Cloning and Drought-Stress Expression Analysis of the SpLEA1 Gene from *Selaginella pulvinata* (Postprint)

Authors: Zhou Xuan, Peng-Hua Gao, Yan Bo

Date: 2022-03-30T21:00:37Z

Abstract

Late embryogenesis abundant (LEA) proteins are closely associated with plant stress resistance, protecting plant cells under drought stress, reducing plant damage, and are widely distributed across organisms. Selaginella pulvinata is a fern species with exceptional survival capacity under drought stress and remarkable recovery ability. To investigate the molecular mechanism and expression characteristics of the SpLEA1 gene from S. pulvinata in drought-tolerant plants, this study employed the highly drought-tolerant plant S. pulvinata as experimental material. Based on transcriptome sequencing data, the cDNA sequence of the SpLEA1 gene was obtained via RT-PCR, the promoter sequence was obtained via HiTail-PCR, bioinformatics analysis was performed on these sequences, and the expression pattern of the SpLEA1 gene under drought stress was analyzed using qRT-PCR. The results demonstrated: (1) The full-length SpLEA1 from S. pulvinata was 476 bp, with an open reading frame (ORF) of 279 bp encoding 92 amino acids. Online tools predicted a protein molecular weight of 9,491.46 Da and an isoelectric point of 5.45. Protein structure prediction analysis indicated that this protein is hydrophilic, contains 10 phosphorylation sites, and its secondary structure is primarily composed of -helices and random coils. (2) The conserved domain of SpLEA1 protein was predicted to be Lea-5, originating from the LEA1 family. Based on phylogenetic tree and genetic distance matrix analysis, S. pulvinata SpLEA1 exhibited high homology with Lea-5 proteins from Cicer arietinum and Trifolium pratense. (3) Prediction and analysis of cis-acting elements in the promoter sequence using online tools revealed that the SpLEA1 gene promoter contains five categories of hormone-responsive elements and functional elements associated with drought stress response. (4) Under natural drought treatment, SpLEA1 gene expression was upregulated and peaked at 12 h; following rehydration treatment after 24 h of drought, the expression level was significantly downregulated. In summary, the SpLEA1 gene in S. pulvinata likely participates in the regulatory mechanisms associated with drought stress

response. These findings establish a foundation for further investigation into the function and expression regulatory mechanisms of the S. pulvinata SpLEA1 gene under drought stress.

Full Text

Cloning and Expression Analysis of the *SpLEA1* Gene from *Selaginella pulvinata* Under Drought Stress

Xuan Zhou, Penghua Gao, Bo Yan* College of Landscape Architecture and Horticulture, Southwest Forestry University, Kunming 650224, China

Abstract

Late embryogenesis abundant (LEA) proteins are closely associated with plant stress resistance, protecting plant cells and reducing damage under drought stress, and are widely distributed across organisms. *Selaginella pulvinata*, a fern species commonly known as resurrection plant, exhibits exceptional survival and recovery capabilities under drought conditions. To investigate the molecular mechanisms and expression characteristics of the *SpLEA1* gene in this highly drought-tolerant species, we cloned the *SpLEA1* cDNA sequence using RT-PCR based on transcriptome sequencing data, isolated its promoter region via HiTail-PCR, performed comprehensive bioinformatics analyses, and examined its expression patterns under drought stress using qRT-PCR. The results revealed: (1) The full-length *SpLEA1* gene spans 476 bp, with an open reading frame (ORF) of 279 bp encoding 92 amino acids. Bioinformatic prediction indicated a molecular weight of 9,491.46 Da and an isoelectric point of 5.45. Structural analysis predicted that the protein is hydrophilic, contains 10 phosphorylation sites, and is predominantly composed of -helices and random coils in its secondary structure. (2) The conserved domain was identified as LEA-5, placing *SpLEA1* in the LEA1 family. Phylogenetic analysis and genetic distance matrix revealed high homology with Lea-5 proteins from chickpea (*Cicer arietinum*) and red clover (*Trifolium pratense*). (3) Promoter analysis identified five classes of hormone-responsive cis-acting elements and functional elements associated with drought stress response. (4) Under natural dehydration treatment, *SpLEA1* expression was upregulated, peaking at 12 h, and significantly downregulated upon rehydration after 24 h of drought. These findings suggest that *SpLEA1* likely participates in the regulatory mechanisms of drought stress response in *S. pulvinata*, providing a foundation for further functional characterization and expression regulation studies of this gene.

Keywords: *Selaginella pulvinata*; *SpLEA1*; gene cloning; promoter cloning; expression analysis

Introduction

Plants have evolved sophisticated physiological and molecular response mechanisms to cope with biotic and abiotic stresses. Research has demonstrated that late embryogenesis abundant (LEA) proteins are associated with plant stress resistance and play widespread roles in plant responses to abiotic stress (Li, 2016). LEA proteins were first isolated and cloned from cotton seeds (Dure et al., 1981), and subsequent studies revealed their broad distribution across plants, invertebrates, and prokaryotes. LEA genes are expressed throughout plant development, with particularly abundant accumulation in plant tissues and cells during environmental stresses such as drought and high temperature, enabling plants to cope with adverse conditions (Wise, 2003; Silveira et al., 2008). In 2010, Hunault & Jaspard established the LEA protein database (LEAPdb), which classifies LEA proteins based on eight conserved PFAM domains in their amino acid sequences. Identification of the PFAM number for a LEA protein through this database enables determination of its family membership, providing a basis for subsequent experimental investigations.

The LEA1 family (group 1 late-embryogenesis-abundant proteins) comprises hydrophilic proteins that exist in a largely unstructured form and are widely distributed in plants. Typical representatives include cotton D-19, wheat EM protein, and barley B19 protein. LEA1 family members are characterized by their hydrophilicity and variable numbers of a highly conserved 20-amino-acid motif (GGETRKEQLGEEGYREMGRK) (Stacy et al., 1995). Cloning a gene' s promoter is essential for systematic functional analysis. Plant promoters are DNA sequences containing transcription start sites that regulate gene expression, with transcription frequency, direction, and initiation site being critical for transcriptional control (Liu et al., 1997; Wang et al., 2011; Zhang et al., 2019). Promoters contain specific regulatory motifs (Mei, 2018) and can be categorized as constitutive, inducible, or tissue-specific (Yang et al., 2018). Studies have shown that inducible promoters cause transient or sustained upregulation of gene expression when environmental conditions change (D' urzo et al., 2013). Zheng et al. (2019) isolated an atypical LEA gene (*LpLEA*) and its promoter from cabbage, demonstrating the presence of unique cis-regulatory elements associated with abiotic stress in the *LpLEA* promoter and enhanced expression in various tissues under different abiotic stresses and ABA induction.

Selaginella pulvinata, commonly known as "Jiuhuan Huncao"(resurrection grass), is primarily distributed in arid regions of China, typically growing on exposed limestone surfaces or rock crevices. As a terrestrial or lithophytic resurrection plant belonging to the family Selaginellaceae, it exhibits remarkable drought tolerance (Wu, 2004). Research indicates that *S. pulvinata* possesses unique reactive oxygen species generation and scavenging pathways, enhanced abscisic acid biosynthesis, and potential ABA signaling and response mechanisms. Analysis of its chloroplast genome revealed unique structural rearrangements and complete loss of chloroplast NAD(P)H dehydrogenase (NDH) genes (Saucedo et al., 2017). LEA1 proteins are induced by drought, salt stress, ABA, and low temperature during seedling stages, protect lactate dehydrogenase activity, and positively regulate the expression of some calcium-dependent protein kinases (Zou, 2011; Xiang et al., 2018). The role of LEA1 proteins as important regulators of plant tolerance warrants investigation in highly drought-tolerant species such as *S. pulvinata*. Currently, research on LEA proteins in ferns is limited, and studies on LEA1 genes in *S. pulvinata* are virtually nonexistent. Therefore, cloning and analyzing the molecular mechanisms and expression characteristics of the LEA1 gene in *S. pulvinata* is significant for understanding its regulatory mechanisms during drought response. This study isolated and cloned the *SpLEA1* gene, performed bioinformatics analysis of its sequence and promoter cis-acting elements, constructed phylogenetic trees and genetic distance matrices, aligned homologous protein sequences, and examined expression patterns in young leaves under different drought conditions using real-time quantitative PCR. These investigations lay the foundation for exploring the function and molecular mechanisms of *SpLEA1* under drought stress and provide genetic resources for improving drought resistance in ornamental horticultural plants.

Materials and Methods

1.1 Plant Materials and Treatments *Selaginella pulvinata* plants were collected from the suburbs of Kunming, Yunnan Province, and cultivated in growth chambers (16 h photoperiod; 25°C; 20% relative humidity). Fresh leaves and shoots were used for DNA and cDNA extraction. For gene expression analysis, plants with low soil content collected from rock surfaces were fully watered at the roots, then subjected to natural dehydration treatment. Fresh young leaves and shoots were collected at specific time points, divided into six treatment groups: 0 h (fully watered control), 2 h dehydration, 4 h dehydration, 12 h dehydration, 24 h dehydration, and rehydration for 2 h after 24 h dehydration. Each treatment group consisted of six uniformly sized plants with three biological replicates. All samples were snap-frozen in liquid nitrogen and stored at -80°C.

1.2 Molecular Cloning and Analysis 1.2.1 Genomic DNA Cloning of *SpLEA1*

Total DNA was extracted from *S. pulvinata* using a plant leaf DNA extraction kit (Tiangen Biotech, Beijing) according to the manufacturer’ s instructions. Based on our laboratory’ s transcriptome sequencing data, full-length primers for *SpLEA1* were designed (Table 1). PCR amplification was performed using *S. pulvinata* genomic DNA as template in a 40 L reaction containing 2.0 L DNA template, 16 L 2×Taq PCR MasterMix, 1.0 L each of forward and reverse primers, and 20 L ddH O. The amplification program consisted of: 94°C for 2 min; 38 cycles of 94°C for 30 s, 57°C for 30 s, 72°C for 90 s; and final extension at 72°C for 10 min. PCR products were detected by 1% agarose gel electrophoresis, purified using a gel extraction kit (OMEGA), ligated into the PMD18-T cloning vector (Takara), and transformed into *E. coli* DH5 competent cells (Tiangen).

Positive clones identified by colony PCR were sequenced by Sangon Biotech.

1.2.2 cDNA Cloning of *SpLEA1*

Total RNA was extracted using an RNA extraction kit (OMEGA) following the manufacturer' s protocol. First-strand cDNA was synthesized using a reverse transcription kit (TransGen Biotech) with extracted RNA as template. The *SpLEA1* cDNA was amplified using the same primers and PCR conditions as described in section 1.2.1, followed by cloning and sequencing to obtain the full-length cDNA.

1.2.3 Promoter Cloning of *SpLEA1*

Using genomic DNA as template, promoter-specific primers *SpLEA1*Q1/2/3 (Table 1) were designed based on the *LEA1* gene sequence for HiTail-PCR amplification. These primers were used sequentially in three rounds of PCR combined with random primers LAD1-1/2/3/4 (Table 1) designed by Liu & Chen (2007). The third-round PCR product was analyzed by electrophoresis, and the target band was excised, purified, cloned, and sequenced.

1.2.4 Bioinformatics Analysis

Amino acid physicochemical properties were analyzed using ExPASy-ProtParam (https://web.expasy.org/protparam/). Gene structure information was obtained using Softberry (http://linux1.softberry.com/berry.phtml). Phosphorylation sites were predicted using NetPhos 3.1 Server (http://www.cbs.dtu.dk/services/NetPhos/). Protein hydrophilicity/hydrophobicity was analyzed using Protscale (https://web.expasy.org/protscale/). Secondary structure prediction was performed using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html). Tertiary structure modeling was conducted using Swiss-Model (https://swissmodel.expasy.org/interactive) via homology modeling. Multiple sequence alignment and genetic distance matrix analysis were performed using DNAMAN9. Phylogenetic trees were constructed using MEGA X software. Cis-acting elements in the promoter were analyzed using PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

1.2.5 Quantitative Real-Time PCR Analysis of *SpLEA1* Expression

The expression pattern of *SpLEA1* under drought stress was analyzed by qRT-PCR. Gene-specific primers were designed based on the *SpLEA1* sequence, with *Selaginella* actin as the reference gene (Table 1). Using cDNA as template, qRT-PCR was performed with TB Green® Premix Ex Taq™ II (Takara, RR820A) following the manufacturer' s instructions. Each sample was run in triplicate to minimize error. The amplification program consisted of: 95°C for 30 s; 40 cycles of 95°C for 5 s, 60°C for 30 s. Relative expression levels were calculated using the 2^(-ΔΔCt) method.

Table 1 Primer sequences used in this study

Primer Name	Sequence (5 →3)
SpLEA1-F	ATGGCTTCTGCACAGGAAAAG
SpLEA1-R	TTAATCAGTCTTCTTAAACTTGC
*SpLEA1*Q1	CTCATCGATGTCAATCCCACGC
*SpLEA1*Q2	CCTTCCTTGCCTAACTGCTCTG
*SpLEA1*Q3	CTTCAGCAAGCCTTTCCTGCG
QREJEM-1	CAGCACAGGGCAGAGCAGTTAG
QRTJEM-2	CCTCTCCTCCTTCCGCACCAG
LAD1-1	ACGATGGACTCCAGAGCGGCCCGCVNVNNNGGAA
LAD1-2	ACGATGGACTCCAGAGCGGCCCGCBNBNNNGGTT
LAD1-3	ACGATGGACTCCAGAGCGGCCCGCVVNVNNNCCAA
LAD1-4	ACGATGGACTCCAGAGCGGCCCGCBDNBNNNCGGT

Results

2.1 Cloning of the *SpLEA1* Gene Using *S. pulvinata* genomic DNA as template with primers *SpLEA1*-F/R, PCR amplification and sequencing yielded a 475 bp genomic fragment of *SpLEA1*. Amplification from cDNA template produced a 279 bp *SpLEA1* cDNA sequence (Fig. 1A). Sequence analysis revealed that *SpLEA1* contains one intron (196 bp) and two exons (115 bp and 164 bp), with the 279 bp cDNA encoding 92 amino acids (Fig. 1B). Querying the NCBI PFAM database identified a conserved LEA-5 domain spanning amino acid positions 2-88, with PFAM accession PF00477, confirming that *SpLEA1* belongs to the LEA1 family.

Secondary structure prediction using SOPMA indicated that the SpLEA1 protein comprises 40.22% -helices, 18.48% -turns, 36.96% random coils, and 4.35% extended strands (Fig. 1C). Tertiary structure modeling via Swiss-Model demonstrated that SpLEA1 is primarily composed of -helices and random coils, with minor contributions from -turns and extended strands.

Figure 1 Sequence analysis of the *SpLEA1* gene. (A) Amplification of *SpLEA1* cDNA. M: DL2000 DNA marker; lane 1: *SpLEA1* gene. (B) *SpLEA1* cDNA sequence and encoded amino acid sequence. (C) Secondary structure prediction of SpLEA1 protein. Blue: -helix; green: -turn; purple: random coil; red: extended strand.

2.2 Bioinformatics Analysis of SpLEA1 ProtParam analysis predicted that the SpLEA1 protein has a molecular weight of 9,491.46 Da, theoretical isoelectric point of 5.45, and molecular formula C H N O S . The protein contains 16 negatively charged residues (Asp + Glu) and 14 positively charged residues (Arg + Lys), with an instability index of 28.72, classifying it as a stable protein. The amino acid composition includes hydrophilic residues Thr (4.3%), Lys (9.8%), Gln (5.4%), Gly (18.5%), and Glu (13.0%), along with hydrophobic residues Ala (12.0%), Met (2.2%), Val (3.3%), Ile (3.3%), and Leu

(6.5%). Hydrophilic amino acids account for 51% of the total, while hydrophobic residues comprise 27.4%, yielding an average hydrophilicity index of -0.838.

Protscale analysis confirmed the predominantly hydrophilic nature of SpLEA1, with most amino acids showing values below 0. The strongest hydrophobicity occurred at position 68 (score: 0.892), while maximum hydrophilicity was observed at position 43 (score: -2.433) (Fig. 2), confirming SpLEA1 as a hydrophilic protein.

Figure 2 Hydrophilicity prediction of SpLEA1 protein.

NetPhos 3.1 Server prediction identified 10 potential phosphorylation sites in SpLEA1: six serine residues (positions 3, 29, 57, 66, 67, 86), three tyrosine residues (positions 19, 25), and one threonine residue (position 76).

Phylogenetic analysis using MEGA X and genetic distance matrix analysis using DNAMAN revealed that SpLEA1 clusters with high homology to Lea-5 proteins from chickpea (*Cicer arietinum*) and red clover (*Trifolium pratense*) (Fig. 3). Multiple sequence alignment of SpLEA1 with 15 closely related but distinct species, including chickpea, flax, and rye, demonstrated the presence of the conserved LEA-5 domain (Fig. 4). The genetic distance matrix showed the closest relationship with red clover and chickpea (distance: 0.272) and the most distant relationship with *Capsella rubella* (distance: 0.348).

Figure 3 Phylogenetic analysis of SpLEA1 protein. **Figure 4** Multiple sequence alignment of SpLEA1 protein with homologs from 15 plant species. Red lines indicate conserved domain intervals. Species abbreviations: Sp: *Selaginella pulvinata*; Ss: *Salvia splendens*; Bd: *Brachypodium distachyon*; Cm: *Cucurbita moschata*; Eg: *Eucalyptus grandis*; Cn: *Cocos nucifera*; Egs: *Elaeis guineensis*; Tp: *Trifolium pratense*; Si: *Sesamum indicum*; Mc: *Momordica charantia*; Cr: *Capsella rubella*; Ts: *Telopea speciosissima*; Td: *Triticum dicoccoides*; Ca: *Cicer arietinum*; Cs: *Cucumis sativus*; Sc: *Secale cereale.*

2.3 Promoter Cloning and Functional Element Analysis of *SpLEA1*

HiTail-PCR amplification using genomic DNA as template yielded a ~2,000 bp fragment in the third-round PCR (Fig. 5). Sequencing revealed a 2,018 bp region upstream of the *SpLEA1* start codon (ATG). PlantCARE analysis identified core promoter elements including TATA-box and CAAT-box, along with numerous cis-acting elements related to abiotic stress responses (Fig. 6, Table 2). These included hormone-responsive elements: abscisic acid response elements (ABRE), methyl jasmonate response elements (CGTCA-motif and TGACG-motif), gibberellin response elements (GARE-motif and P-box), salicylic acid response element (TCA-element), and auxin response element (TGA-element). Additionally, drought-associated functional elements such as MYB binding sites (MBS) involved in drought induction were identified. These findings suggest that *SpLEA1* expression significantly influences the drought survival capacity of *S. pulvinata*.

Figure 5 HiTail-PCR amplification of the *SpLEA1* gene promoter. M: DL2000 DNA marker; lanes 1-4 represent third-round PCR products using specific primer *SpLEA1*Q3 combined with random primers LAD1-1/2/3/4. The product in the black box from lane 3 was recovered for cloning.

Figure 6 Sequence and partial cis-acting elements of the *SpLEA1* gene promoter. Boxes indicate cis-elements; bold black sequences represent TATA-box; italicized sequences represent CAAT-box.

Table 2 Predicted cis-acting elements in the *SpLEA1* promoter

Cis-acting element	Core sequence	Function
TATA-box	TATA/ATATAT	Core promoter element around -30 of transcription start
CAAT-box	CAAT/CCAAT/CAAAT	Common cis-acting element in promoter and enhancer regions
I-box	CCGAAA	Part of light-responsive element
G-Box	CACGTT	Cis-acting regulatory element involved in light responsiveness
GARE-motif	TCTGTTG	Gibberellin-responsive element
P-box	CCTTTTG	Gibberellin-responsive element
ABRE	ACGTG/TACGGTC/GC	Cis-acting element involved in abscisic acid responsiveness
MBS	CAACAG/CAACCA/CAACTG	MYB binding site involved in drought-inducibility

Cis-acting element	Core sequence	Function
TCA-element	TCAGAAGAGG	Cis-acting element involved in salicylic acid responsiveness
TGACG-motif	TGACG	Cis-acting regulatory element involved in MeJA responsiveness
CGTCA-motif	CGTCA	Cis-acting regulatory element involved in MeJA responsiveness
TGA-element	AACGAC	Auxin-responsive element

2.4 Quantitative Real-Time PCR Analysis of *SpLEA1* Expression qRT-PCR analysis revealed that *SpLEA1* expression was upregulated during natural dehydration, peaking at 12 h and subsequently declining. Rehydration after 24 h of drought caused a significant decrease in expression (Fig. 7). These results indicate that *SpLEA1* participates in the drought stress response of *S. pulvinata*.

Figure 7 Expression of *SpLEA1* gene under drought stress. * indicates significant difference compared to 0 h treatment ($P<0.05$). 0, 2, 4, 12, 24, and re24 represent dehydration treatment times (h), where re24 indicates samples rehydrated for 2 h after 24 h of dehydration.

Discussion and Conclusion

In this study, the SpLEA1 protein was identified as belonging to the LEA1 family with a conserved LEA-5 domain (PF00477). Sequence comparison with 15 LEA proteins from NCBI revealed two conserved sequences at the N- and C-termini, consistent with the report by Battaglia et al. (2008). Phylogenetic analysis did not identify closely related homologs within the same genus, but showed high homology with Lea-5 proteins from chickpea and red clover. Previous studies on chickpea *CarLEA793* and *CarLEA4* genes have advanced understanding of the molecular mechanisms by which LEA proteins protect plant cells under drought stress (Gu, 2010). Structural studies of soybean LEA1 proteins have shown that conserved motifs Em-C and Em-2M adopt different structures and aggregation behaviors under various environmental conditions, yet primarily maintain an unstructured form, which may relate to the structural characteristics of Em proteins and the functional significance of key regions (Xue et al., 2012). Zou (2011) demonstrated that LEA1 proteins can form different spatial structures in response to environmental changes, with these special structures interacting with -helices to protect enzyme activities and protein structures in cells. Our structural predictions for SpLEA1 using SOPMA and Swiss-Model revealed a predominance of -helices and random coils, consistent with structural features of the LEA1 family, suggesting that SpLEA1 participates in drought response through the formation of -helical and unstructured regions. The qRT-PCR results showing upregulation of *SpLEA1* during drought treatment, peaking at 12 h and decreasing upon rehydration, indicate that SpLEA1 is induced by drought stress, which may contribute to the high drought tolerance of *S. pulvinata*. The spatiotemporal expression patterns of *SpLEA1* provide a basis for future overexpression studies.

Promoter cloning via HiTail-PCR and analysis using PlantCARE revealed core elements TATA-box and CAAT-box, indicating stable expression capability. The promoter region contains numerous inducible elements, including five classes of hormone-responsive elements and abiotic stress-responsive elements. Notably, seven ABRE elements, four MYB sites, and four MYC sites associated with water stress were identified, correlating with the upregulation of *SpLEA1* expression under drought treatment observed in our qRT-PCR experiments. These findings confirm that *SpLEA1* participates in drought stress response, though the detailed mechanisms require further investigation.

In conclusion, we cloned the full-length *SpLEA1* cDNA from *S. pulvinata* and demonstrated through protein structure analysis that SpLEA1 is a hydrophilic, stable protein. Promoter isolation and cis-element analysis suggest that *SpLEA1* functions in the drought tolerance mechanism of *S. pulvinata*, with qRT-PCR confirming its high expression under drought conditions. We propose that *SpLEA1* is associated with the exceptional drought tolerance of *S. pulvinata* and participates in drought stress response regulation. Future studies should construct overexpression vectors for yeast or *Arabidopsis* to validate *SpLEA1* function in prokaryotic and eukaryotic systems. Further exploration of this research will provide insights for plant drought resistance studies and enhance survival rates of ornamental horticultural plants under drought conditions.

References

Battaglia M, Olvera-Carrillo Y, Garcirrubio A, et al., 2008. The enigmatic LEA proteins and other hydrophilins. *Plant Physiology*, 148(1): 6-24.

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

D' urzo N, Martinelli M, Nenci C, et al., 2013. High-level intracellular expression of heterologous proteins in *Brevibacillus choshinensis* SP3 under the control of a xylose inducible promoter. *Microbial Cell Factories*, 12(1): 12.

Gu HY, 2010. *Cloning, expression, and subcellular localization of two LEA protein genes, CarLEA793 and CarLEA4 in chickpea (Cicer arietinum L.).* Master's thesis, Nanjing Agricultural University, Nanjing, China.

Hunault G, Jaspard E, 2010. LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics*, 11(1): 221.

Jaspard E, Macherel D, Hunault G, et al., 2012. Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve late embryogenesis abundant protein classes. *PLoS ONE*, 7(5): e36968.

Li X, 2016. *Molecular evolutionary research of the plant LEA gene family and preliminary functional analysis of an atypical LEA gene in Arabidopsis.* Master's thesis, Jiangsu University, Zhenjiang, China.

Liu YG, Chen Y, 2007. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *BioTechniques*, 43(5): 649.

Liu ZB, Hagen G, Guilfoyle TJ, 1997. A G-box-binding protein from soybean binds to the E1 auxin-response element in the soybean GH3 promoter and contains a proline-rich repression domain. *Plant Physiology*, 115(2): 397-407.

Mei YQ, 2018. *Effect of DAD gene promoter sequence variation on upland cotton domestication.* Master' s thesis, Qufu Normal University, Qufu, China.

Saucedo AL, Hernández-Domínguez EE, de Luna-Valdez LA, et al., 2017. Insights on structure and function of a late embryogenesis abundant protein from *Amaranthus cruentus*: an intrinsically disordered protein involved in protection against desiccation, oxidant conditions, and osmotic stress. *Frontiers in Plant Science*, 8: 497.

Silveira V, Santa-Catarina C, Balbuena TS, et al., 2008. Endogenous abscisic acid and protein contents during seed development of *Araucaria angustifolia*. *Biologia Plantarum*, 52: 101-104.

Stacy RA, Espelund M, Sæbø-Larssen S, et al., 1995. Evolution of the group 1 late embryogenesis abundant (Lea) genes: analysis of the Lea B19 gene family in barley. *Plant Molecular Biology*, 28(6): 1039-1054.

Wang Z, Zhao L, Li WB, 2011. Advances in the study of inducible promoters in plants. *Soybean Science and Technology*, (3): 5-9.

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

Wu ZY, 2004. *Flora Reipublicae Popularis Sinicae.* Science Press, Beijing, China.

Xiang DJ, Man LL, Zhang CL, et al., 2018. A new Em-like protein from *Lactuca sativa*, LsEm1, enhances drought and salt stress tolerance in *Escherichia coli* and rice. *Protoplasma*, 255(4): 1089-1106.

Xue R, Zou YD, Zheng YZ, et al., 2012. Structures and assembly of conserved domains from soybean late embryogenesis abundant protein Em. *Analytical Chemistry*, 40(4): 563-568.

Yang RJ, Bai JR, Li R, et al., 2018. Research progress of the inducible promoters in plant genetic engineering. *Shanxi Agricultural Sciences*, 46(2): 292-298.

Zhang XY, He H, Li Z, et al., 2019. Functional analysis of the A7-FT promoter in *Brassica napus* L. *Acta Agriculturae Boreali-Sinica*, 34(2): 59-65.

Zheng J, Su H, Lin R, et al., 2019. Isolation and characterization of an atypical LEA gene (*IpLEA*) from *Ipomoea pes-caprae* conferring salt/drought and oxidative stress tolerance. *Scientific Reports*, 9(1): 14838.

Zou YD, 2011. *Molecular mechanism for the enhanced salt tolerance of Arabidopsis thaliana, Saccharomyces cerevisiae and Escherichia coli by soybean EM(LEA1) protein*. PhD dissertation, Northeast Normal University, Changchun, China.

Note: Figure translations are in progress. See original paper for figures.