEA-4	defined by Pfam database	-
XP 003567984.1/*Brachypodium distachyon*	LEA-4	defined by Pfam database	-
BAJ99283.1/*Hordeum vulgare vulgare*	LEA-4	defined by Pfam database	-
AHZ35571.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
AQM73731.1/*Festuca arundinacea*	LEA-4	defined by Pfam database	-
XP 014754957.1/*Brachypodium distachyon*	LEA-4	defined by Pfam database	-
XP 020200916.1/*Aegilops tauschii tauschii*	LEA-4	defined by Pfam database	-
AQM73729.1/*Festuca arundinacea*	LEA-4	defined by Pfam database	-
BAC80266.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
BAF79927.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
XP 020173192.1/*Aegilops tauschii tauschii*	LEA-4	defined by Pfam database	-
BAD22767.1/*Bromus inermis*	LEA-4	defined by Pfam database	-
AAN74638.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
XP 014754956.1/*Brachypodium distachyon*	LEA-4	defined by Pfam database	-
AHZ35570.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
ADC55280.1/*Ampeloca lamuscalcareus*	LEA-4	defined by Pfam database	-
XP 020200902.1/*Aegilops tauschii tauschii*	LEA-4	defined by Pfam database	-
XP 006654728.1/*Oryza brachyantha*	LEA-4	defined by Pfam database	-
XP 014754955.1/*Brachypodium distachyon*	LEA-4	defined by Pfam database	-
XP 014754954.1/*Brachypodium distachyon*	LEA-4	defined by Pfam database	-
XP 021303222.1/*Sorghum bicolor*	LEA-4	defined by Pfam database	-
AFO85387.1/*Sorghum bicolor*	LEA-4	defined by Pfam database	-
EMS54158.1/*Triticum urartu*	LEA-4	defined by Pfam database	-
ACV91271.1/*Sorghum bicolor*	LEA-4	defined by Pfam database	-
EMS46889.1/*Triticum urartu*	LEA-4	defined by Pfam database	-
KXG22444.1/*Sorghum bicolor*	LEA-4	defined by Pfam database	-
XP 004961353.1/*Setari italica*	LEA-4	defined by Pfam database	-
AAD02421.1/*Oryza sativa*	LEA-4	defined by Pfam database	-
XP 015637572.1/*Oryza sativa Japonica*	LEA-4	defined by Pfam database	-
A2Y720.1/*Oryza sativa Indica*	LEA-4	defined by Pfam database	-
ABS44867.1/*Oryza sativa Japonica*	LEA-4	defined by Pfam database	-

Table 2: Continued

Protein/ species	LEA protein grouping		*E* valuate
AFP33238.1/*Sorghum halepense*	LEA-4	defined by Pfam database	-
BAK03304.1/*Hordeum vulgare vulgare*	LEA-4	defined by Pfam database	-
CAA92106.1/*Oryza sativa Indica*	LEA-4	defined by Pfam database	-
BAD22766.1/*Bromus inermis*	LEA-4	defined by Pfam database	-
AFO85389.1/*Sorghum bicolor*	LEA-4	defined by Pfam database	-
BAJ89497.1/*Hordeum vulgare vulgare*	LEA-4	defined by Pfam database	-
PAN18358.1/*Panicum hallii*	LEA-4	defined by Pfam database	-
Q42376.1/*Zea mays*	LEA-4	defined by Pfam database	-
ACH69969.1/*Agropyron mongolicum*	LEA-4	defined by Pfam database	-
ALM25914.1/*Sorghum arundinaceum*	LEA-4	defined by Pfam database	-
OEL32771.1/*Dichanthe liumoligosanthes*	LEA-4	defined by Pfam database	-
ADF36679.1/*Pogonatherum paniceum*	LEA-4	defined by Pfam database	-
ADF36676.1/*Pogona therumpaniceum*	LEA-4	defined by Pfam database	-
APU66353.1/*Oryza sativa Indica*	LEA-4	defined by Pfam database	-
BAA05537.1/*Oryza sativa Japonica*	LEA-4	defined by Pfam database	-
APU66352.1/*Oryza nivara*	LEA-4	defined by Pfam database	-
XP 015613488.1/*Oryza sativa Indica*	LEA-4	defined by Pfam database	-
XP 006644582.1/*Oryza brachyantha*	LEA-4	defined by Pfam database	-
XP 020179481.1/*Aegilops tauschii tauschii*	LEA-4	defined by Pfam database	-
ADF36678.1/*Pogona therumpaniceum*	LEA-4	defined by Pfam database	-
XP 004969685.1/*Setaria italica*	LEA-4	defined by Pfam database	-
XP 003569641.1/*Brachypodium distachyon*	LEA-4	defined by Pfam database	-
PAN28951.1/*Panicum hallii*	LEA-4	defined by Pfam database	-
EMS66582.1/*Triticum urartu*	LEA-4	defined by Pfam database	-
AMW07364.1/*Triticum aestivum*	LEA-4	defined by Pfam database	-
EEE64550.1/*Oryza sativa Japonica*	LEA-4	defined by Pfam database	-
BAJ85410.1/*Hordeum vulgare vulgare*	LEA-4	defined by Pfam database	-
CAA71751.1/*Sporobolus stapfianus*	LEA-4	defined by Pfam database	-
BAS95139.1/*Oryza sativa Japonica*	LEA-4	defined by Pfam database	-
XP 020173203.1/*Aegilops tauschii tauschii*	LEA-3	Similar LEA3 WHEAT protein	4.30E-52
EMS46890.1/*Triticum urartu*	LEA-3	Similar to LEA3 WHEAT protein	2.20E-82
NP 001146945.1/*Zea mays*	LEA-3	Similar to LEA3 MAIZE protein	8.00E-100
EMS46888.1/*Triticum urartu*	LEA-3	Similar to LEA3 WHEAT protein	3.90E-48
ADF36681.1/*Pogona therumpaniceum*	LEA-3	Similar to J7FJG8 SORBI protein	2.10E-69
EMS53654.1/*Triticum urartu*	LEA-3	Similar to LEA3 WHEAT protein	7.80E-31
CAA55976.1/*Hordeum vulgare*	LEA-3	Similar to A7VL25 WHEAT protein	8.50E-88
BAF79926.1/*Triticum aestivum*	LEA-3	Similar to A7VL25 WHEAT protein	7.10E-102
XP 021313237.1/*Sorghum bicolor*	LEA-3	Similar to B6UI06 MAIZE protein	2.00E-99
AAB18208.1/*Triticum urartu*	LEA-3	Similar to A7VL25 WHEAT protein	7.70E-53
XP 020150225.1/*Aegilops tauschii tauschii*	LEA-3	Similar to A7VL25 WHEAT protein	1.80E-53
ADF36680.1/*Pogona therumpaniceum*	LEA-3	Similar to J7FJG8 SORBI protein	1.30E-59
CAA03925.1/*Hordeum vulgare vulgare*	LEA-3	Similar to A7VL25 WHEAT protein	4.10E-64
ADF36677.1/*Pogona therumpaniceum*	LEA-3	Similar to J7FJG8 SORBI protein	3.60E-59
XP 020150224.1/*Aegilops tauschii tauschii*	LEA-3	Similar to A7VL25 WHEAT protein	1.10E-78
XP 020200903.1/*Aegilops tauschii tauschii*	LEA-19	Similar to LEA19 ORYSJ protein	7.60E-55
ACV91270.1/*Zizani alatifolia*	LEA-14	Similar to LEA14 ORYSJ protein	1.40E-59
XP 020174962.1/*Triticum aestivum*	LEA-14	Similar to LEA14 ORYSJ protein	2.30E-70
AKG51641.1/*Triticum aestivum*	LEA-14	Similar to LEA14 ORYSJ protein	6.50E-69
BAN15016.1/*Bromus inermis*	LEA-14	Similar to LEA14 ORYSJ protein	2.60E-63
AKG51643.1/*Triticum aestivum*	LEA-14	Similar to LEA14 ORYSJ protein	6.00E-65
KQK05114.1/*Brachypodium distachyon*	LEA	Similar to A0A024CKY0 WHEAT protein	3.20E-24

Motif analysis of LEA protein sequences revealed the most conserved five motifs conserved in all proteins (Supplementary 3). Motif-1 presents in 79 LEA protein sequences, motif-2 presents in 74 proteins, motif-3 presents in 76 proteins, motif-4 presents in all, and motif-5 presents in 89 proteins (Supplementary 3). Also, members within the same LEA protein group shared several group-specific conserved protein motifs. Moreover, conserved motif analysis suggested the origination of the LEA proteins from the gene duplication-expansion

scenario in the Poaceae. Whole-genome duplication, tandem duplications, and segmental duplications of protein-coding genes play an essential role in gene dispersal and diversification (see Singh *et al.,* 2019).

A multiple sequence alignment of all the identified LEA protein sequences carried out using Clustal Omega software (http://www.ebi.ac.uk/Tools/msa/clustalo/) with the default parameters. The conserved amino acid domains were appeared not only in the species but also between species (Supplementary 4). Identical columns and identical residues in the same column colored as blue and blue-shaded. Consequently, the lowest BIC score (3976.736472) between the amino acid substitution belongs to the Jones-Taylor-Thornton (JTT) model (Supplementary 5). Therefore, in this study, we use the Jones-Taylor-Thornton (JTT) amino acid substitution model in our phylogenetic and evolutionary analysis. We also computed the probability of substitution from everyone's amino acid to another one. Substitution patterns and rates were estimated using the Jones-Taylor-Thornton (1992) model. We extract the amino acid substitution matrix from the LEA protein sequences (Supplementary 6) for further homology searches and find the other homologous sequences. Additionally, average evolutionary divergence over all LEA protein sequence pairs estimated as 0.4022 amino acid substitutions per site from averaging over all sequence pairs (from 0 up to 0.7303 amino acid differences per site). The amino acid composition in each LEA protein sequence and the average of each amino acid composition among the LEA protein sequences were also computed (Supplementary 7). In the LEA protein sequences, the results showing a high percentage of Gly residues (9.328%), lack Cys and Trp residues (0.1495 and 0.1089%). For revealing the evolutionary history and phylogenetic relationship of the LEA protein sequences, the evolutionary history deduced using the JTT matrix-based method (Jones *et al.,* 1992). The JTT matrix-based method is the units of the number of amino acid substitutions per site in the Clustal Omega aligned sequences using the Neighbor-Joining method (Saitou *et al.,* 1987), calculated by the MEGA 7.0.26 tool (Kumar *et al.,* 2016; Fig. 3).

To improve the quality of phylogenetic analysis bootstrapping test was done up to 500 replications. However, due to long sequence similarities, some nodes of the constructed tree have smaller bootstrap values. The small bootstrap values in the phylogenetic tree potentially show relatively high similarity in the protein sequences. The result implies that there is not a considerable variability in the LEA gene sequences, even though the protein motifs are less conserved. It infers the close evolutionary relationship of the LEA protein in the Poaceae species. The phylogenetic tree revealed the sequence association as eight main clades in the evolution. Additionally, the phylogenetic tree strongly supports the existence of monophyletic LEA protein clades that originated in the early history of Poaceae. The aggregation pattern of the LEA proteins within the constructed tree indicates the common origin of the LEA protein sequences. The majority of the Poaceae family LEA proteins belong to an LEA-4 protein family. However, between all identified LEA protein groups, a close relationship was observed. Moreover, all identified groups shared a common point of origin. The proposed ancestral sequence of the Poaceae family members also confirmed that the identified LEA protein groups share a common ancestral (Fig. 4). Consequently, the LEA protein sequences from the identified groups are homologs (orthologues and, or paralogs) among the species. The result showed that LEA proteins rise from the similar LEA-4 grouped ancestral sequence, and the other groups evolve from the common LEA-4 ancestor. To estimate the time of the most recent common ancestor among LEA proteins, the time tree analysis was performed by the RelTime method by selecting Andropogoneae species as an outgroup based on the time scale and the divergence time of Poaceae inferred from Poaceae timetree database (Fig. 5).

Fig. 3. Evolutionary relationships of LEA protein sequences: The evolutionary history was inferred using the Neighbor-Joining method (Saitou *et al.,* 1987). The optimal unrooted tree with the sum of branch length = 8.92617397 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method (Jones *et al.,* 1992) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 100 amino acid sequences. All ambiguous positions were removed from each sequence pair. There were 450 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.,* 2016).

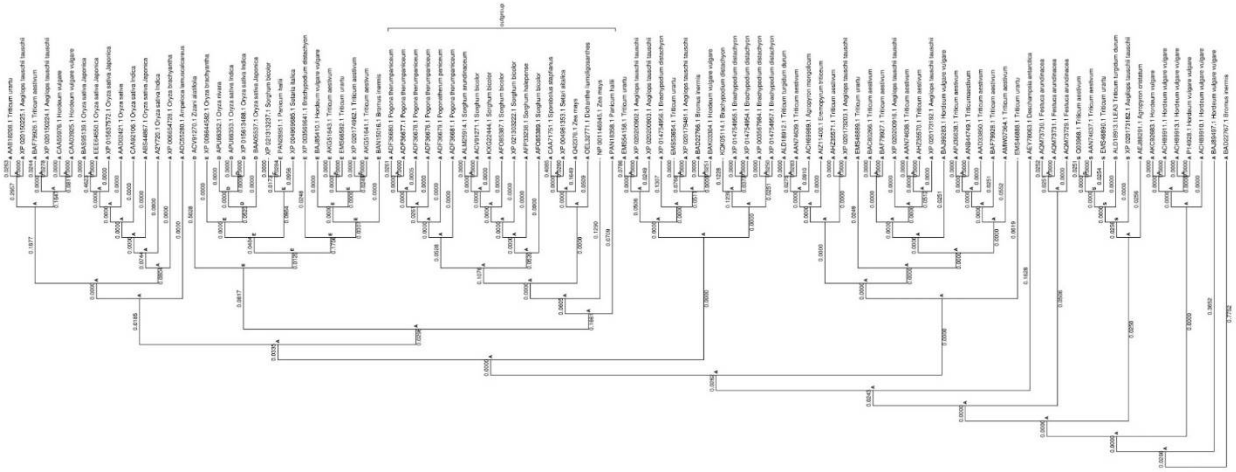

Fig. 4. Inferred Ancestral Sequences of Poaceae species: Ancestral states were inferred using the Maximum Likelihood method (Nei *et al.,* 2000) under the JTT matrix-based model (Jones *et al.,* 1992). The tree shows a set of possible amino acids (states) at each ancestral node based on their inferred likelihood at site 1. The set of states at each node is ordered from most likely to least likely, excluding states with probabilities below 5%. The initial tree was inferred using the method. The rates among sites were treated as a Gamma distribution with invariant sites using 0 Gamma Categories (Gamma with invariant sites option). The analysis involved 100 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were 38 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.,* 2016).

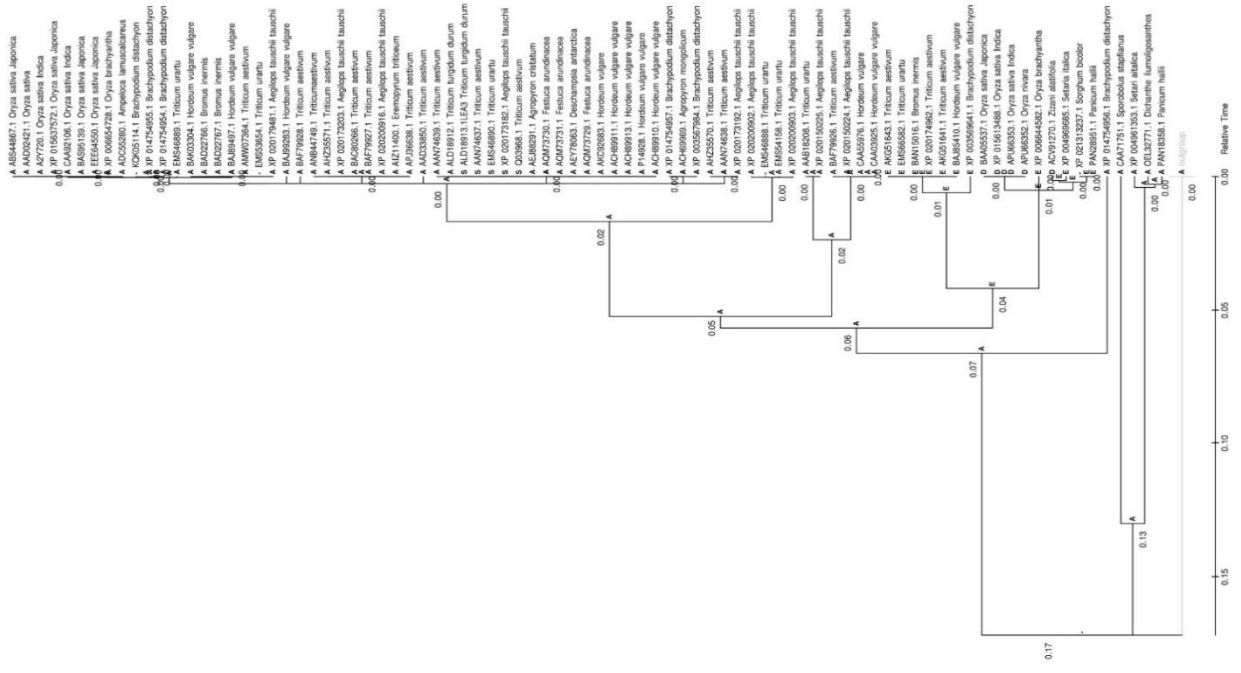

Fig. 5. Timetree analysis using the RelTime method: A timetree inferred using the RelTime method (Tamura *et al.* 2012) and the JTT matrix-based model (Jones *et al.,* 1992). The estimated log-likelihood value is -1097.04. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 18.8199)). The rate variation model allowed some sites to be evolutionarily invariable ([+I], 0.00% sites). The analysis involved 100 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were 38 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.,* 2016).

LEA gene proliferation analysis shows that LEA proteins are particularly prone to gene amplification (Fig. 5). Gene duplication leads to the production of homologous genes in the genome and the expansion of gene families. This result is consistent with Ibrahim and his colleagues (Ibrahim *et al*., 2019), which showed that LEA genes generate by replication. In addition, mutation and selection modes in repetitive cases lead to the emergence of a new gene with diverse functions. The results presented here suggest more than 53 gene duplication events (Fig. 6).

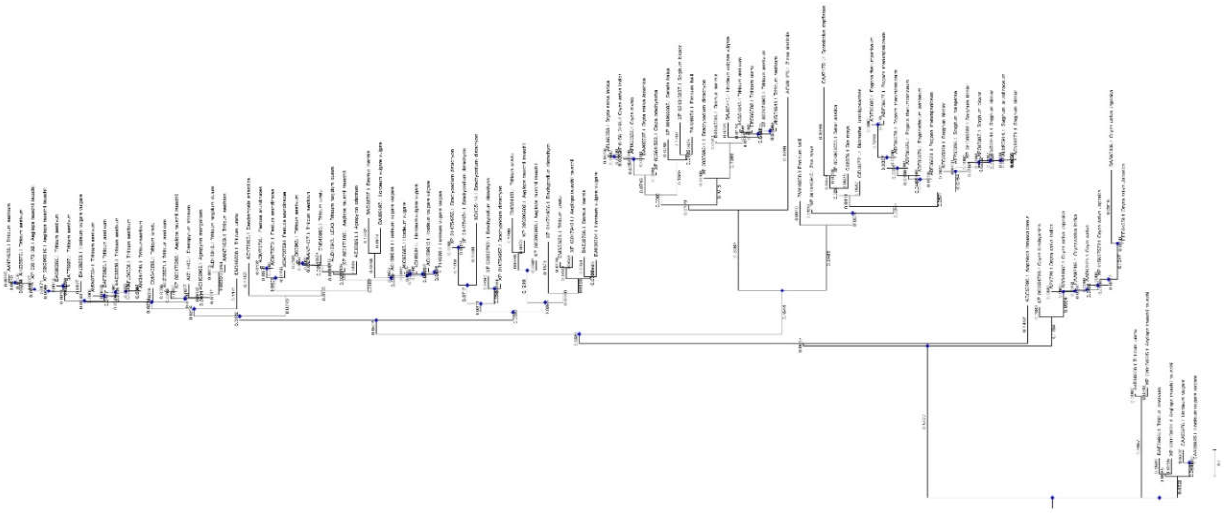

Fig. 6. Evolutionary relationships of Poaceae species: The evolutionary history shown is from a constructed Neighbor-joining tree. Gene duplications are identified by searching for all branching points in the topology with at least one species that is present in both subtrees of the branching point. There are 53 gene duplications (closed diamonds) identified in the tree. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.,* 2016).

Also, recent duplications observe in a given species of Poaceae. It seems that the expansion of the gene number in Poaceae was mainly by the duplicate of preexisting genes in the common ancestral species rather than by the appearance of the new LEA gene again reflects orthologous relationships between LEA protein in the Poaceae. Consequently, the representative of a homolog in the species occurred before divergence from their ancestor by gene duplication. To suggest plausible scenarios for the evolution of specific LEA protein properties and assigning polarity to character changes, we rooted the LEA proteins tree by the clade containing Andropogoneae species as an outgroup based on the time scale, and the divergence time of Poaceae inferred from Poaceae timetree database (Supplementary 8).

Discussion

The late embryogenesis abundant (LEA) gene family is an essential group of functional proteins, which is widely found in not only plant species but also fungi, bacteria, and even animal kingdom (Sasaki *et al*., 2014; Jin *et al*., 2019). LEA proteins have no significant enzymatic activity and intrinsically unstructured but partially fold into mainly α-helical structures under dehydration conditions (Hand *et al*., 2011; Hincha and Thalhammer, 2012). They play crucial roles during embryonic development and response to abiotic stresses to reduce cell damage and protect cells throughout the plant species (Ibrahim *et al*., 2019).

LEA proteins are small hydrophilic proteins with 10-30 kDa molecular weight (He *et al*., 1996). They are unstructured in the hydrated state. They induced under abiotic stress conditions like drought, high temperature, cold, and salinity in many plant species, function as chaperones via preventing functional proteins against aggregation, having a protective role from the damaging effects of abiotic stress consequently can protect the plant (Magwanga *et al*., 2018; Ibrahim *et al*., 2019). Also, LEA proteins ubiquitous within cells and their respective

tissues and present in all subcellular compartments, including the Nucleus, Cell membranes, Cell walls, Mitochondria, Chloroplasts, Golgi apparatus, and Cytoplasm (Liu *et al*., 2019), suggesting the fact that their function requires in all cellular compartments during stress.

LEA proteins classify into eight groups, including LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6, dehydrin (DHN), and seed maturation protein (SMP). In this study, based on an NCBI search of protein database and homologous sequence alignment, a total of 100 LEA proteins have identified five subfamilies based on their conserved domains and their sequence similarity with previously known protein sequences defined by the Pfam database. All the LEA proteins from Poaceae are found Gly and Lys hydrophilic amino acids abundant proteins and are hydrophilic like reported on the other plants (Hong-Bo, *et al*., 2005). Motif analysis of the LEA proteins by the MEME program showed that each LEA group within families contains specific conserved motifs compositions, but they vary among different LEA families. However, several group-specific conserved protein motifs resulted among the LEA proteins, supporting group-specific functions in the same group. It suggested that the protein structures have conserved among members of the same group. This result agrees with Cao and Li's study (Cao and Li, 2015). It indicates that the motifs encoding the LEA domains are conserved, and the conserved motifs observed within each LEA group determine the probable origin within the groups. The similar motif distribution in the LEA protein groups is likely to indicate the gene family expansion within the group (Kong *et al*., 2007; Cao and Li 2015; Zeng *et al*., 2018). However, the different motif composition among diverse groups suggests the evolution of one group from distinct ancestors with various motif structures among evolutionary duration. To reveal the phylogenetic relationships among the LEA proteins in the Poaceae species, in the first, the full-length protein sequences aligned by the Clustal Omega. The aligned LEA sequences were highly similar, and the conserved regions are determined.

During evolution, the two principal gene duplication mechanisms, including small-scale tandem and large segmental duplications, with polyploidization contribute to novel gene generation in higher plants. Gene duplication is the essential and efficient feature for plant genomic evolution, which provides new raw genetic materials for genetic drift throughout the mutation-selection scenario. These mechanisms result in the emergence of new functional genes and finally evolution in gene networks, which leads to the expansion of genome content and diversification of gene function. In this study, several lines of evidence point to the genome duplication of the LEA gene (Fig. 6), which remains the chief impetus for developing the LEA genes in the Poaceae family. Sequence similarity in each group of the LEA protein sequences of the Poaceae species provides supporting evidence. The same results have previously identified in several plant species such as cotton (Magwanga *et al*., 2018), solanum (Cao and Li, 2015), oilseed rape (Liang *et al*., 2016), and Arabidopsis (Hundertmark *et al*., 2008), which LEA-encoding genes have expanded through segmental or tandem duplications. Homologous and orthologous proteins are the result of gene duplication events. Therefore, the LEA protein orthologues and paralogs expansion in the Poaceae family are mostly the results of gene duplication events occurring in the closely related species.

Here, we provide phylogenetic analysis and evolutionary history for the Poaceae LEA protein sequences. Phylogenetic analysis of the LEA protein among 26 species of the Poaceae family pointed out that these proteins have a high level of conservation and perform similar functions among Poaceae. The neighbor-joining phylogenetic analysis showed the evolutionary relationship in the eight major clades. The evolutionary relationship in each subgroup indicates the direct homolog of the LEA protein group with each other. As a main result, the LEA protein family appears a monophyletic origin in the Poaceae. In the Poaceae, the LEA protein evolved from the Poaceae member's common ancestor during the evolutionary history. Our data will provide novel insights into further studies of the Late Embryogenesis Abundant protein family in Poaceae. However, additional

studies are necessary to further elucidating and verifying the functional role of the LEA genes in the Poaceae.

Conclusion

Our study aimed to provide a comparative phylogenetic analysis of LEA protein sequences in the 26 species of the Poaceae family by using bioinformatics tools. Our study provides several lines of evidence in sequence analysis, motif composition, gene duplication, phylogeny, and evolutionary history of the LEA proteins in the Poaceae family. It contains valuable theoretical knowledge for further study about the LEA protein features in diverse plant groups and species.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been supported by the University of Zabol in Grant code: UOZ-GR-9618-158

References

Abba S, Ghignone S, Bonfante P. 2006. A dehydration inducible gene in the truffle Tuber borchii identifies a novel group of dehydrins. *BMC Genom* 7: 1-15.

Kader AA, Almeslemani M, Baghdady A, Alzubi H, Alasaad N, Basha NA, Hassan F. 2012. Isolation, characterization of the hva1 gene from Syrian barely varieties and cloning into binary plasmid vector. *Int J Botany* 8(3): 117-126.

Alpert P, Oliver MJ. 2002. Drying without dying. Desiccation and survival in plants. Wallingford: CABI Publishing, 3-43.

Alpert P. 2005. The limits and frontiers of desiccation-tolerant life. *Integr Comp Biol* 45: 685-695.

Altunoglu YC, Baloglu MC, Baloglu P, Yer EN, Kara S. 2017. Genome-wide identification and comparative expression analysis of LEA genes in watermelon and melon genomes. *Physiol Mol Biol Plants* 23: 5-21.

Altunoglu YC, Baloglu PC, Nurten E et al. 2016. Identification and expression analysis of LEA gene family members in cucumber genome. *Plant Growth Regul* 80:225.

Bailey TL, Boden M, 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: W202-8.

Bray EA. 1997. Plant responses to water deficit. *Trend Plant Sci* 2: 48-54.

Browne JA, Dolan KM, Tyson T, Goyal K, Tunnacliffe A, Burnell AM. 2004. Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode *Aphelenchus avenae. Eukaryot Cell* 3: 966-975.

Campos F, Cuevas-Velazquez C, Fares MA, Reyes JL, Covarrubias AA. 2013. Group 1 LEA proteins, an ancestral plant protein group, are also present in other eukaryotes, and in the archeae and bacteria domains. *Mol Genet Genomics* 288(10): 503-517.

Cao J, Li X. 2015. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum). Planta* 241: 757-772.

Cuming AC. 1999. LEA proteins. In R Casey, PR Shewry, eds, Seed Proteins. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 753-780.

Dellaporta SL, Wood J, Hicks JB. 1983. A plant DNA minipreparation: version II. *Plant Mol Biol Rep* 1: 19-21.

Dure L, Chilan CA. 1981. Development biochemistry of cottonseed embryogenesis and germination. XII. Purification and properties of principle storage proteins. *Plant Physiol* 68: 180-186.

Dure L, Crouch M, Harada J, Ho TH, Mundy J, Quatrano R, Thomas T, Sung ZR. 1989. Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol Biol* 12:475-486.

Dure L, Greenway, SC, Galau GA. 1981. Developmental biochemistry of cottonseed embryogenesis and germination, changing messenger ribonucleic acid populations as shown by in vitro protein synthesis. *Biochemistry* 20: 4162-4168.

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

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.

Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Eddy SR. 2015. HMMER web server: 2015 update. *Nucleic Acids Res* 43: W30-W38.

Gao J, Lan T. 2016. Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabuliformis* (Pinaceae) in *Escherichia coli*. *Sci Rep* 19467.

Garay-Arroyo A, Colmenero-Flores JM, Garciarrubio A, Covarrubias AA. 2000. Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J Biol Chem* 275: 5668-5674.

Hand SC, Menze MA, Toner M, Boswell L, Moore D. 2011. LEA proteins during water stress: not just for plants anymore. *Annu Rev Physiol* 73, 115-134.

He JX, Fu JR. 1996. The research progresses in LEA proteins of seeds. *Plant Physiol Commun* 32(4): 241-246.

Hincha DK, Thalhammer A. 2012. LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem Soc T* 40:1000-1003

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

Hunault G, Jaspard E. 2010. LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics* 11: 1-9.

Hundertmark M, Hincha DK. 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9: 1-22.

Ibrahime M, Kibar U, Kazan K, Özmen CY, Mutaf F, Asci SD, Ergül A. 2019. Genome-wide identification of the LEA protein gene family in grapevine (*Vitis vinifera* L.). *Tree Genet Genomes* 15: 1-4.

Ingram J, Bartels D. 1996. The molecular basis of dehydration tolerance in plants. *Annu Rev Plant Biol* 47: 377-403.

Iturriaga G, Cushman MAF, Cushman JC. 2006. An EST catalog from the resurrection plant *Selaginella lepidophylla* reveals abiotic stress-adaptive genes. *Plant Sci* 170: 1173-1184.

Jaspard E, Macherel D, Hunault G. 2012. Computational and statistical analyses of amino acid usage and physicochemical properties of the twelve-late embryogenesis abundant protein classes. *Plos One* 16; 7(5): e36968.

Jin X, Cao D, Wang Z, Ma L, Tian K, Liu Y, Gong Z, Zhu X, Jiang C, Li Y. 2019. Genome-wide identification and expression analyses of the LEA protein gene family in tea plant reveal their involvement in seed development and abiotic stress responses. *Sci Rep* 9: 1-5.

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

Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, DePamphilis CW. 2007. The pattern of gene duplication in the plant *SKP1*gene family in angiosperms: evidence for multi mechanisms of rapid gene birth. *Plant J* 50:873-885.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 3:1870-1874.

Liang Y, Xiong Z, Zheng J, Xu D, Zhu Z, Xiang J, Li M. 2016. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*. *Sci Rep* 2:4265.

Liu H, Xing M, Yang W, Mu X, Wang X, Lu F, Zhang L. 2019. Genome-wide identification of and functional insights into the late embryogenesis abundant (LEA) gene family in bread wheat (*Triticum aestivum*). *Sci Rep* 9: 1-11.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47(W1): 636-641.

Magwanga RO, Lu P, Kirungu JN, Lu H, Wang X, Cai X, Liu F. 2018. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress

tolerance in upland cotton. *BMC Genetics* 19: 1-31.

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Oliver MJ, Dowd SE, Zaragoza J, Mauget SA, Payton PR. 2004. The rehydration transcriptome of the desiccation-tolerant bryophyte *Tortula ruralis*: transcript classification and analysis. *BMC Genomics* 5: 89-107.

Oliver MJ, Tuba Z, Mishler BD. 2000. The evolution of vegetative desiccation tolerance in land plants. *Plant Ecol* 151: 85-100.

Proctor MC, Oliver MJ, Wood AJ, Alpert P, Stark LR, Cleavitt NL, Mishler BD. 2007. Desiccation-tolerance in bryophytes: a review. *Bryologist* 110: 595-621.

Reynolds TL, Bewley JD. 1993. Characterization of protein synthetic changes in a desiccation-tolerant fern, *Polypodium virginianum*: comparison of the effects of drying, rehydration, and abscisic acid. *J Exp Bot* 44: 921-928.

Rorat T. 2006. Plant dehydrins--tissue location, structure, and function. *Cell Mol Biol Lett* 11(4): 536-556.

Saavedra L, Svensson J, Carballo V, Izmendi D, Welin B, Vidal S. 2006. A dehydrin gene in *Physcomitrella patens* is required for salt and osmotic stress tolerance. *Plant J* 45: 237-249.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425.

Sasaki K, Christov NK, Tsuda S, Imai R. 2014. Identification of a novel LEA protein involved in freezing tolerance in wheat. *Plant Cell Physiol* 55:139-147.

Singh TR, Bansal A. 2019. Gene duplication and speciation. *Encycl Bioinf Comput Biol* 3: 965-974.

Solomon A, Salomon R, Paperna I, Glazer I. 2000. Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat-stable protein. *Parasitology* 121: 409-416.

Stacy RAP, Aalen RB. 1998. Identification of Sequence Homology Between the Internal Hydrophilic Repeated Motifs in Group 1 Late-Embryogenesis-Abundant Proteins in Plants and Hydrophilic Repeats of the General Stress Protein GsiB of *Bacillus subtilis*. *Planta* 206: 476-478.

Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci* 109:19333-19338.

Tanaka S, Ikeda K, Miyasaka H. 2004. Isolation of a new member of group 3 late embryogenesis abundant protein gene from a halotolerant green alga by a functional expression screening with cyanobacterial cells. *FEMS Microbiol Lett* 236: 41-45.

Thomashow MF. 1998. Role of cold-responsive genes in plant freezing tolerance. *Plant Physiol* 118: 1-7.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinf* 25(9): 1189-91.

Wise MJ. 2003. LEAping to conclusions: a computational reanalysis of late embryogenesis abundant proteins and their possible roles. *BMC Bioinf* 4: 1-19.

Zeng X, Ling H, Yang J, Li Y, Guo S. 2018. LEA proteins from *Gastrodia elata* enhance tolerance to low-temperature stress in *Escherichia coli*. *Gene* 646:136-142.