

RESEARCH ARTICLE

Genome-wide identification and expression analysis of late embryogenesis abundant protein-encoding genes in rye (*Secale cereale* L.)Mengyue Ding¹ , Lijian Wang^{1,2} *, Weimin Zhan¹, Guanghua Sun¹, Xiaolin Jia¹, Shizhan Chen¹, Wusi Ding¹, Jianping Yang¹***1** College of Agronomy, Henan Agricultural University, Zhengzhou, China, **2** Department of Criminal Science and Technology, Henan Police College, Zhengzhou, China These authors contributed equally to this work.* jpyang@henau.edu.cn (JY); wanglijian2012@163.com (LW) OPEN ACCESS**Citation:** Ding M, Wang L, Zhan W, Sun G, Jia X, Chen S, et al. (2021) Genome-wide identification and expression analysis of late embryogenesis abundant protein-encoding genes in rye (*Secale cereale* L.). PLoS ONE 16(4): e0249757. <https://doi.org/10.1371/journal.pone.0249757>**Editor:** Aimin Zhang, Institute of Genetics and Developmental Biology Chinese Academy of Sciences, CHINA**Received:** January 21, 2021**Accepted:** March 24, 2021**Published:** April 8, 2021**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0249757>**Copyright:** © 2021 Ding et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.**Data Availability Statement:** All rye RNA-Seq files are available from the NCBI Sequence Read Archive Database (accession number SRP293835).

Abstract

Late embryogenesis abundant (LEA) proteins are members of a large and highly diverse family that play critical roles in protecting cells from abiotic stresses and maintaining plant growth and development. However, the identification and biological function of genes of *Secale cereale* LEA (*ScLEA*) have been rarely reported. In this study, we identified 112 *ScLEA* genes, which can be divided into eight groups and are evenly distributed on all rye chromosomes. Structure analysis revealed that members of the same group tend to be highly conserved. We identified 12 pairs of tandem duplication genes and 19 pairs of segmental duplication genes, which may be an expansion way of *LEA* gene family. Expression profiling analysis revealed obvious temporal and spatial specificity of *ScLEA* gene expression, with the highest expression levels observed in grains. According to the qRT-PCR analysis, selected *ScLEA* genes were regulated by various abiotic stresses, especially PEG treatment, decreased temperature, and blue light. Taken together, our results provide a reference for further functional analysis and potential utilization of the *ScLEA* genes in improving stress tolerance of crops.

Introduction

Plants are affected by a variety of biotic and abiotic stresses during their lifetime due to their inability to escape harmful environmental conditions. Late embryogenesis abundant (LEA) proteins are highly hydrophilic proteins with antioxidant and ion binding properties that function in membrane and protein stabilization, osmotic regulation, and as a protective buffer against dehydration [1]. Thus, they play important roles in protecting cells from abiotic stress, and in normal growth and development. LEA proteins were first isolated from cotyledons of cotton (*Gossypium hirsutum* L.) at late embryonic development stages [2], and in recent decades, have also been found in species from algae to angiosperms, such as cyanobacteria [3], *Arabidopsis thaliana* [4], rice (*Oryza sativa* L.) [5], wheat (*Triticum aestivum* L.) [6], maize (*Zea mays* L.) [7], and in prokaryotes and invertebrates, such as rotifers [8,9]. In plants, *LEA*

Funding: This study was funded by the National Natural Science Foundation of China to JP (General Program, 31871709).

Competing interests: The authors have declared that no competing interests exist.

genes are highly expressed at the late stage of seed maturation [2] and LEA proteins accumulate in roots, stems, and other organs throughout the plant growth duration [10].

In higher plants, based on the similarity of amino acid sequences and differences in conserved domains, the LEA family is generally divided into eight groups: LEA1, LEA2, LEA3, LEA4, LEA5, LEA6, Dehydrin, and seed maturation protein (SMP) [11]. In recent studies abscisic acid (ABA) stress ripening (ASR) proteins were classified as LEA7 group proteins [12,13], extending the LEA family to nine groups. The division of LEA family members varies among species [14]. Most *LEA* genes encode small proteins with molecular weights ranging from 10 to 30 kDa, and they are predominantly composed of a repeating arrangement of hydrophilic amino acids that form a highly hydrophilic structure [15]. Most LEA proteins have low hydrophobicity and a large net charge, which allow them to be “completely or partially disordered” [16,17], these proteins may act as a novel form of molecular chaperone, or ‘molecular shield’, to help prevent the formation of damaging protein aggregates during water stress [18]. Each group of LEA proteins has its own unique conserved motif [19]. For example, LEA1 proteins contain a 20-amino-acid motif (GGETRKEQLGEEGYREMGRK) [20]. Dehydrin proteins contain a motif called the K-segment (EKKGIMDKIKEKLP) [21,22]. LEA4 proteins have a conserved domain of 11-amino-acid sequences (TAQAAKEKAGE) [21]. These conserved sequences (motifs) have been preserved through long-term evolution, and play an important role in plant responses to environmental stress.

Previous studies have shown that LEA proteins are ubiquitous and localized in the cytoplasm, nucleus, chloroplast, mitochondria, and endoplasmic reticulum [23]. Thus, not all LEA proteins are localized in the same part of the cell, and that their particular functions depend on their intra-cellular locations. The expression of LEA proteins is often induced by abiotic stresses such as drought, heat, cold, and exogenous hormone at different development stages and tissues of plants [1,4]. Moreover, overexpression of the *LEA* genes can improve stress tolerance of transgenic plants. For example, overexpression of the barley (*Hordeum vulgare* L.) *HVA1* gene promoted drought and salt stress tolerance in rice [24] and wheat [25]. Overexpression of the wheat *DHN-5* enhanced osmotic stress tolerance in *Arabidopsis* [26], and activated pathogenesis-related protein expression [27]. The above studies demonstrate that the *LEA* genes could potentially be used to improve the abiotic stress tolerance of crops.

Rye (*Secale cereale* L.) is a diploid Triticeae species closely related to wheat and barley, and an important crop for feed and food [28]. Rye has excellent disease and stress resistance, and it has been used as a source to improve wheat resistance to pathogens in the past few decades [29]. Rye-wheat chromosome translocation has been widely used in wheat breeding around the world, and translocation lines have notably increased yield [30,31]. Studies on LEA proteins in wheat and barley have been reported [6,25]. However, the distribution and function of LEA proteins in rye have been rarely reported [32]. In this study, we identified 112 *ScLEA* genes based on the rye genome sequence. The phylogenetic analysis, structural characterizations, evolutionary relationships, and expression profiles of the *ScLEA* genes were systematically analyzed. These results will contribute to our understanding of the LEA family in rye and to the utilization of these *LEA* genes in homologous or heterologous systems.

Materials and methods

Identification of *LEA* genes in the rye genome

The nucleotide and protein sequences, as well as the gene annotation files, were downloaded from the rye genome database (<http://pgsb.helmholtz-muenchen.de/plant/rye/index.jsp>). The HMM profiles (LEA1: PF03760, LEA2: PF03168, LEA3: PF03242, LEA4: PF02987, LEA5: PF00477, LEA6: PF10714, Dehydrin: PF00257, SMP: PF04927) were downloaded from the

Pfam database (<http://pfam.sanger.ac.uk/>). Then, HMMER 3.0 was used to search the encoded protein sequences with the default parameters and a filter threshold of 0.01. In addition, the obtained amino acid sequences were used as queries in searches against both the CDD database (<https://www.ncbi.nlm.nih.gov/cdd/>), and SMART database (<http://smart.embl-heidelberg.de/>), and the repeated or non-LEA domain sequences were eliminated manually. Finally, the identified rye LEA proteins were named according to the group and gene ID order. The physical and chemical properties of ScLEA proteins were analyzed using the ProtParam online tool (<https://web.expasy.org/protparam/>).

Phylogenetic analysis of LEA proteins in rye

To investigate the evolutionary relationships of LEA proteins in rye, multiple alignments of the full-length protein sequences were performed using MAFFT [33] (<http://mafft.cbrc.jp/alignment/software/>) with the default parameters. A phylogenetic tree was constructed based on these alignments using the Neighbor-Joining (NJ) method with a bootstrap test of 1000 replicates for assessing internal clade reliability using MEGA X software [34]. The Evolview website (<https://www.evolgenius.info/evolview/#login>) was used to modify the phylogenetic tree.

Analysis of gene structures and conserved domains of LEA genes in rye

The exon-intron structures of the ScLEA family genes were determined by aligning the coding sequences with the corresponding genomic sequences, and visualized using the online software GSDS (<http://gsds.gao-lab.org/index.php>). Conserved domains of the ScLEA proteins were predicted using MEME (<http://meme-suite.org/index.html>).

Distribution of LEA genes on rye chromosomes

ScLEA genes were mapped on rye chromosomes according to the positional information from the rye genome annotation database, and the chromosome physical location map was displayed using Mapchart [35] software. According to Holub [36], chromosomal region of 200 kb containing two or more genes are defined as a tandem duplicated event. The Multiple Collinearity Scan toolkit (MCScanX) [37] with default parameters was used to analyze gene duplication events.

Expression profile analysis of rye LEA genes

The rye RNA-Seq datasets used for generating gene expression levels were downloaded from the NCBI Sequence Read Archive Database (accession number SRP293835) (<https://www.ncbi.nlm.nih.gov/sra/?term=SRP293835>). The transcripts of the datasets were aligned, followed by merging and removal of redundant sequences using Hisat (<http://ccb.jhu.edu/software/hisat/index.shtml>) (version 2.0.4) and Stringtie (<http://ccb.jhu.edu/software/stringtie/>) (version 1.2.3). The software featureCounts [38] was used to calculate the number of reads and normalized gene expression levels. Expression heatmaps were generated using TBtools software [39].

Plant materials and stress treatments

For abiotic stress treatments, rye seedlings were grown in hydroponics with a temperature (18°C) for 10 days, and then treated with 20% (w/v) PEG6000, 200 mM NaCl, 100 mM mannitol, 100 μM ABA, or temperatures of 0°C or 4°C. Leaves were collected after 0 h, 3 h, 6 h, 9 h, 12 h, and 24 h. Leaf samples were treated with sterile water and similarly collected as a control. For different light treatments, seedlings were grown in the dark for seven days and subsequently transferred to far-red (FR) light (5 μmol·m⁻²·s⁻¹), red (R) light (17.56 μmol·m⁻²·s⁻¹),

blue (B) light ($13 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), or white (W) light ($85 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) for 4 h. For each sample, three leaves treated in parallel represented three biological replicates. All treated tissue samples were stored at -80°C for subsequent analysis.

RNA extraction and gene expression analysis

Total RNA was extracted from the leaves of rye using Easstep® Super RNA Kit (Promega). The GoScript™ Reverse Transcriptase Kit (Promega) was used to reverse-transcribe RNA into cDNA. Quantitative real-time PCR (qRT-PCR) was carried out in the Roche Lightcycler® 480 instrument using SYBR Green (TaKaRa). The *ScActin* gene was used as an internal control. The thermal cycler program used was 95°C for 10 min, and followed by 45 cycles of 95°C for 10 s, 60°C for 10 s, and 72°C for 20 s. Each reaction was performed in three technical replicates, and the data from qRT-PCR were calculated using the $2^{-\Delta\Delta\text{Ct}}$ method [40]. The significant differences between data were calculated using Student's *t*-test or two-way analysis of variance (ANOVA), and indicated with asterisks (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Sequences of the primers used in this study are shown in S1 Table.

Cis-acting element analysis

The Plant CARE website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) was used to predict *cis*-acting elements in the regions 1000 bp upstream of the gene start codon and a diagram was visualized using TBtools software [39].

Results

Genome-wide identification and phylogenetic analysis of *ScLEA* genes

Based on genome-wide blast searches, 112 *ScLEA* genes were identified from the rye genome database. The identity of the sequences was verified by checking for the presence of conserved domains using the CDD and SMART tools. These 112 genes were divided into eight groups (LEA1–LEA6, Dehydrin, and SMP) based on the phylogenetic analysis (Fig 1). The largest group was LEA2, which contained 56 members, while the smallest groups, LEA4 and LEA6, had only two members each. Groups Dehydrin, LEA1, LEA5, SMP, and LEA3 contained 18, 14, 8, 7, and 5 members, respectively.

The physical and chemical parameters of the 112 *ScLEA* proteins were calculated using the ProtParam online tool (S2 Table). The smallest *ScLEA* protein had 78 amino acid residues with a molecular weight of 8.24 kDa (*ScLEA*5-8) and the largest had 965 amino acid residues with a molecular weight of 99.64 kDa (*ScDehydrin*-12). The isoelectric point (pI) values ranged from 4.22 (*ScSMP*-6) to 11.07 (*ScLEA*2-17), with an average of 8.37. Grand average of hydropathicity (GRAVY) index analysis indicated that most of the *ScLEA* proteins were hydrophilic (S2 Table). However, the *ScLEA*2 group was atypical with 67.8% of proteins were predicted as hydrophobic. The aliphatic index reflects the thermal stability of the protein.

Phylogenetic analysis revealed two major clades of the *ScLEA* family (Fig 1). The LEA2 group proteins were clustered in one clade, while the proteins in the other seven groups formed another clade. According to the evolutionary relationships, one clade including the Dehydrin group (18 numbers) and LEA5 group proteins (8) was closely related to the LEA4 (2), LEA1 (14), and SMP (7) groups, but distant from the LEA6 (2) and LEA3 (5) groups.

Structural characterization of *ScLEA* genes

The structure of genes is important for determining their expression and function. Structural analysis revealed that the *ScLEA* genes contain few introns; 42% of *ScLEA* genes have one

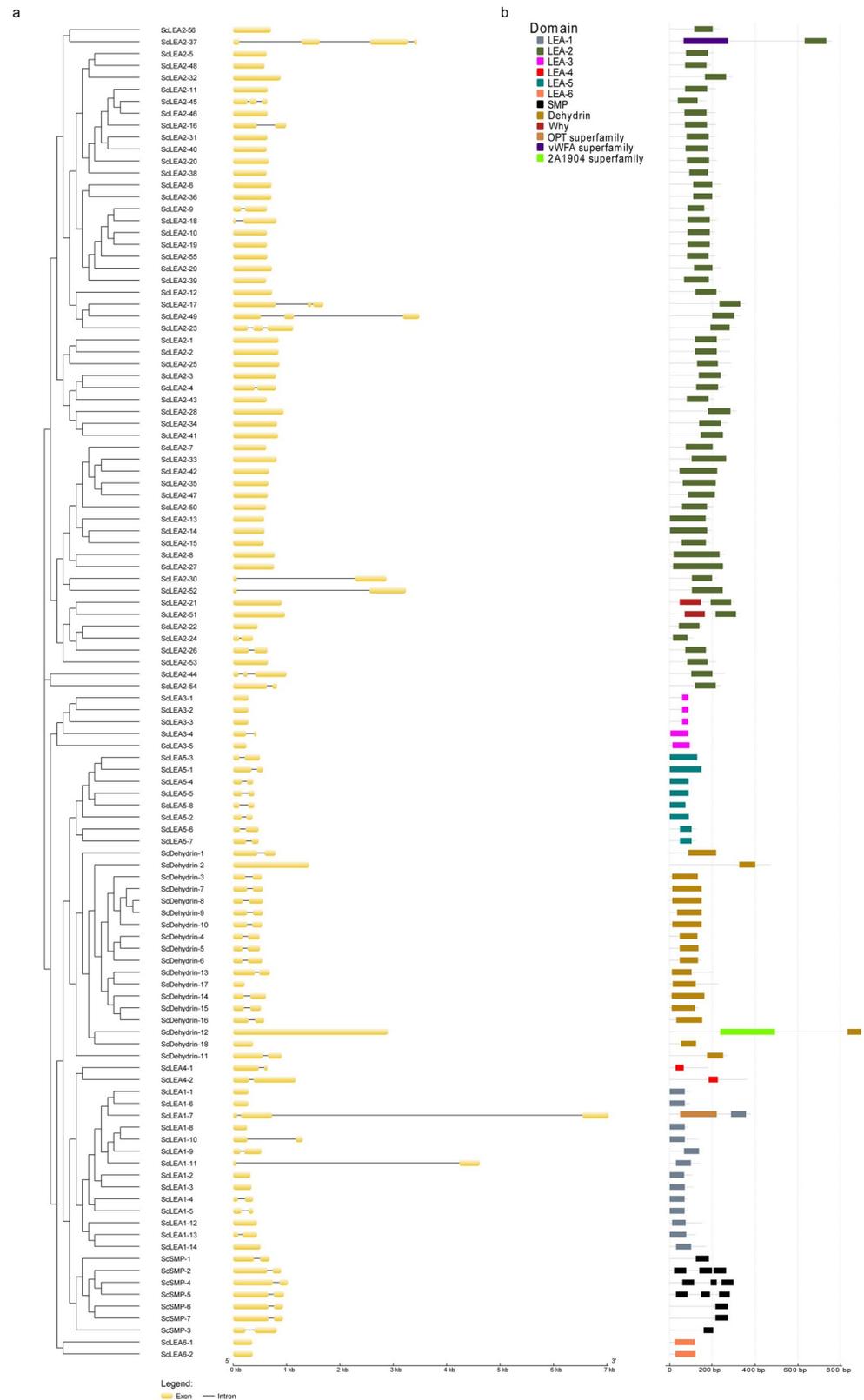


Fig 2. Schematic representation of exon-intron structures of ScLEA genes and conserved domains of ScLEA proteins. (a) The exon-intron structures were examined using the GSDS online tool. The yellow boxes and black lines

represent exons and introns, respectively. (b) The ScLEA protein domains were predicted using the MEME online tool.

<https://doi.org/10.1371/journal.pone.0249757.g002>

Chromosomal locations and gene duplication of ScLEA genes

The distribution of ScLEA genes was determined by mapping their positions on the rye chromosomes. The 112 ScLEA genes were distributed on all seven chromosomes, and their distribution was nearly even among different chromosomes (Fig 3). Chromosome 5 had the most ScLEA genes (17.8%, 20 genes), and chromosome 7 had the fewest (12.5%, 14 genes). Chromosomes 1, 2, 3 and 6 each contained 15 genes. Moreover, genes of the largest group, for example, ScLEA2 group, are distributed across all chromosomes to ensure maximum functionalization. However, other LEA groups have a limited distribution and are mainly found on specific chromosomes.

High-density ScLEA gene clusters were identified in certain chromosomal regions. Twenty-eight ScLEA genes (*ScLEA5-1/2*, *ScLEA5-4/5*, *ScLEA2-9/10*, *ScLEA2-13/14/15*, *ScLEA2-18/19*, *ScLEA1-2/3*, *ScLEA3-2/3*, *ScLEA2-31/32*, *ScSMP-1/2*, *ScDehydrin-7/8/9/10*, *ScDehydrin-12/13*, and *ScDehydrin-14/15/16*) were clustered into 12 tandem duplicated regions on chromosomes 1, 2, 3, 4, 5, and 6 (S3 Table). Chromosome 2 had four clusters, indicating that it is a hot spot for ScLEA genes. In addition to the tandem duplication events, 19 segmental duplication events with 27 ScLEA genes were also identified using BLASTP and MCScanX (Fig 4 and S4 Table). These results indicate that some ScLEA genes are possibly generated by tandem duplication and segmental duplication events.

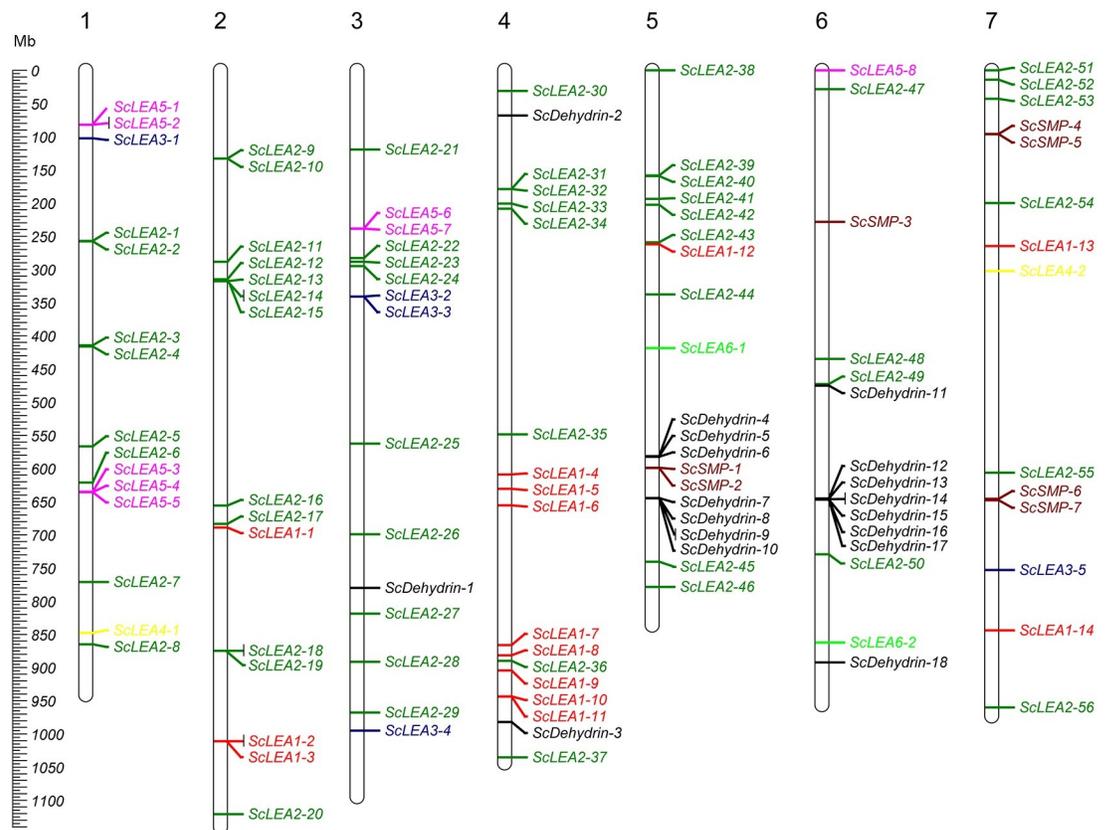


Fig 3. Distribution of ScLEA genes on rye chromosomes visualized using Mapchart software.

<https://doi.org/10.1371/journal.pone.0249757.g003>

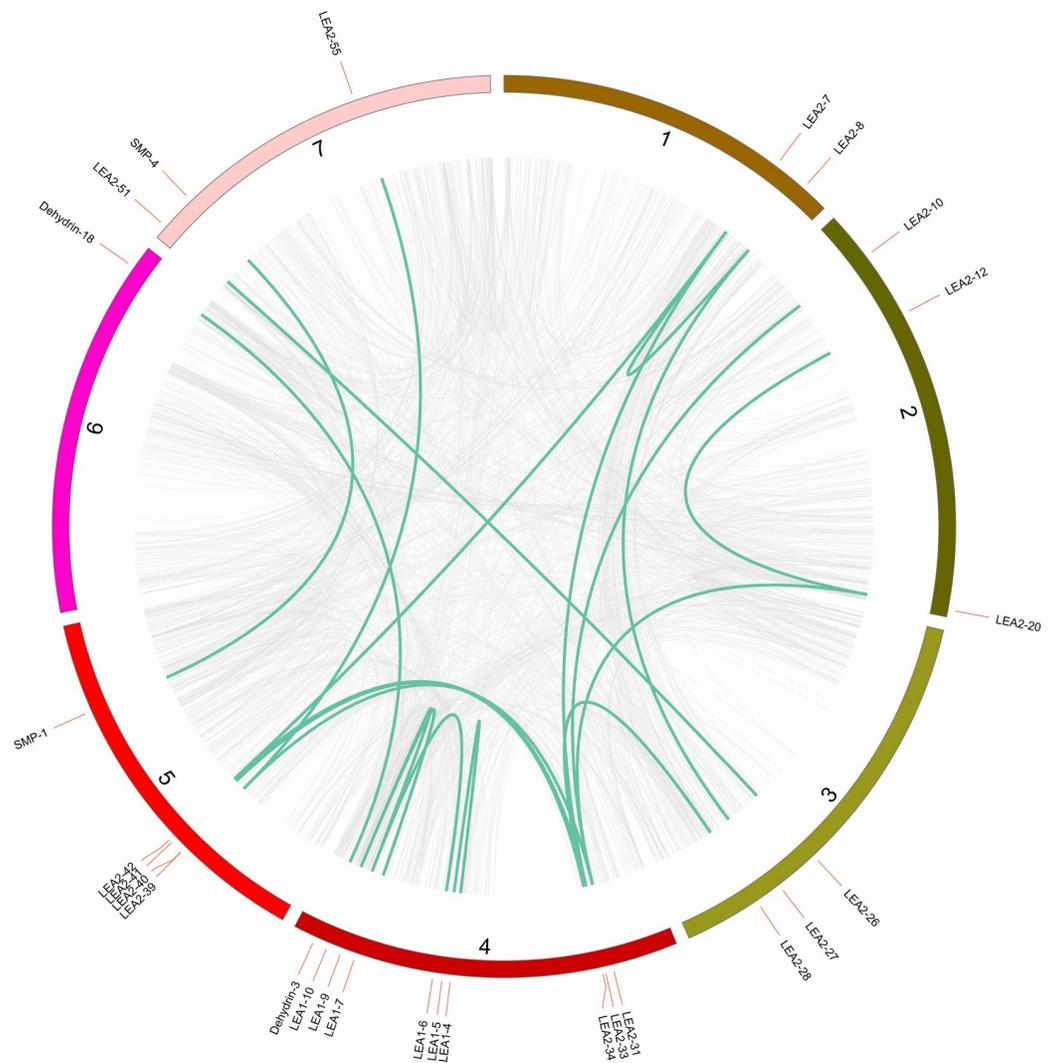


Fig 4. Schematic representation of the chromosomal distribution and interchromosomal relationships of *ScLEA* genes derived from segmental duplication. The gray lines indicate all syntenic gene pairs in the rye genome, and the dark turquoise lines indicate segmentally duplicated *ScLEA* gene pairs. The number of each chromosome is indicated.

<https://doi.org/10.1371/journal.pone.0249757.g004>

The expression profiles of *ScLEA* genes

RNA-seq data for different tissues of rye (Weining) (root, stem, leaf, spikelet, and grain) were used to explore the potential biological functions of *ScLEA* genes in growth and development. Expression profiles showed that most *ScLEA* genes were expressed with obvious spatial specificity (Fig 5). *ScLEA2-26* and *ScLEA2-43* were only expressed in spikelets, while *ScLEA2-3*, *ScLEA2-4*, *ScLEA2-27*, *ScLEA2-28*, and *ScLEA2-29* were only expressed in roots. The expression levels of *ScLEA1-3*, *ScLEA1-11*, *ScLEA2-22*, *ScDehydrin-5*, and *ScDehydrin-13* increased continuously during grain development, and reached a maximum at 40 DAA, indicating that these genes may be involved in grain grouting. *ScDehydrin-11*, *ScLEA2-20*, *ScLEA2-31*, and *ScLEA2-51* were highly expressed in different tissues, indicating that they are important at all stages of rye growth and development.

By analyzing the expression profiles of 112 *ScLEA* genes after drought treatment (S1 Fig), we found that the expression levels of *ScDehydrin-10*, *ScDehydrin-11*, *ScDehydrin-13*, and

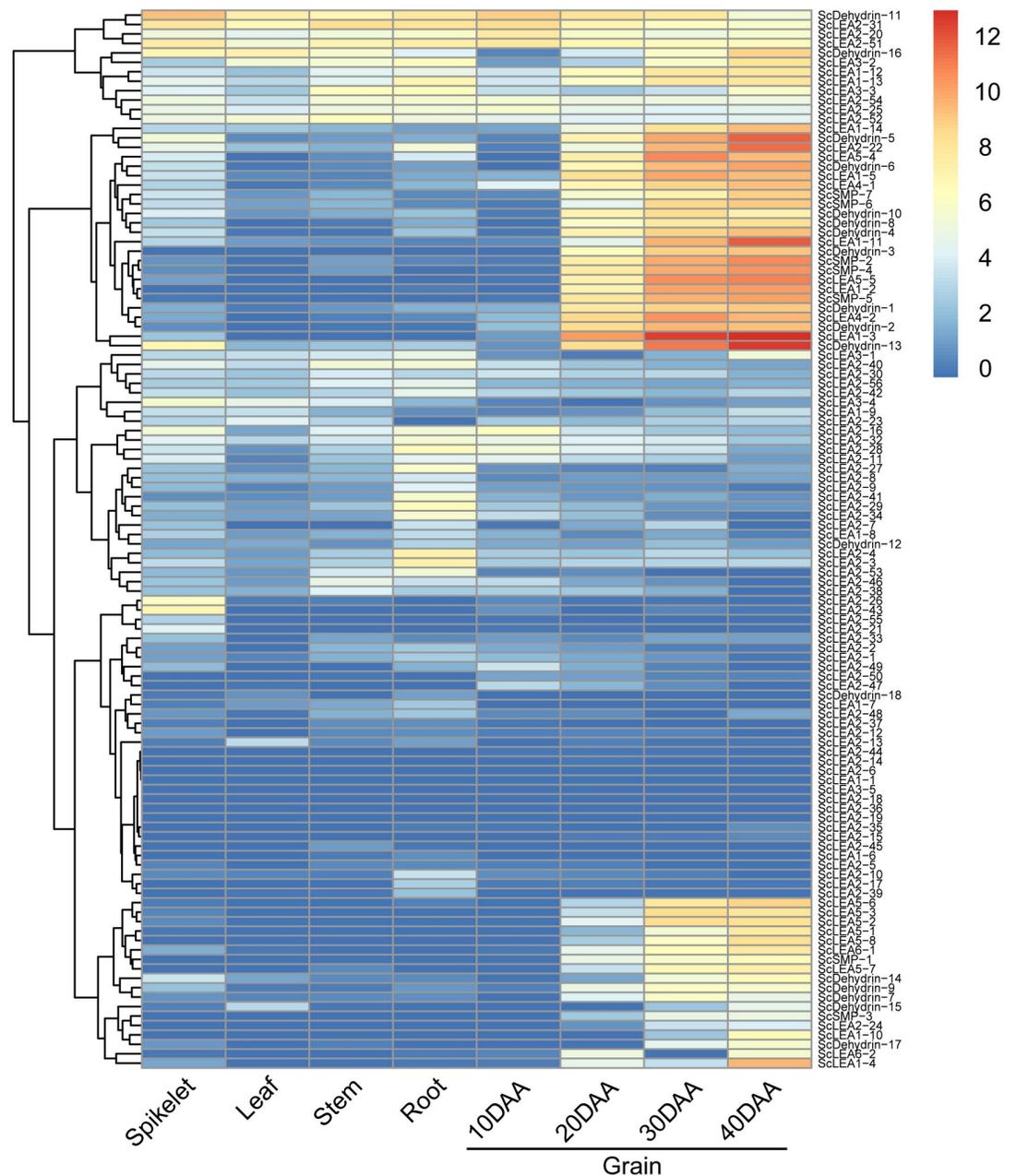


Fig 5. RNA-seq expression profiles of 112 *ScLEA* genes in different tissues of rye (Weining). The heatmap was constructed using TBtools. The color scale on the right represents relative expression levels: Red represents high level and blue represents low level.

<https://doi.org/10.1371/journal.pone.0249757.g005>

ScDehydrin-16 increased the most, followed by *ScLEA1-11*, *ScLEA1-13*, and *ScLEA1-14*, thus, these genes may play major roles in drought stress response. The responses of *ScLEA2* group genes were diverse due to its large number of members. For example, *ScLEA2-22* and *ScLEA2-31* respectively showed a significant upregulation and downregulation to drought stress. The responses of the *ScLEA5* and *ScLEA6* group genes were not obvious, with only the induction of *ScLEA5-5* being prominent.

The responses of *ScLEA* genes to abiotic stress

To verify the potential roles of *ScLEA* genes in abiotic stress, we randomly selected 12 *ScLEA* genes from eight groups, and their expression in rye leaves after treatment with PEG, ABA, NaCl, or mannitol after 0 h, 3 h, 6 h, 9 h, 12 h, and 24 h were analyzed by qRT-PCR. Similar to the results of RNA-seq under drought treatment (S1 Fig), the results of qRT-PCR showed that the expression of these *ScLEA* genes was induced by different stress treatments (Fig 6 and S2 Fig). Among all treatments, *ScLEA* genes showed the most intense response to PEG, and

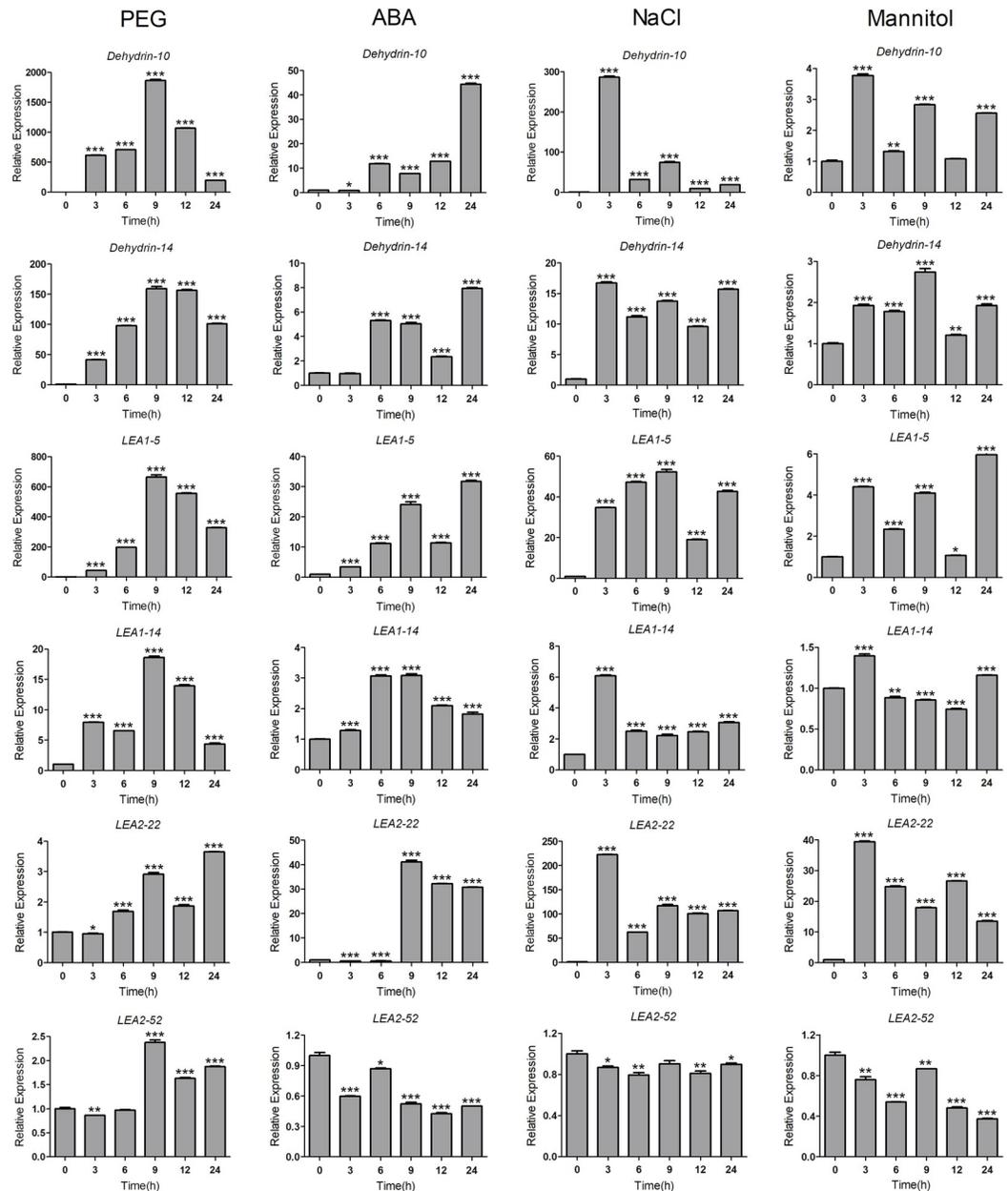


Fig 6. The relative expression levels of six *ScLEA* genes under different abiotic stresses were analyzed by qRT-PCR. Ten-day-old seedling leaves were sampled after 0 h, 3 h, 6 h, 9 h, 12 h, and 24 h under 20% PEG6000, 100 μ M ABA, 200 mM NaCl, or 100 mM mannitol. The values represent mean \pm SEM of three replicates. The significant differences between data were calculated using Student's *t*-test, and indicated with asterisks (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

<https://doi.org/10.1371/journal.pone.0249757.g006>

compared with the control group, the level of *ScLEA* gene upregulation hundreds of times higher (except for *ScLEA2-22* and *ScLEA2-52*) and lasted for a long time (12–24 h). The 12 *ScLEA* genes did not show the same response to ABA treatment, and some genes were even inhibited (*ScLEA2-22* and *ScLEA3-4*). The expression levels of most genes reached a maximum within 3 h after NaCl treatment, and the subsequent changes were quite different between genes. For some genes, expression was maintained at a continuously high level (*ScDehydrin-14*, *ScLEA1-5*, *ScLEA4-2*, and *ScSMP-4*), while the expression levels of others decreased rapidly (*ScDehydrin-10* and *ScSMP-6*). Under mannitol treatment, only the expression of *ScLEA2-22* was significantly upregulated and the expression of some genes was inhibited (*ScLEA2-52* and *ScLEA3-4*).

The results of qRT-PCR analysis showed that genes of the *ScLEA1*, *ScLEA2*, *ScDehydrin*, and *ScSMP* groups were highly sensitive to the four treatments, with significant upregulation of gene expression. Interestingly, the expression of genes in the *ScLEA3* group decreased under three treatments (ABA, NaCl, and mannitol), indicating that these treatments repressed gene expression. The above results showed that the gene expression patterns were different under the four treatments, and that the expression levels of genes in each group were also significantly different under the same treatment. The expression of these genes validates the RNA-seq results and also reflects the diversity of *ScLEA* gene functions in response to abiotic stress.

The response of *ScLEA* genes to cold stress

RNA-seq data analysis showed that *ScLEA* genes were responsive to different durations of low temperature stress treatment (S3 Fig). We analyzed the expression patterns of the same 12 genes under cold stress. *ScLEA* genes expression in rye leaves incubated at 4°C and 0°C for 0 h, 3 h, 6 h, 9 h, 12 h, and 24 h were analyzed by qRT-PCR. The qRT-PCR results showed that most of genes were sensitive to cold stress and showed different degrees of response (Fig 7). The results showed that the expression of *ScLEA* genes increased with increasing duration of cold treatment, and expression was significantly higher than that of the control by 12 h. The responses of *ScLEA* genes to 0°C treatment were significantly stronger than those to 4°C treatment at almost all treatment time points. Among the 12 genes, *ScDehydrin-10*, *ScLEA2-22*, and *ScSMP-6* were strongly responsive to low temperature, and their expression levels were upregulated to hundreds of times compared with the control (0 h). Similar to PEG, ABA, NaCl, and mannitol treatment, the gene expression of *ScLEA2-52* and *ScLEA3-4* to low temperature were inhibited at some time points, suggesting that these genes may be involved in other stress responses, such as biotic stress, or plant growth and development processes. In conclusion, the response of *ScLEA* genes to cold stress reflects the tolerance of rye to low temperatures.

The response of *ScLEA* genes to photon switching

To elucidate the possible regulatory mechanisms of *ScLEA* genes involved in environmental stress, we analyzed the stress-responsive elements in the promoter regions of 12 *ScLEA* genes (S4 Fig). The ABRE, which is involved in ABA responsiveness, the LTRE, which responds to low temperature, the MBS (MYB binding site) element, which is involved in drought response, the G-box, Box 4 and GT1-motif, which participate in light responsiveness. The ABRE and G-box elements were found in most of the 12 genes, and auxin, salicylic acid, and jasmonic acid responsive elements were also identified.

The responses of 12 *ScLEA* genes to PEG, ABA, NaCl, mannitol, and low temperature were confirmed. The role of *LEA* genes in photon switching response has rarely been reported. To verify whether the *ScLEA* genes respond to photon switching, the expression of *ScLEA* genes

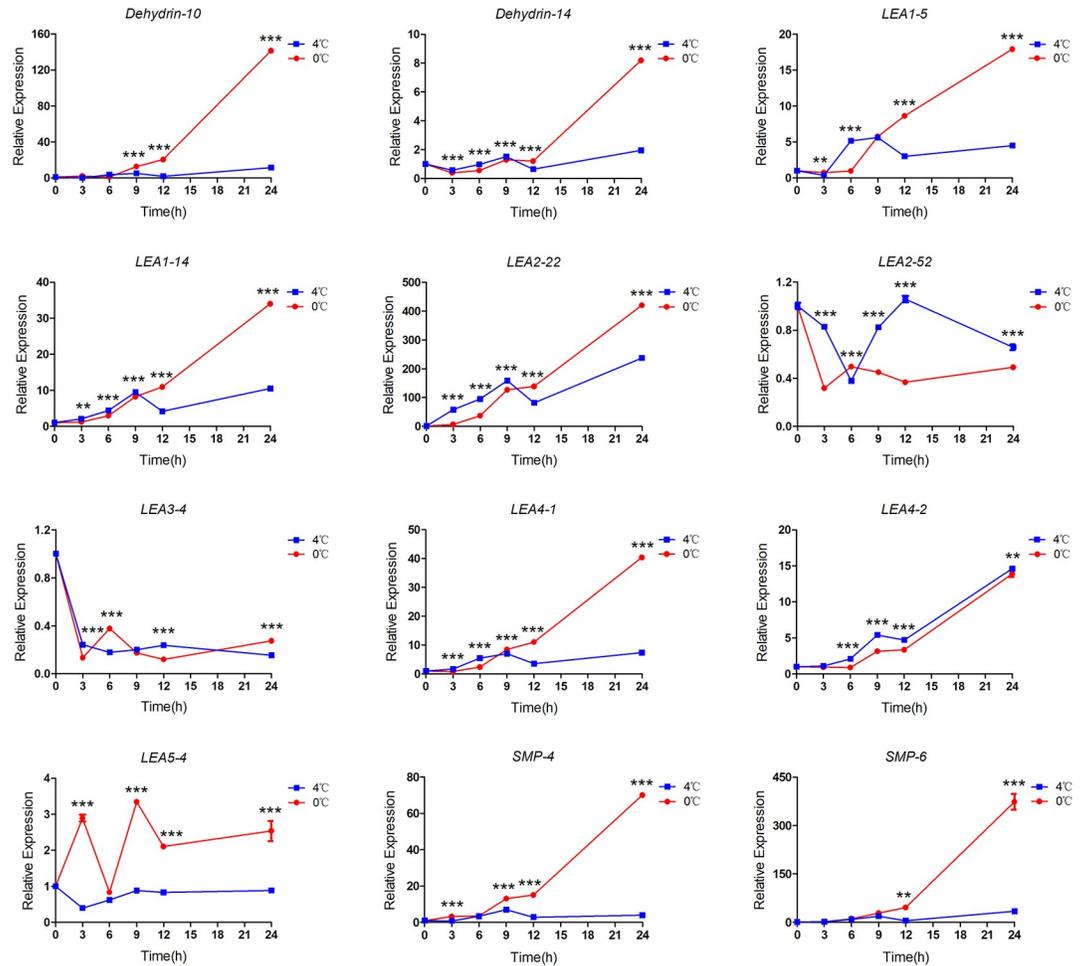


Fig 7. The relative expression levels of 12 *ScLEA* genes under cold stress were analyzed by qRT-PCR. Ten-day-old seedling leaves were sampled after 0 h, 3 h, 6 h, 9 h, 12 h, and 24 h at 0°C or 4°C. The values represent mean \pm SEM of three replicates. The significant differences between data were calculated using two-way ANOVA, and indicated with asterisks (** $P < 0.01$, *** $P < 0.001$).

<https://doi.org/10.1371/journal.pone.0249757.g007>

in rye leaves treated with FR, R, B and W light for 4 h were analyzed by qRT-PCR (Fig 8). Most of these *ScLEA* genes (*ScLEA1-5*, *ScLEA1-14*, *ScLEA2-22*, *ScLEA2-52*, *ScLEA4-1*, *ScLEA4-2*, *ScDehydrin-10*, and *ScDehydrin-14*) showed obvious responses to blue light. Compared with W light, *ScSMP-4*, *ScLEA4-1*, and *ScLEA4-2* showed a stronger response to R light. The results showed that *ScLEA* genes respond to different types of light, and that the degree of response to the four types of light varied greatly. Thus, plants may express various genes including *LEA* in response to stress.

Discussion

Identification and analysis of rye *LEA* genes

LEA genes have been reported in many species; however, the genomic identification and annotation of *LEA* genes has not been reported in rye. In this study, 112 *ScLEA* genes were identified, and were divided into eight groups. LEA2, the largest group in the rye family, accounts for about 50% of the total number of family members. The LEA2 group of wheat is also the largest, accounting for 159 of all 281 *LEA* genes [6]. However, the largest group in rice [5] and

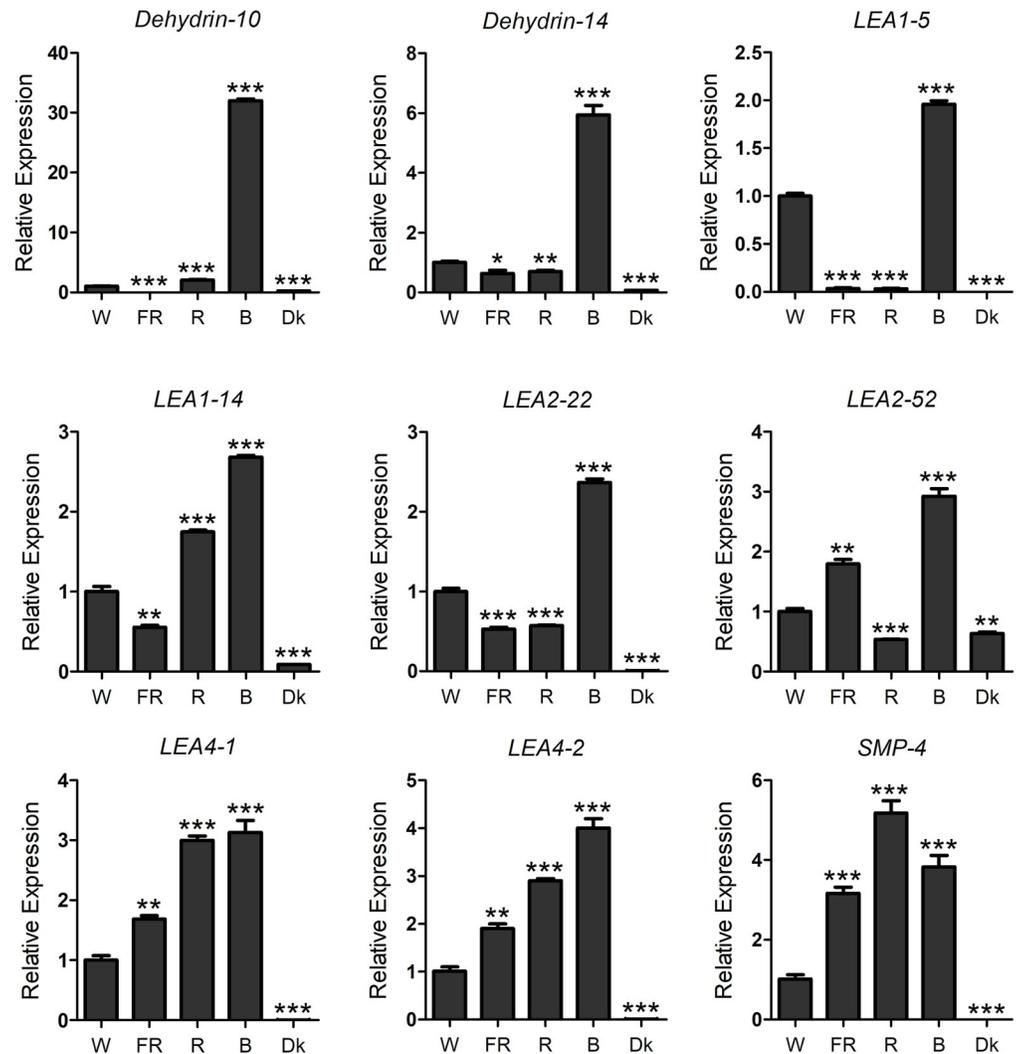


Fig 8. The relative expression levels of nine *ScLEA* genes under different light conditions were analyzed by qRT-PCR. Seedlings were grown in the dark for seven days and subsequently transferred to FR light ($5 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), R light ($17.56 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), B light ($13 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), or W light ($85 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) for 4 h. The values represent mean \pm SEM of three replicates. The significant differences between data were calculated using Student's *t*-test, and indicated with asterisks (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

<https://doi.org/10.1371/journal.pone.0249757.g008>

Brassica napus [41] is Dehydrin and LEA4, respectively. The distributions of *LEA* family members in different species indicate that *LEA* genes play unique roles in stress response and regulation of plant growth and development in different species. These differences also suggest that the *LEA* family genes may have evolved independently after the divergence of these species.

Normally, the diversity of structures and conserved domains caused the evolution of multi-gene families [42]. Our analysis revealed that the majority of *LEA* proteins in the same group shared similar domains, suggesting that these conserved domains may play crucial roles in group-specific functions. However, high divergence was also found in the structures between different groups. Identical results were reported earlier for *LEA* proteins in *Arabidopsis* [43], rice [5], and tomato (*Solanum lycopersicon* L.) [44]. Rye has the largest genome (7.86 Gbp) among all diploid species in Triticeae, with more than 90% repetitive sequences [45]. Gene duplication is considered to be one of the primary driving forces in the evolution of genomes

and genetic systems [46]. Twenty-eight *ScLEA* genes were clustered into 12 tandem duplication blocks, and 27 *ScLEA* genes were clustered into 19 segmental duplication blocks in rye. These results indicate that members of the same group may have originated via gene duplication, and that different groups may have evolved from different ancestral genes with various domain structures [47].

Expression analysis and functional prediction of rye *LEA* genes

Analysis of expression profiles in different tissues revealed obvious temporal and spatial specificity in *ScLEA* genes expression, with the highest expression for most genes mainly observed in grains (Fig 5). The expression levels of *LEA* genes in *Arabidopsis* [4] and wheat [48] were also reported to be generally higher in seeds. Transcriptome data for rye subjected to drought stress showed that, except for *ScDehydrin-3*, all Dehydrin group genes were responsive to drought treatment (S1 Fig). Dehydrin group genes were also specifically and highly expressed in grains (Fig 5). These results showed that the expression levels of Dehydrin group genes are upregulated during seed maturation, suggesting that the expression of these genes might be induced by dehydration at the later stage of seed maturation [48]. The correlation between *LEA* protein and desiccation tolerance had been confirmed in many orthodox seeds [24,49]. During the mature dehydration of seeds, the Dehydrin protein will be present as space filler in cells, maintaining the dissolved character of cell fluid to avoid damage to cell architecture [15]. The molecular function of Dehydrin proteins in the process of seed maturation will provide new ideas for the genetic improvement of crops.

Light affects plant growth and developmental processes including seed germination, dormancy, circadian rhythms, flower induction, plant architecture, and shade avoidance [50]. Studies have found that *LEA* protein expression in *Escherichia coli* promotes tolerance to UV radiation [51,52]. At present, studies of *LEA* proteins in plants generally focus on drought, salt, plant hormone, and low temperature stresses, including UV radiation stress, but few studies have been conducted on the response to different light qualities. Our results showed that *ScLEA* genes responded to FR, R, B, and W light, and the transcription level was almost the highest under blue light. Thus, *ScLEA* genes may play certain roles in blue light signal, and those finding may provide a new perspective for the future study of new functions of *ScLEA* proteins. Rye is mainly grown in northern Europe and North Africa [53], and the mountainous or colder northern parts of China. Most of the regions experience a wide range of temperatures, altitudes, rainfall, and light levels. Therefore, the analysis of different stress treatments in this study lays a foundation for exploring the function of *LEA* proteins in improving the stress tolerance of rye.

Cis-acting elements involved in the mediation of rye *LEA* gene expression

Plants frequently adapt to environmental stresses by regulating the rate of transcription [54]. *Cis*-acting elements of genes are binding sites for transcription factors that activate or repress transcription [55]. Some promoter elements, such as the TATA, GC, and CCAAT boxes, are common to many genes. In response to abiotic stresses such as heat shock and hormones, a large number of specific elements are also involved in transcriptional regulation [55]. In *Arabidopsis*, most genes encoding *LEA* proteins were found to have ABRE and LTRE elements in their promoters and many genes containing these elements were induced by ABA, cold, or drought treatment [4]. In this study, we also found ABRE and LTRE elements in the *ScLEA* promoters (S4 Fig). The results of qRT-PCR showed that the expression of 12 *ScLEA* genes was induced by ABA treatment and low temperature, and that the expression levels were significantly upregulated with a decrease of temperature. Moreover, light responsive elements were

also found in the *ScLEA* promoter, and *ScLEA* genes responded to different light qualities. Therefore, *cis*-acting elements that may influence the expression of *ScLEA* should also be investigated when analyzing the role of *ScLEA* genes under various stresses.

Conclusion

LEAs are important proteins that respond to biotic and abiotic stresses and have been described in numerous plants. In this study, a total of 112 *ScLEA* genes were identified in rye and classified into eight groups. All *ScLEA* genes contained few or no introns, and all encoded proteins containing the conserved LEA domain. *ScLEA* genes were distributed on all rye chromosomes with some clustering. By analyzing promoter elements in combination with RNA-seq data and qRT-PCR results, we showed that *LEA* genes in rye are strongly responsive to abiotic stresses such as drought, low temperature, light quality, ABA, and NaCl. These results lay a foundation for further investigating the functions of LEA proteins and their potential use in genetic improvement of crops.

Supporting information

S1 Fig. RNA-seq expression profiles of 112 *ScLEA* genes after drought treatment for 0 h, 3 h, 6 h, and 12 h. The heatmap was constructed using TBtools. The color scale on the right represents relative expression levels: red represents high level and blue represents low level. (TIF)

S2 Fig. The relative expression levels of six *ScLEA* genes under different abiotic stresses were analyzed by qRT-PCR. Ten-day-old seedling leaves were sampled after 0 h, 3 h, 6 h, 9 h, 12 h, and 24 h under 20% PEG6000, 100 μ M ABA, 200 mM NaCl, or 100 mM mannitol. The values represent mean \pm SEM of three replicates. The significant differences between data were calculated using Student's *t*-test, and indicated with asterisks (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (TIF)

S3 Fig. RNA-seq expression profiles of 112 *ScLEA* genes after incubation at -10°C for 0 h, 1 h, 4 h, and 8 h. The heatmap was constructed using TBtools. The color scale on the right represents relative expression levels: red represents high level and blue represents low level. (TIF)

S4 Fig. Predicted *cis*-acting elements in the promoters of 12 *ScLEA* genes. (TIF)

S1 Table. List of qRT-PCR primers for the 12 *ScLEA* genes in this study. (XLSX)

S2 Table. List of the 112 *ScLEA* genes identified in this study. (XLSX)

S3 Table. Tandem duplicated *ScLEA* genes. (XLSX)

S4 Table. Segmental duplicated *ScLEA* genes. (XLSX)

Author Contributions

Conceptualization: Jianping Yang.

Formal analysis: Mengyue Ding, Weimin Zhan.

Investigation: Guanghua Sun.

Methodology: Xiaolin Jia, Shizhan Chen, Wusi Ding.

Writing – original draft: Mengyue Ding, Lijian Wang.

References

1. Chen Y, Li C, Zhang B, Yi J, Yang Y, Kong C, et al. The role of the late embryogenesis-abundant (LEA) protein family in development and the abiotic stress response: a comprehensive expression analysis of potato (*Solanum Tuberosum*). *Genes (Basel)*. 2019; 10(2):148. <https://doi.org/10.3390/genes10020148> PMID: 30781418.
2. Dure L, Greenway SC, Galau GA. Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochemistry*. 1981; 20(14):4162–4168. <https://doi.org/10.1021/bi00517a033> PMID: 7284317.
3. Close TJ, Lammers PJ. An osmotic stress protein of cyanobacteria is immunologically related to plant dehydrins. *Plant Physiol*. 1993; 101(3):773–779. <https://doi.org/10.1104/pp.101.3.773> PMID: 8310057.
4. Hundertmark M, Hincha DK. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*. 2008; 9:118. <https://doi.org/10.1186/1471-2164-9-118> PMID: 18318901.
5. Wang XS, Zhu HB, Jin GL, Liu HL, Wu WR, Zhu J. Genome-scale identification and analysis of LEA genes in rice (*Oryza sativa* L.). *Plant Sci*. 2007; 172(2):414–420. <https://doi.org/10.1016/j.plantsci.2006.10.004>.
6. Zan T, Li L, Li J, Zhang L, Li X. Genome-wide identification and characterization of late embryogenesis abundant protein-encoding gene family in wheat: evolution and expression profiles during development and stress. *Gene*. 2020; 736:144422. <https://doi.org/10.1016/j.gene.2020.144422> PMID: 32007584.
7. Amara I, Odena A, Oliveira E, Moreno A, Masmoudi K, Pages M, et al. Insights into maize LEA proteins: from proteomics to functional approaches. *Plant Cell Physiol*. 2012; 53(2):312–329. <https://doi.org/10.1093/pcp/pcr183> PMID: 22199372.
8. Tunnacliffe A, Lapinski J, McGee B. A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers. *Hydrobiologia*. 2005; 546(1):315–321. <https://doi.org/10.1007/s10750-005-4239-6>.
9. Hand SC, Menze MA, Toner M, Boswell L, Moore D. LEA proteins during water stress: not just for plants anymore. *Annu Rev Physiol*. 2011; 73:115–134. <https://doi.org/10.1146/annurev-physiol-012110-142203> PMID: 21034219.
10. Hatanaka R, Gusev O, Cornette R, Shimura S, Kikuta S, Okada J, et al. Diversity of the expression profiles of late embryogenesis abundant (LEA) protein encoding genes in the anhydrobiotic midge *Polypedilum vanderplanki*. *Planta*. 2015; 242(2):451–459. <https://doi.org/10.1007/s00425-015-2284-6> PMID: 25822798.
11. Hunault G, Jaspard E. LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics*. 2010; 11:221. <https://doi.org/10.1186/1471-2164-11-221> PMID: 20359361.
12. Gonzalez RM, Iusem ND. Twenty years of research on Asr (ABA-stress-ripening) genes and proteins. *Planta*. 2014; 239(5):941–949. <https://doi.org/10.1007/s00425-014-2039-9> PMID: 24531839
13. Wetzler DE, Fuchs Wightman F, Bucci HA, Rinaldi J, Caramelo JJ, Iusem ND, et al. Conformational plasticity of the intrinsically disordered protein ASR1 modulates its function as a drought stress-responsive gene. *PLoS One*. 2018; 13(8):e0202808. <https://doi.org/10.1371/journal.pone.0202808> PMID: 30138481.
14. Tunnacliffe A, Wise MJ. The continuing conundrum of the LEA proteins. *Naturwissenschaften*. 2007; 94(10):791–812. <https://doi.org/10.1007/s00114-007-0254-y> PMID: 17479232.
15. Battaglia M, Olvera-Carrillo Y, Garcíarrubio A, Campos F, Covarrubias AA. The enigmatic LEA proteins and other hydrophilins. *Plant Physiol*. 2008; 148(1):6–24. <https://doi.org/10.1104/pp.108.120725> PMID: 18772351.
16. Daughdrill GW, Kashtanov S, Stancik A, Hill SE, Helms G, Muschol M, et al. Understanding the structural ensembles of a highly extended disordered protein. *Mol BioSyst*. 2012; 8(1):308–319. <https://doi.org/10.1039/c1mb05243h> PMID: 21979461.
17. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*. 2002; 11(4):739–756. <https://doi.org/10.1110/ps.4210102> PMID: 11910019.
18. Goyal K, Walton LJ, Tunnacliffe A. LEA proteins prevent protein aggregation due to water stress. *Biochem J*. 2005; 388(Pt 1):151–157. <https://doi.org/10.1042/BJ20041931> PMID: 15631617.

19. Shao HB, Liang ZS, Shao MA. LEA proteins in higher plants: structure, function, gene expression and regulation. *Colloids Surf B Biointerfaces*. 2005; 45(3–4):131–135. <https://doi.org/10.1016/j.colsurfb.2005.07.017> PMID: 16199145.
20. Espelund M, Larssen SS, Hughes DW, Galau GA, Larsen F, Jakobsen KS. Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differentially by abscisic acid and osmotic stress. *Plant J*. 1992; 2(2):241–252. <https://doi.org/10.1046/j.1365-313X.1992.t01-46-00999.x> PMID: 1302052.
21. Dure L 3rd. A repeating 11-mer amino acid motif and plant desiccation. *Plant J*. 1993; 3(3):363–369. <https://doi.org/10.1046/j.1365-313x.1993.t01-19-00999.x> PMID: 8220448.
22. Close TJ. Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physiol Plant*. 1996; 97(4):795–803. <https://doi.org/10.1034/j.1399-3054.1996.970422.x>.
23. Candat A, Paszkiewicz G, Neveu M, Gautier R, Logan DC, Avelange-Macherel MH, et al. The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in *Arabidopsis* offers tailored protection against abiotic stress. *Plant Cell*. 2014; 26(7):3148–3166. <https://doi.org/10.1105/tpc.114.127316> PMID: 25005920.
24. Xu D, Duan X, Wang B, Hong B, Ho T, Wu R. Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol*. 1996; 110(1):249–257. <https://doi.org/10.1104/pp.110.1.249> PMID: 12226181.
25. Sivamani E, Bahieldin A, Wraith JM, Al-Niemi T, Dyer WE, Ho THD, et al. Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA1* gene. *Plant Sci*. 2000; 155(1):1–9. [https://doi.org/10.1016/s0168-9452\(99\)00247-2](https://doi.org/10.1016/s0168-9452(99)00247-2) PMID: 10773334.
26. Brini F, Yamamoto A, Jlaiel L, Takeda S, Hobo T, Dinh HQ, et al. Pleiotropic effects of the wheat Dehydrin DHN-5 on stress responses in *Arabidopsis*. *Plant Cell Physiol*. 2011; 52(4):676–688. <https://doi.org/10.1093/pcp/pcr030> PMID: 21421569.
27. Hanin M, Brini F, Ebel C, Toda Y, Takeda S, Masmoudi K. Plant dehydrins and stress tolerance. *Plant Signal Behav*. 2011; 6(10):1503–1509. <https://doi.org/10.4161/psb.6.10.17088> PMID: 21897131.
28. Bauer E, Schmutzer T, Barilar I, Mascher M, Gundlach H, Martis MM, et al. Towards a whole-genome sequence for rye (*Secale cereale* L.). *Plant J*. 2017; 89(5):853–869. <https://doi.org/10.1111/tpj.13436> PMID: 27888547
29. Crespo-Herrera LA, Garkava-Gustavsson L, Åhman I. A systematic review of rye (*Secale cereale* L.) as a source of resistance to pathogens and pests in wheat (*Triticum aestivum* L.). *Hereditas*. 2017; 154:14. <https://doi.org/10.1186/s41065-017-0033-5> PMID: 28559761.
30. Kim W, Johnson JW, Baenziger PS, Lukaszewski AJ, Gaines CS. Agronomic effect of wheat-rye translocation carrying rye chromatin (1R) from different sources. *Crop Sci*. 2004; 44(4):1254–1258. <https://doi.org/10.2135/cropsci2004.1254>.
31. Sharma S, Bhat PR, Ehdaie B, Close TJ, Lukaszewski AJ, Waines JG. Integrated genetic map and genetic analysis of a region associated with root traits on the short arm of rye chromosome 1 in bread wheat. *Theor Appl Genet*. 2009; 119(5):783–793. <https://doi.org/10.1007/s00122-009-1088-0> PMID: 19544051.
32. NDong C, Danyluk J, Wilson KE, Pocock T, Huner NP, Sarhan F. Cold-regulated cereal chloroplast late embryogenesis abundant-like proteins. Molecular characterization and functional analyses. *Plant Physiol*. 2002; 129(3):1368–1381. <https://doi.org/10.1104/pp.001925> PMID: 12114590.
33. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30(4):772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690.
34. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018; 35(6):1547–1549. <https://doi.org/10.1093/molbev/msy096> PMID: 29722887.
35. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered*. 2002; 93(1):77–78. <https://doi.org/10.1093/jhered/93.1.77> PMID: 12011185.
36. Holub EB. The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat Rev Genet*. 2001; 2(7):516–527. <https://doi.org/10.1038/35080508> PMID: 11433358.
37. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012; 40(7):e49. <https://doi.org/10.1093/nar/gkr1293> PMID: 22217600.
38. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014; 30(7):923–930. <https://doi.org/10.1093/bioinformatics/btt656> PMID: 24227677.

39. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools—an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020; 13(8):1194–1202. <https://doi.org/10.1016/j.molp.2020.06.009> PMID: 32585190
40. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc*. 2008; 3(6):1101–1108. <https://doi.org/10.1038/nprot.2008.73> PMID: 18546601.
41. Liang Y, Xiong Z, Zheng J, Xu D, Zhu Z, Xiang J, et al. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*. *Sci Rep*. 2016; 6:24265. <https://doi.org/10.1038/srep24265> PMID: 27072743.
42. Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G. Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. *BMC Plant Biol*. 2010; 10:145. <https://doi.org/10.1186/1471-2229-10-145> PMID: 20630103.
43. Bies-Ethève N, Gaubier-Comella P, Debures A, Lasserre E, Jobet E, Raynal M, et al. Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Mol Biol*. 2008; 67:107–124. <https://doi.org/10.1007/s11103-008-9304-x> PMID: 18265943.
44. Cao J, Li X. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*). *Planta*. 2015; 241(3):757–772. <https://doi.org/10.1007/s00425-014-2215-y> PMID: 25491641.
45. Li G, Wang L, Yang J, He H, Jin H, Li X, et al. A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat Genet*. 2021; Available from: <https://doi.org/10.1038/s41588-021-00808-z> PMID: 33737755.
46. Moore RC, Purugganan MD. The early stages of duplicate gene evolution. *PNAS*. 2004; 100(26):15682–15687. <https://doi.org/10.1073/pnas.2535513100> PMID: 14671323.
47. Ling H, Zeng X, Guo S. Functional insights into the late embryogenesis abundant (LEA) protein family from *Dendrobium officinale* (Orchidaceae) using an *Escherichia coli* system. *Sci Rep*. 2016; 6(1):39693. <https://doi.org/10.1038/srep39693> PMID: 28004781
48. Liu D, Sun J, Zhu D, Lyu G, Zhang C, Liu J, et al. Genome-wide identification and expression profiles of late embryogenesis-abundant (LEA) genes during grain maturation in wheat (*Triticum aestivum* L.). *Genes (Basel)*. 2019; 10(9):696. <https://doi.org/10.3390/genes10090696> PMID: 31510067.
49. Xiao B, Huang Y, Tang N, Xiong L. Over-expression of a LEA gene in rice improves drought resistance under the field conditions. *Theor Appl Genet*. 2007; 115(1):35–46. <https://doi.org/10.1007/s00122-007-0538-9> PMID: 17426956.
50. Franklin KA, Quail PH. Phytochrome functions in *Arabidopsis* development. *J Exp Bot*. 2010; 61(1):11–24. <https://doi.org/10.1093/jxb/erp304> PMID: 19815685.
51. He S, Tan L, Hu Z, Chen G, Wang G, Hu T. Molecular characterization and functional analysis by heterologous expression in *E. coli* under diverse abiotic stresses for *OsLEA5*, the atypical hydrophobic LEA protein from *Oryza sativa* L. *Mol Genet Genomics*. 2012; 287(1):39–54. <https://doi.org/10.1007/s00438-011-0660-x> PMID: 22127413.
52. Huwaidi A, Pathak N, Syahir A, Ikeno S. *Escherichia coli* tolerance of ultraviolet radiation by in vivo expression of a short peptide designed from late embryogenesis abundant protein. *Biochem Biophys Res Commun*. 2018; 503(2):910–914. <https://doi.org/10.1016/j.bbrc.2018.06.095> PMID: 29928878.
53. Hübner M, Wilde P, Schmiedchen B, Dopierala P, Gowda M, Reif JC, et al. Hybrid rye performance under natural drought stress in Europe. *Theor Appl Genet*. 2013; 126(2):475–482. <https://doi.org/10.1007/s00122-012-1994-4> PMID: 23090142.
54. Sachs MM, Ho T-HD. Alteration of gene expression during environmental stress in plants. *Ann Rev Plant Physiol*. 1986; 37(1):363–376. <https://doi.org/10.1146/annurev.pp.37.060186.002051>.
55. Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*. 1989; 245(4916):371–378. <https://doi.org/10.1126/science.2667136> PMID: 2667136.