



Molecular characterization of late embryogenesis abundant proteins encoding genes from wild sorghum genotype IS-18,909 expressed under drought stress

N. R. Ujjainkar¹ · A. U. Panpatil¹ · V. P. Chimote¹ · B. D. Pawar¹  · A. A. Kale¹

Received: 2 April 2023 / Revised: 5 February 2024 / Accepted: 23 February 2024
© The Author(s) under exclusive licence to Society for Plant Research 2024

Abstract

As a drought-tolerant crop, sorghum is an ideal plant for identifying genes conferring drought tolerance. In the present study, LEA1 and LEA 3 genes expressing in response to PEG-6000 induced osmotic stress were isolated from drought and heat tolerant wild sorghum genotype IS-18,909. Gene-specific primers were designed and used for cDNA synthesis of late embryogenesis abundant (LEA) proteins encoding *LEA1* (547 bp) and *LEA3* (817 bp) genes. Sequence analysis of *LEA1* (547 bp) and *LEA3* (817 bp) cDNA confirmed the presence of corresponding full-length coding sequences. LEA1 and LEA3 proteins had specific 20 nucleotides and 11 nucleotides consensus sequences, respectively, and other conserved sequences. These drought-induced proteins are extremely hydrophilic, resistant to dehydration, and high in amino acid residues with the sulfhydryl group (serine, threonine). They are composed largely of the amino acids glycine, alanine, and glutamine and lack cysteine and tryptophan. The genes conferring drought tolerance may provide a foundation for improving sorghum productivity under water deficit conditions.

Keywords Drought stress · Hydrophilic proteins · Conserved motifs, LEA1, LEA3

Introduction

Sorghum bicolor (L.) Moench, is an important staple food crop with high ethanol production potential. Sorghum crop is grown mostly under rainfed conditions and exposed to abiotic stress at different stages of its life cycle, although it can cope with many such stresses (Badigannavar et al. 2018; Sanchez et al. 2002). Considering ongoing climatic changes caused principally by global warming, the pressure on crop production in water-limited environments is expected to increase. In India, lower yields of *Rabi* sorghum grown on residual moisture is mainly due to low adoption of improved varieties to abiotic stress like drought. Genetic improvement of sorghum cultivars can further increase their yields under rainfed conditions (Ongom et al. 2016).

Drought resistance in sorghum is a complex trait influenced by many genes coding for various contributing factors towards drought tolerance (Ali et al. 2011). A high-quality sorghum genome sequence data has been developed and identified 34,211 genes which can help future molecular studies (McCormick et al. 2018). The genes conferring drought resistance can provide a foundation for the scientific improvement of sorghum productivity under water deficit conditions. Understanding the drought tolerance mechanism and expression of drought resistance genes can help breed high-yielding drought-tolerant varieties. Plants overcome drought stress, by modifying several physiological and biochemical mechanisms. Drought and desiccation tolerance in plants is correlated with the presence of considerable quantities of compatible osmolytes, antioxidative enzymes, and specific desiccation tolerance proteins (Damame et al. 2016; Pardo et al. 2020; Oliver et al. 2020).

Late Embryogenesis Abundant (LEA) proteins are thermo-stable, highly hydrophilic and induced by abiotic stress. They help to prevent crystallization of cellular components under water deficit. LEA proteins function as hydration buffers and ion chelators to protect macromolecules or

✉ B. D. Pawar
bhau.raje@gmail.com

¹ State Level Biotechnology Centre, Mahatma Phule Krishi Vidyapeeth, 413 722 Rahuri, Ahmednagar, Maharashtra, India

stabilize membranes at the onset of dehydration. In stress tolerant sorghum genotypes, LEA proteins play a critical role in stabilization of cell membrane (Vincour and Altman, 2005). Under drought stress condition, levels of reactive oxygen species (ROS) increased. For protection of hydrophobic component of enzymes from ROS derivatives, several amino acids such as glycine, glutamine, glutamic acids, threonine, asparagine, serine, and aspartic were synthesized (Ogbaga et al. 2016). LEA proteins are rich in these amino acids. Nagraja et al., (2019), identify 68 LEA genes belonging to eight families in sorghum bicolor. Majority of those LEA genes are intron less. Abdel-Ghany et al. (2020) identified 180 genes from sorghum that were differentially expressed in response to drought, of which most (70%) of them were up-regulated under drought stress. The earliest genes to be up-regulated encode for transcription factors. *LEA* genes were top up-regulated genes in response to drought stress in RNA-sequencing-based differential gene expression studies (Johnson et al. 2014; Fracasso et al. 2016; Abdel-Ghany et al. 2020). Under mild to severe drought stresses, 11 *LEA* genes were up-regulated, while the expression of these genes was reduced on re-watering (Deng-feng et al. 2019). This result indicates an important role of *LEA* protein genes under drought stress in sorghum.

The wild sorghum genotype IS-18,909 possesses incredible drought and heat tolerance (Hinge et al. 2015, Gokul et al. 2014). It can tolerate extreme drought stress and hence this genotype can be considered as a good gene source to combat various abiotic stress. Keeping in mind the above facts, the present research program involved studies on genes encoding late embryogenesis abundant (LEA) proteins, during drought stress in sorghum.

Materials and methods

Collection of plant materials and RNA isolation

The present investigation was conducted at State Level Biotechnology Centre, Mahatma Phule Krishi Vidyapeeth (MPKV), Rahuri, District-Ahmednagar, India. The seeds of wild sorghum genotype IS-18,909 (*Sorghum bicolor* (L.) Moench subsp. *verticilliflorum* (Steud.) de Wet ex Wiersema & J. Dahlb.) were obtained from the Senior Sorghum Breeder, All India Coordinated Sorghum Improvement Project (AICSIP), MPKV Rahuri. Seeds were germinated

on vermiculite under laboratory conditions. The 15 days old seedlings of each cultivar were subjected to PEG-6000 induced drought stress to obtain -0.4 bar (0.4Mpa) osmotic pressure as per Michel and Kaufmann (1973). Leaves of seedling of each cultivar from control and stressed condition were immediately used for total RNA isolation using the QiaRNeasy Miniprep isolation kit. Formaldehyde agarose (FA) gel electrophoresis (0.8% agarose) was used to confirm the integrity of the RNA. Total RNA extracted was quantified spectrophotometrically to dilute the RNA samples to working concentration.

cDNA synthesis by RT-PCR (reverse transcription PCR)

Single Step GeNei AMV RT-PCR (Reverse Transcription PCR) kit was used to synthesize complementary cDNA from sorghum leaf RNA; which was further amplified by using Hot-start Taq DNA polymerase. First step reverse transcription for the synthesis of cDNA was carried out at 50°C for 1 h. Hot start Taq DNA polymerase was activated at 95°C for 15 min. The amplification of cDNA was done by performing 40 cycles of denaturation at 94 °C for 1 min, annealing at 50–60 °C (Table 1) for 1 min, extension at 72 °C for 1.30 min in a thermal cycler (Eppendorf Master Cycler Gradient, Germany). The final extension was carried out at 72 °C for 1 min, and 4 °C holds up to retrieval. The 10µL RT-PCR reaction mixture consists of components provided with single step GeNei AMV RT-PCR kit along with 25 pmoles of both gene-specific primers and 5 ng template RNA. RT-PCR products were analyzed by 1.2% Agarose gel electrophoresis in TBE buffer, and imaged was captured by a gel documentation system (Flour Chem.™ Alpha Innotech, USA). Different gene-specific cDNA bands (*LEA1*:547 bp and *LEA3P*:817 bp), were carefully sliced from the gel using the separate sterile blades, and the DNA sample was gel eluted by using Hiper™ Minispin Gel Extraction Kit.

Cloning of cDNA fragments

Gel eluted *LEA1* and *LEA3* gene-specific cDNA bands were clone in vector pTZ57R/T (2886 bp). Ligation reactions were prepared on ice with InsTAclone™ PCR Cloning Kit, and this ligation mixture was incubated at 4°C overnight. Competent cells of *Escherichia coli* strain JM109 were transformed with this ligation mix and incubated overnight

Table 1 Designing of gene-specific primers for genes encoding for desiccation tolerance proteins

Genes	Source accession No.	Primer	Sequences of primers designed	T annealing °C	cDNA Fragments obtained
LEA1	DQ855277	F	ATGGCGTCCGGTCAGG	60°C	547 bp
		R	CTAGGACTTGGTCCTGAACTTG		
LEA3	NM01111828	F	ATGGCTTCCCACCAGGACAA	57.5°C	817 bp
		R	CTAGTGATCCCTGGTGATGGT		

at 37°C on Luria Bertani (LB) agar plates having ampicillin (100 µg/mL) and nalidixic acid (30 µg/mL). To get distinct single colonies, transformed bacterial colonies were streaked and incubated overnight at 37°C on LB-ampicillin (100 µg/mL) agar plates. Single bacterial colonies from these plates were used for further sub-culturing in a 5 mL LB nalidixic acid broth medium and incubated overnight at 37°C. Transformed *Escherichia coli* culture was pelleted and resuspended in glycerol (0.5 mL of 60% glycerol), LB broth (0.5 mL), and DMSO (90 µL). They were chilled at -196°C in liquid nitrogen and stored at -80°C deep freezer till further work.

Confirmation of cloned fragments by PCR amplification and restriction analysis

Gene JET™ plasmid Miniprep kit (MBI Fermentas Life Science Ltd.) was used for plasmid DNA isolation from transformed *E. coli* (5 mL) culture in LB broth. All recombinant plasmids were double digested with restriction enzymes *EcoRI* and *BamHI* (M/s Bangalore Genei Ltd.) together at 37°C for 1 h. Isolated plasmid DNA of each recombinant clone was also PCR amplified. Products of both, i.e., PCR amplification and restriction digestion of the plasmid, were analyzed on the 1.2% agarose gel electrophoresis (Fig. 1).

Sequencing analysis of cloned and eluted fragments

Custom sequencing was done from M/s Bangalore Genei Ltd. from both directions using universal sequencing primers M13F and M13R for cloned fragments in PCR cloning vector pTZ57R/T (2886 bp) (pTZ57R/T::500, pTZ57R/T::806 bp). Sequenced results were analyzed using ChromasLite 2.01 software. To study the phylogenetic relationship of sequence results generated from the cloning of LEA1 and LEA3 were evaluated by BLAST homology search. Sequence data was submitted at the online site www.ncbi.nlm.nih.gov as BankIT submissions. GenBank accession numbers KJ637318.1 and KT030731.1 were assigned to *LEA1* (547 bp) and *LEA3* (816 bp) was respectively. Consensus motifs from different LEA groups, as previously

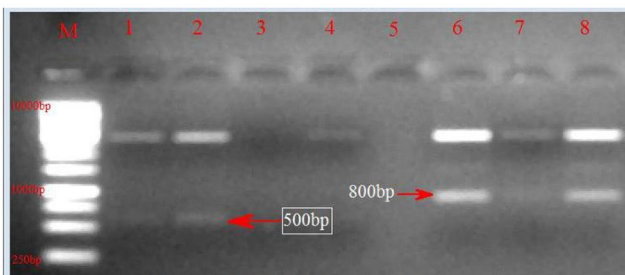


Fig. 1 Restriction digestion of cloned cDNA (*LEA1* 547 bp, *LEA3*:816 bp) into pTZ57R/T (2886 bp) vector

identified (Dure 2001; Battaglia et al. 2008; Pedrosa et al. 2015) were searched to classify the derived LEA protein sequences.

Analysis of sorghum LEA amino acid sequence

General features LEA proteins' such as molecular weight, isoelectric point, amino acid (aa) composition, molecular extinction coefficient, and half-life, were predicted using the ExPASy-ProtParam software (<http://web.expasy.org/protparam/>). Motifs of LEA1 and LEA3 proteins were identified using the MotifScan software (http://myhits.isb-sib.ch/cgi-bin/motif_scan) (Sigrist et al. 2010). Hydrophilicity/hydrophobicity of LEA1 and LEA3 proteins was analyzed using ProtScale (Sweet et al. 1983, Gasteiger et al. 2005).

The secondary structure of LEA1 and LEA3 was predicted using PredictProtein (<https://www.predictprotein.org/>). Then, determination of the topologic structural characteristics of LEA1 and LEA3 proteins was performed using the online service SMART (Letunic et al. 2012), and their 3-dimensional (3D) structure was estimated by modeling using the online service SWISS-MODEL (Bienert et al. 2017). A 3D model of LEA1 and LEA3 was constructed using the program Swiss-PDB Viewer (Guex et al. 2009; Marchler-Bauer et al. 2017) and Phyre2 (Kelley et al. 2015) and assessed using Verify_3D (http://services.mbi.ucla.edu/Verify_3D/). SPIDER² (Sequence-based Prediction of Local and Nonlocal Structural Features for Proteins) (<http://sparks-lab.org/server/SPIDER2/>) was used to predict region forming helical structure.

Results

In the present study, attempts were made to investigate genes encoding for LEA proteins expressed under drought stress in sorghum. Gene specific primers were designed and cloned full-length cDNA of these genes was sequenced for further characterization.

cDNA sequence homology analysis

On the BLAST homology-based megablast alignment, 547 bp *LEA1* insert clone showed complete 100% coverage along with 99% and 95% homology (Gaps=0%) with *sorghum LEA* mRNA, complete cds (DQ855277.1) and *Zea mays LEA1* mRNA, complete cds (EU961259.1), respectively. This confirmed that the 547 bp cloned PCR fragment has the *LEA1* gene's full-length coding sequence.

The BLAST homology search of the 816 bp *LEA3* cDNA insert clone showed 100% query coverage with six *LEA3* gene accessions, including five sorghum *LEA3* genes

(97.55–99.51% homology with no gap) and one *Pogonatherum LEA3* gene (84.8% homology with 7% gaps). It also showed 99.18–100% sequence homology with 75–89% coverage (Gaps = 12%) with three *sorghum LEA3* cds.). It also showed 90.76–93.0% sequence homology with 38–52% coverage (Gaps = 12%) with three of five *Zeamays LEA3* cds; with rest 2 accessions showing 86.3–89.1% sequence homology with 19–36% coverage (Gaps = 0–1%). This confirmed that the 817 bp cloned PCR fragment has the *LEA3* gene's complete coding sequence.

LEA1 and LEA3 protein homology analysis

On blastp analysis, the 181 amino acid LEA1 protein showed complete query coverage in 44 hits, with another 59 entries showing over 90% coverage. Only a single sorghum entry LEA B19.3 showed 69.1% homology (125 out of 181 aa residues) with a 44 amino acid gap (between 44 and 87 aa residues position) and 12 mismatches.

On blastp analysis, the 271 amino acid LEA3 protein showed complete query coverage in 43 hits, of which five were from sorghum that exhibited 73.4 to 74.9% homology. The top five homologous sorghum entries matched for 199–203 aa residues, with all of them showing two-gap regions of 31 (in between 21 and 51 aa residues position) and 37 amino acids gap (in between 186 and 222 aa residues position), respectively. Another sorghum entry showed over 90% coverage and 71.2% homology.

LEA protein conserved domains and motifs analysis

LEA1 protein showed an LEA domain in from 1 to 177 aa residues position with a gap from 48th to 108th aa positions (Suppl. Figure 1a and Fig. 2a). Typical LEA1 specific internal 20 bp sequence (TRKEQ [L/M] G [T/E] EGY [Q/K]

EMGRKGG [L/E]) is present in three copies in between 95 and 154 amino acid positions; with other two LEA1 specific 20-mer motifs present in a single copy each (Table 2). Typical LEA1 N-terminal motif (TVVPGGTGGKSLEAQE[H/N] LAE) was located at 20–39 residue position, while LEA1-C-terminal motif D[K/E]SGGERA[A/E][E/R]EGI[E/D] IDESK[F/Y] was located at 158–177 residue position.

The 271 amino acid LEA3 protein showed two overlapping LEA domains, from 89 to 132 and 122–165 aa residues position (Suppl. Figure 1b and Fig. 2b). Typical LEA3 specific internal 11-mer sequence (ATEAAKQKASE) as described by Dure (2001) is present in five copies in between 100 and 153 amino acid positions; while SYKAG-ETKGRKT motif is present twice along with additional two 11-mer LEA3 specific motifs present in a single copy each. Further, this protein appeared to be of the LEA3-D-7 subgroup (Table 2).

Structural features 181 aa LEA1 protein

On Conserved domain structure analysis of LEA protein, two distinct domains covering 1 to 40 and 109 to 177 amino acid residue regions exhibited LEA 5 superfamily type small hydrophilic seed protein-specific domains with a wide 60 amino acid wide gap in between (Fig. 2a). On Motifs identification using the MotifScan software, it had 7 Casein kinase II phosphorylation sites (each amino acid tetramer); 5 Protein kinase C phosphorylation sites (4 had amino acid trimers), six each of which was six amino acid residue long (at 24–29, 93–98, 109–114, 133–138, 152–157, 161–166 sites), glycine-rich 88–162 aa region, two amidation sites (at 128–131 and 148–151 amino acid residues).

The SWISS MODEL analysis of the LEA1 protein KT030731 (Based on GenBank: KT030731.1) revealed two alternate pairs of LEA5 family domains with common

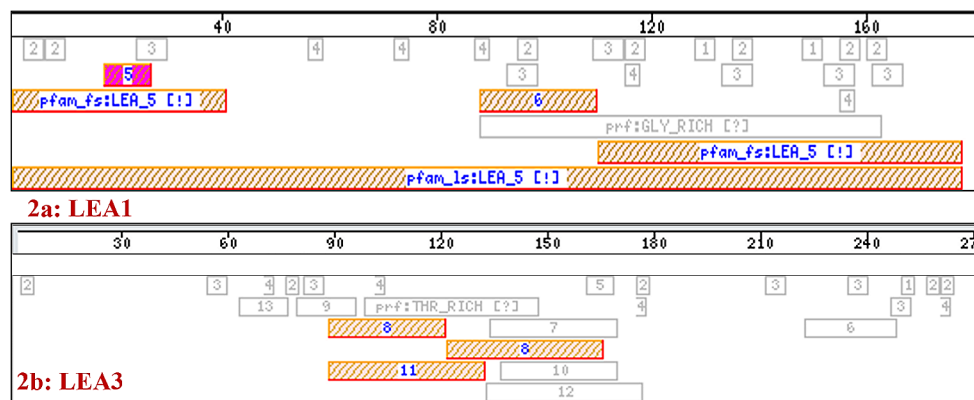


Fig. 2 (a) The conserved motifs with structural features of LEA 1 protein (AID50187) below the bp ruler from the query. Legends of motif scan features: 1: ASN_Glycosylation; 2:CK2_Phospho_Site; 3:Myristyl; 4: PKC_Phospho_Site; 5: TYR_Phospho_Site; 6: CAP160; 7: CRA_rpt; 8: LEA_4 (89–121; 122–165); 9: Oleosin; 10: CRA_rpt; 11: LEA_4 (89–132); 12: LEA_4 (133–176); 13: NUMOD3

(KT030731.1) below the bp ruler from query. Legends of motif scan features: 1: ASN_Glycosylation; 2:CK2_Phospho_Site; 3:Myristyl; 4: PKC_Phospho_Site; 5: TYR_Phospho_Site; 6: CAP160; 7: CRA_rpt; 8: LEA_4 (89–121; 122–165); 9: Oleosin; 10: CRA_rpt; 11: LEA_4 (89–132); 12: LEA_4 (133–176); 13: NUMOD3

Table 2 Evaluation of consensus motifs of amino acids (aa) sequences from different LEA groups

Group	Motif	Consensus sequence	Homology with 181aa LEA protein	Homology with 271 aa LEA protein
LEA1	1	TVVPGGTGGKSLEAQE[H/N]LAE	YES (20–39)	No
D-19	2	TRKEQ [L/M] G [T/E] EGY [Q/K] EMGRKGG [L/E])	YES (95–113, 115–133, 135–154)	No
	3	D[K/E]SGGERA[A/E][E/R]EGI[E/D]IDESK[F/Y]	YES (158–177)	No
LEA2	K	EKKGIMDKIKEKLPG	No	No
	S	LHRSGSWSSSSDDDD	No	No
	Y	R [V/T]D[E/Q]YGNPVH	No	No
LEA3	1	GGVLQQTGEQV	No	No
D-7	2	AADAVKHTLGM	No	Yes (235–245)
	3*	(TAQ [A/S] AK [D/E] KT[S/Q] E)	No	YES(64–74)
	5*	ATEAAKQKASE	No	Yes (111–120, 122–131, 133–142, 144–153, 100–109)
	4	SYKAGETKGRKT	No	Yes (9–18, 94–99)
LEA3	1*	TAEKAGEYKDY	No	No
D-29	4*	TVEKAKEAKDT	No	NO
	2*	AYEKAGSAKDM	No	NO
	3*	AAQKAKDYAGD	No	NO
	5	ESWTEWAKEKI	No	NO
LEA4	1	AQEKAEMTA[R/HDPXKEMAHERR[E/K][A/E][K/R]	No	YES (74–93)
D-113	2	MQSAKEKAASNMAASAKAGMEKTKAK	No	YES(112–135)
	3	EAEMDKHQAKAHAAEKQ	No	NO
	4	PTGTHQMSALPGHGTGQPTGHVVEG	No	YES(98–111)
LE6	1	LEDYKMQGYGTQGHQPKPGRG	No	No
	2	GSTDAPTLGGAV	YES(18–28)	No
	3	TDAINRHGVP	No	No
	4	GLPTETSPYV	No	No
LEA7	1	AAGAYALHEKHKAKKDPEHAHRHKI	No	No
	2	EIAAAAAVGAGGFAPHEHHEKKDEAK	No	No
	3	DYKKEEKHHKHMEHLGELGAV	No	No
	4	HHHHHLFHHHKD	No	No
	5	EEEEEAHGKHHHLF	No	No

*11mers described by Dure 2001; rest all from Battaglia et al. 2008, Pedrosa et al. 2015

Domain-I (from 1 to 47 position); along with either Domain-II-A of 55 aa (from 80 to 134 position) or Domain-II-B of 60 aa (from 118 to 177 position). There are two internal repeats of 26 amino acids, from position 92–117 and 132 to 157. The protein structure is highly complex, with low complexity region only in between 158 and 172 residue position. On 3D Structure analysis of protein using Pyre software, 94% of the LEA1 protein couldn't be meaningfully predicted and is highly speculative as 66% of the protein's sequence was predicted disordered (Fig. 3a and b).

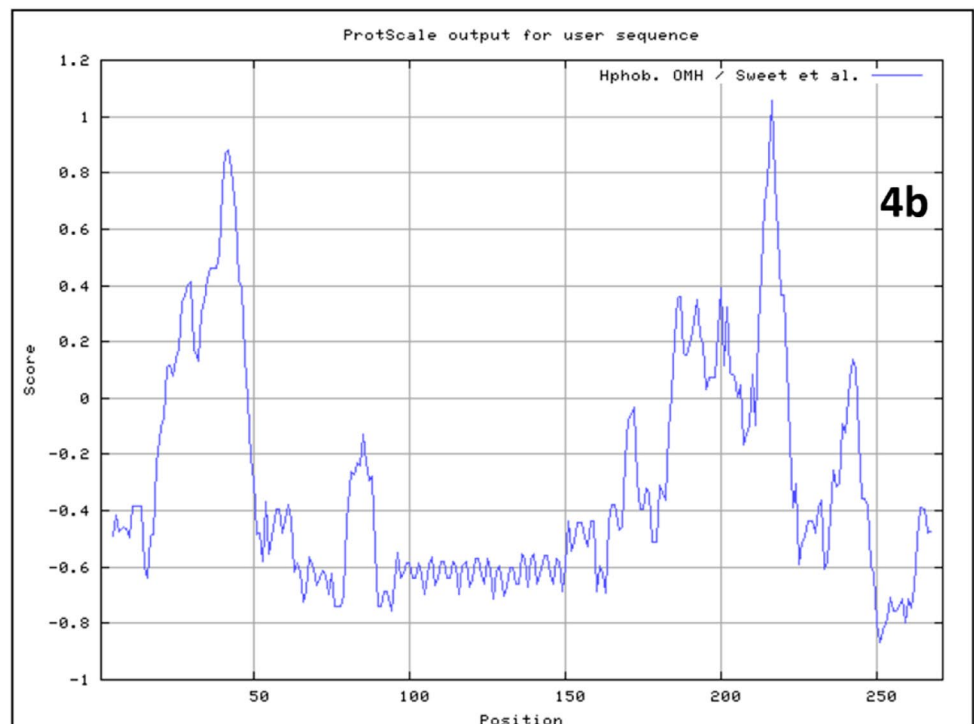
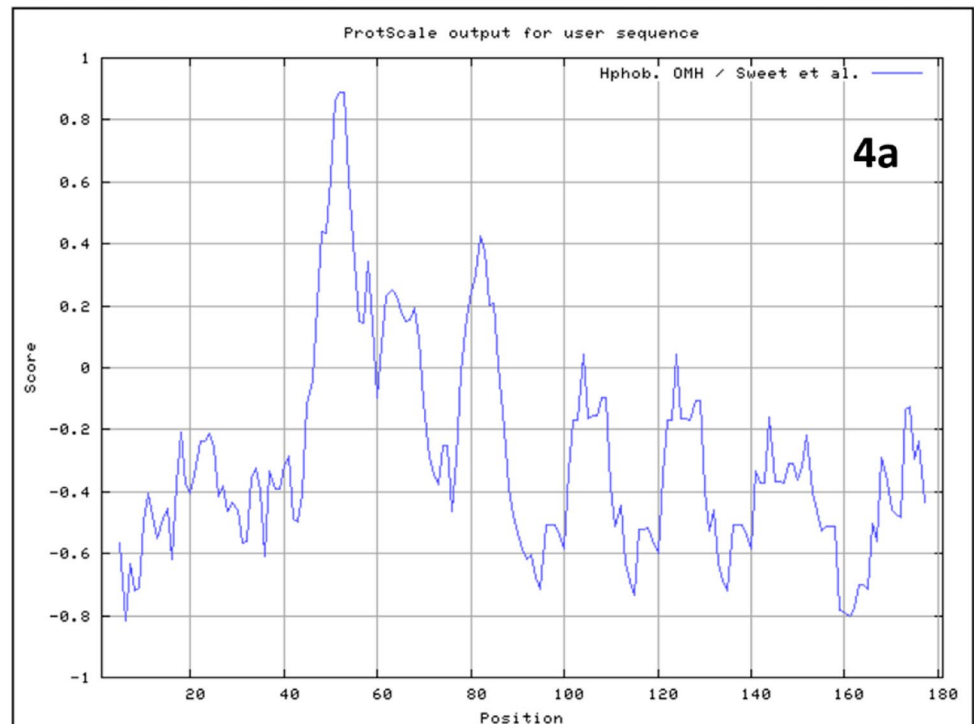
Structural features 271 aa LEA3 protein

On Conserved domain structure analysis of LEA3 protein, two partially overlapping domains viz. 89 to 132 and 122 to 165 amino acid residue regions exhibited LEA 4 superfamily-specific protein-specific domains (Fig. 2b). On Motifs identification using the MotifScan software, it had 5 Protein kinase C phosphorylation sites (4 had amino acid trimers),

single Tyrosine kinase phosphorylation sites at the end of second LEA domain (in between 161 and 168 amino acid residues), 5 myristoylation sites outside LEA region. The 5 myristoylation site was six amino acid residue long (at 55–60, 82–87, 211–216, 234–239, 246–251 aa positions), threonine rich 99–147 aa region, asparagine glycosylation site at 249–253 amino acid residues. It has an oleosin site at 80–96 aa region (that has a role in membrane desiccation tolerance), CRA_rpt site at 137–169 aa, CAP at 222–247 aa site.

The SWISS_MODEL analysis of the protein KT030731 generated 7 alternate models with most exhibiting helical structure (Fig. 3c and d). There is a transmembrane domain-like region from position 26–48. The protein structure has a low complexity region only in 103–165 and 248–255 residue positions. On 3D Structure analysis of protein using Pyre² software, 94% of the LEA3 protein formed an alpha-helical structure (which excluded both terminal regions); with 26% of protein being disordered mostly in the terminal

Fig. 4 (a) Prediction of hydrophobic domains of LEA 1 protein (AID50187); (b) Prediction of hydrophobic domains of LEA 3 protein (KT030731.1)



against 35 positively charged residues (Arg+Lys). LEA2 protein had an aliphatic index of 57.79 with a grand average of hydropathicity (GRAVY) value of -0.738 as per ProtParam (Fig. 4b). This protein was classified as stable as its instability index (II) was 37.61.

Discussion

In the climate change scenario, water stress will be a major factor limiting crop production in this century. Transcriptome profiling of drought-stressed sorghum plants from various stages revealed the complex nature of drought

stress (Varoquaux et al. 2019). The expression of 10,727 genes being significantly influenced, of which 75% reverted to their original expression levels; if stress is removed (Varoquaux et al. 2019).

LEA or dehydrins are small, extremely hydrophilic proteins having a role in imparting desiccation tolerance in plants (Close 1996; Cuming 1999; Nagaraju et al. 2019). LEA proteins are protective proteins that prevent the denaturation of cellular components, crystallization induced damages (Abdul et al. 2021) and stabilize membrane and proteins against dehydration (Fracasso et al. 2016). They accumulate in the vegetative tissues and recalcitrant seeds, where they protect the phospholipid membranes. Grasses tolerate desiccation stress during growth through various conserved mechanisms that are shared with the seed dehydration process, with an overlapping pattern of the LEA expression under both conditions (Pardo et al. 2020). Based on the high polar amino acid residues, they are thought to provide preferential hydration to intracellular macromolecules (Damame et al. 2016). On further dehydration, LEAs would provide a layer of their own hydroxylated residues to interact with the surface proteins, acting as replacement water (Close 1996).

The LEA genes have been classified into eight groups based on their conserved domains and homology analysis (Hunault and Jaspard 2010). However, 32 maize LEA genes were classified into nine groups, which were distributed throughout the genome via transposition (Li and Cao 2016). Nagaraju et al. (2019) further identified a total of 68 LEA genes in sorghum, which are evenly distributed on all ten chromosomes, with chromosomes 1, 2, and 3 being the hot spots. The majority of the *sorghum* LEA genes were intronless or have fewer introns. The majority of the LEA proteins are basic, with chloroplast sub-cellular localization. Their promoter analysis revealed the presence of abiotic stress-responsive, biotic stress-responsive, hormone-responsive, and development-responsive *cis*-elements. Gene expression analysis revealed their tissue-specific expression, with higher expression noticed in stems than roots and leaves. Most the LEA family members were up-regulated at least in one tissue under different stress conditions (Nagaraju et al. 2019).

In the present study, the potential sorghum LEA1 (543 bp) and LEA3 (817 bp) cDNAs were cloned, and their physiological and biochemical characters were predicted. They were found to encode for a protein with 181 (19.65 kD) and 271 amino acids, matching LEA1 and LEA3 consensus sequence motifs, respectively. LEA1 and LEA3 proteins are characterized by a repeating motif of 20-mer and 11-mer amino acids, respectively (Amara et al. 2014). LEA1 proteins are seed embryo-specific, while LEA3 proteins are abscisic acid (ABA) inducible during specific

developmental stages under stress conditions that interact with membranes during desiccation (Amara et al. 2014).

In sorghum LEA1, there is high glycine content (15.5%) and charged amino acids (27.6%), with a coiled structure that agrees to typical LEA1 description as reviewed by Battaglia et al., (2008). As per Zhang et al. (2014), LEA 1 protein (AID50187) consisted of 181 amino acids long (19.65 kD size), with two distinct LEA5 type domains with 60 amino acid wide gap. It had a glycine-rich nonapeptide LEA signature in the N-terminal region with a consensus pattern. It had a highly disordered/complex structure lacking any secondary structure, comprising two internal 26 amino acids repeats. LEA 3 proteins are mostly devoid of secondary structure, being largely in a random coil conformation in solution (Gasteiger et al. 2005). This protein had an aliphatic index of 57.18 with a hydropathicity (GRAVY) value of -0.865. The LEA 3 protein (KT030731.1) consisted of 271 amino acids long (29.2 kD size) with an alpha-helix structure with two partially overlapping LEA 4 type domains and transmembrane domain. LEA2 protein had an aliphatic index of 57.79 with a hydropathicity (GRAVY) value of -0.738.

Although both these LEA proteins are rich in hydrophilic amino acids with two LEA domains; however, they were distinct in the sense of being wide apart vs. overlapping; helical vs. disordered in their structure. A high portion of random coil structures affects their exceptional water binding capacity, and the conserved segments give rise to amphipathic α -helices, which form lipid-binding domains and thus can associate with and protect lipid aggregates and hydrophobic domains of proteins. This suggests that LEA1 protein may serve as a dehydrin to protect cells from stress.

The LEA proteins comprise a diverse collection of multifunctional proteins that play a major role in desiccation tolerance and seed longevity. Each LEA protein group or family harbors one or more copies of unique domains. In a study, transgenic tobacco and maize plants expressing maize *LEA14tv* gene exhibited enhanced drought tolerance (Minh et al. 2019). Understanding these genes will provide further opportunities to elucidate the underlying molecular mechanism involved in drought tolerance in sorghum. Therefore, more research is needed to identify the contribution of these LEA proteins towards drought tolerance; and to decide whether they could be used for imparting drought tolerance.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42535-024-00856-w>.

Acknowledgements Authors acknowledge the Mahatma Phule Krishi Vidyapeeth, Rahuri, India, for providing the necessary infrastructure and facilities to carry out this work. Financial help received from National Agricultural Innovation Project, received under the research

project “Allele mining for abiotic stress tolerance -sorghum” is also acknowledged.

Declarations

Conflict of interest The authors declare no compete of interests.

Ethics approval and consent Not applicable.

Consent for publication All authors agree to publish.

References

- Abdel-Ghany SE, Ullah F, Ben-Hur A, Reddy AS (2020) Transcriptome analysis of drought-resistant and drought-sensitive sorghum (*Sorghum bicolor*) genotypes in response to PEG-induced drought stress. *Int J Mol Sci* 21(3):772. <https://doi.org/10.3390/ijms21030772>
- Abdul Aziz M, Sabeem M, Mullath SK, Brini F, Masmoudi K (2021) Plant Group II LEA proteins: intrinsically disordered structure for multiple functions in response to environmental stresses. *Biomolecules* 11(11):1662. <https://doi.org/10.3390/biom11111662>
- Ali MA, Jabran K, Awan SI, Abbas A, Zulkiffal M, Acet T, Farooq J, Rehman A (2011) Morpho-physiological diversity and its implications for improving drought tolerance in grain sorghum at different growth stages. *Aust J Crop Sci* 5:311
- Amara I, Zaidi I, Masmoudi K, Ludevid MD, Pagès M, Goday A, Brini F (2014) Insights into late embryogenesis abundant (LEA) proteins in plants: from structure to the functions. *Am J Plant Sci* 5(22):3440–3455. <https://doi.org/10.4236/ajps.2014.522360>
- Badigannavar A, Teme N, de Oliveira AC, Li G, Vaksman M, Viana VE, Ganapathi TR, Sarsu F (2018) Physiological, genetic and molecular basis of drought resilience in sorghum [*Sorghum bicolor* (L.) Moench]. *Ind J Plant Physiol* 23:670–688. <https://doi.org/10.1007/s40502-018-0416-2>
- Battaglia M, Olvera-Carrillo Y, Garciarubio A, Campos F, Covarrubias AA (2008) The enigmatic LEA proteins and other hydrophilins. *Plant Physiol* 148(1):6. <https://doi.org/10.1104/pp.108.120725>
- Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, Schwede T (2017) The SWISS-MODEL repository - new features and functionality. *Nucleic Acids Res* 45(D1):D313. <https://doi.org/10.1093/nar/gkw1132>
- Close TJ (1996) Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiol Plant* 97(4):795. <https://doi.org/10.1111/j.1399-3054.1996.tb00546.x>
- Cuming A (1999) In: Casey R, Shewry PR, Proteins S (eds) LEA proteins. Kluwer Academic, Dordrecht, pp 753–780. https://doi.org/10.1007/978-94-011-4431-5_32
- Damame SV, Naik RM, Chimote VP, Munjal SV (2016) Molecular analysis of rabi sorghum genotypes differing in osmolytes accumulation under water stress. *Int J Bio-Resour Stress Manag* 7(5):11207. <https://doi.org/10.23910/IJBSM/2016.7.5.1544>
- Dure III, L (2001) Occurrence of a repeating 11-mer amino acid sequence motif in diverse organisms. *Protein Pept Let* 8(2):115. <https://doi.org/10.2174/0929866013409643>
- Fracasso A, Trindade LM, Amaducci S (2016) Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE. *BMC Plant Biol* 16(1):1. <https://doi.org/10.1186/s12870-016-0800-x>
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy server. In: Walker JM (ed) *The Proteomics protocols Handbook*. Springer protocols handbooks. Humana. <https://doi.org/10.1385/1-59259-890-0:571>
- Guex N, Peitsch MC, Schwede T (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-Pdb-Viewer: a historical perspective. *Electrophoresis* 30(S1):S162. <https://doi.org/10.1002/elps.200900140>
- Hunault G, Jaspard E (2010) LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genom* 11(1):1–9. <https://doi.org/10.1186/1471-2164-11-221>
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–858. <https://doi.org/10.1038/nprot.2015.053>
- Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40(D1):D302–D305. <https://doi.org/10.1093/nar/gkr931>
- Li X, Cao J (2016) Late Embryogenesis Abundant (LEA) gene family in maize: identification, evolution, and expression profiles. *Plant Mol Biol Rep* 34:15–28. <https://doi.org/10.1007/s11105-015-0901-y>
- Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45(D1):D200–D203. <https://doi.org/10.1093/nar/gkw1129>
- McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, Mattison A (2018) The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J* 93(2):338. <https://doi.org/10.1111/tpj.13781>
- Michel BE, Kaufmann MR (1973) The osmotic potential of polyethylene glycol 6000. *Plant Physiol* 51(5):914–916. <https://doi.org/10.1104/pp.51.5.914>
- Minh BM, Linh NT, Hanh HH, Hien LT, Thang NX, Hai NV, Hue HT (2019) A LEA gene from a Vietnamese maize landrace can enhance the drought tolerance of transgenic maize and tobacco. *Agronomy* 9(2):62. <https://doi.org/10.3390/agronomy9020062>
- Nagaraju M, Kumar SA, Reddy PS, Kumar A, Rao DM, Kavi Kishor PB (2019) Genome-scale identification, classification, and tissue specific expression analysis of late embryogenesis abundant (LEA) genes under abiotic stress conditions in *Sorghum bicolor* L. *PLoS ONE* 14(1):e0209980. <https://doi.org/10.1371/journal.pone.0209980>
- Ogbaga CC, Stepien P, Dyson BC, Rattray NJW, Ellis DI, Goodacre R (2016) Biochemical analyses of sorghum varieties reveal differential responses to drought. *PLoS ONE* 11(5):e0154423. <https://doi.org/10.1371/journal.pone.0154423>
- Oliver MJ, Farrant JM, Hilhorst HW, Mundree S, Williams B, Bewley JD (2020) Desiccation tolerance: avoiding cellular damage during drying and rehydration. *Annu Rev Plant Biol* 29(71):435–460. <https://doi.org/10.1146/annurev-arplant-071219-105542>
- Ongom PO, Volenec JJ, Ejeta G (2016) Selection for drought tolerance in sorghum using desiccants to simulate post-anthesis drought stress. *Field Crops Res* 198:312–321. <https://doi.org/10.1016/j.fcr.2016.03.015>
- Pardo J, Wai CM, Chay H, Madden CF, Hilhorst HW, Farrant JM, VanBuren R (2020) Intertwined signatures of desiccation and drought tolerance in grasses. *Proc Natl Acad Sci USA* 117(18):10079–10088. <https://doi.org/10.1073/pnas.2001928117>
- Pedrosa AM, Martins CDPS, Gonçalves LP, Costa MGC (2015) Late embryogenesis abundant (LEA) constitutes a large and diverse family of proteins involved in development and abiotic stress responses in sweet orange (*Citrus sinensis* L. Osb). *PLoS ONE* 10(12):pe0145785. <https://doi.org/10.1371/journal.pone.0145785>

- Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002) Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L. Moench). *Plant Mol Biol* 48:713–726. <https://doi.org/10.1023/A:1014894130270>
- Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* 38:D161–D166. <https://doi.org/10.1093/nar/gkp885>
- Sweet RM, Eisenberg D (1983) Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *J Mol Biol* 171(4):479–488. [https://doi.org/10.1016/0022-2836\(83\)90041-4](https://doi.org/10.1016/0022-2836(83)90041-4)
- Varoquaux N, Cole B, Gao C, Pierroz G, Baker CR, Patel D, Madera M, Jeffers T, Hollingsworth J, Sievert J, Yoshinaga Y (2019) Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. *Proc Natl Acad Sci USA* 116(52):27124–27132. <https://doi.org/10.1073/pnas.1907500116>
- Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. *Curr Opin Biotechnol* 16(2):123–132. <https://doi.org/10.1016/j.copbio.2005.02.001>
- Zhang NZ, Huang SY, Zhou DH, Xu Y, He JJ, Zhu XQ (2014) Identification and bioinformatic analysis of a putative calcium-dependent protein kinase (CDPK6) from *Toxoplasma Gondii*. *Genet Mol Res* 13(4):10669–10677. <https://doi.org/10.4238/2014>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.