http://communications.science.ankara.edu.tr

Commun.Fac.Sci.Univ.Ank.Ser. C Biology
Volume 33, Number 2, Pages 81-108 (2024)
ISSN 1303-6025 E-ISSN 2651-3749
DOI: 10.53447/communc.1440742

*Research Article; Received: February 21, 2024; Accepted: March 26, 2024*

# DECIPHERING LATE EMBRYOGENESIS ABUNDANT (*LEA*) GENES IN *PHASEOLUS VULGARIS* L. THROUGH BIOINFORMATICS

DARYA FARHOOMAND AKSOY[1], SİMAY EZGİ AKBULUT[1], İLKER BÜYÜK[1]

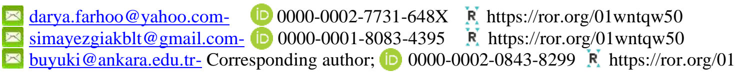
[1]Department of Biology, Faculty of Science, Ankara University, Ankara, TÜRKİYE

**ABSTRACT** The Late Embryogenesis Abundant (*LEA*) gene family is considered vital for plant's ability to survive freezing and desiccation, affecting important developmental and growth processes. These proteins possess notable hydrophilicity and thermal stability, which are essential for preserving cell membrane integrity, forming molecular barriers, aiding in ionic binding, and mitigating oxidative damage during extended periods of exposure to abiotic stress conditions. Although LEA proteins have been extensively studied in numerous plant species, this study represents the initial comprehensive exploration and characterization of LEA proteins in *Phaseolus vulgaris* L. In this context, the biochemical/physicochemical properties of the LEA family at both the gene and protein level have been deeply characterized and defined using various bioinformatics tools. Through comprehensive bioinformatics analyzes, we identified 80 *LEA* genes in common bean and phylogenetically categorized their proteins into eight major groups. Investigating gene duplications, we uncovered 28 events, including 24 segmental and 4 tandem duplications, significantly influencing the evolutionary trajectory of this gene family. In silico micro-RNA (miRNA) target analyzes revealed that 21 *PvLEA* genes were targeted by various miRNAs, with miRN2588 and miR164 being the most prevalent. *PvLEA-63* emerged as the most highly expressed gene across tissues, followed by *PvLEA-27, PvLEA-35, PvLEA-41, PvLEA-49* and *PvLEA-52* genes, demonstrating their ubiquitous expression patterns. Moreover, using publicly available RNAseq data, a comparative expression study of *PvLEA* genes was carried out, and expression alterations in *PvLEA-02, -08, -20, -21, -40, -42, -50* and *-51* genes were detected under both salt and drought stress conditions. These results constitute a substantial resource for future researchers interested in unravelling the functional intricacies of *PvLEA* genes.

**Keywords** *LEA* gene family, *Phaseolus vulgaris* L., bioinformatics, RNAseq

## 1. INTRODUCTION

Common bean is a plant classified under the genus *Phaseolus vulgaris* L. that formed seven thousand years ago in separate regions of the South and North American continents and is known as "new global crop" [1]. The common bean (*Phaseolus vulgaris*, 2n = 22) is an essential component of vegetable protein for millions of people throughout the world and is the most consumed legume in both developing and developed countries [1, 2]. It is sometimes called as "poor man's meat" due to its considerable protein, minerals, soluble fiber, starch, phytochemicals and vitamin content [3]. Furthermore, because of the high



darya.farhoo@yahoo.com- 0000-0002-7731-648X https://ror.org/01wntqw50
simayezgiakblt@gmail.com- 0000-0001-8083-4395 https://ror.org/01wntqw50
buyuki@ankara.edu.tr- Corresponding author; 0000-0002-0843-8299 https://ror.org/01wntqw50

2024 Ankara University
Communications Faculty of Sciences University of Ankara Series C: Biology
The copyright of the works published in our journal belongs to the authors and these works are published as open access under the CC-BY 4.0 license

protein content, people who ate a diet high in common bean had a decreased frequency of cancer, cardiovascular disease (CVD) and diabetes [4]. Several biotic factors, such as bacteria, viruses and fungal infections, as well as abiotic challenges like low temperature, drought, and salt, significantly impact the yield and quality of common bean [5]. The significance of drought and salt resistance in bean cultivation cannot be ignored, as they are essential to produce food for human consumption [6].

In the 21st century, food problems have arisen due to global climate change, drought, population growth and reduction of agricultural lands. Salinity and drought are environmental stress conditions that induce major modifications in controlling how genes are expressed, including the activation or suppression of genes, as well as the modulation of signal transduction pathways, and these changes have serious effects on the life cycle of plants [6, 7, 8]. In common bean production, cultivating plants tolerant to abiotic stress is crucial [6]. However, developing plants resilient to stress requires further investigation into identifying genes that contribute to the way plants react to stress conditions [9]. Furthermore, two fundamental stress-inducing gene sets that are assumed to be crucial for stress response were found in a study on Arabidopsis [10]. The initial category involves regulatory proteins, protein kinases and transcription factors that regulate other signaling molecules. The next category consist of chaperones that regulate osmotin, abiotic stress tolerance, mRNA binding proteins, antifreeze proteins, water channel proteins, and late abundant embryogenesis (LEA) proteins [10].

Activating the *LEA* gene family confers resistance to abiotic stressors, including drought, cold, and salinity, which are critical for plant development and growth [6, 8, 11-13]. LEA proteins show a strong affinity to water and possess a remarkable capacity to resist extreme temperatures. Moreover, they play a vital role in plant cell membrane stabilization, anti-oxidation, establishing molecular barriers and ionic binding under prolonged abiotic stress [14]. During the late phases of embryonic development, LEA proteins stored in seeds were found to respond to diverse abiotic stresses especially drought stress in higher plants [15]. Previous studies also indicated that the upregulation of *LEA* genes in various plant species, including wheat, lettuce, rice, tobacco, and Arabidopsis, enhanced their ability to withstand abiotic stresses [16].

LEA proteins were initially discovered in the latter stages of cotton seed embryogenesis [17]. Subsequent research has revealed the presence of LEA proteins in various plant tissues, including stems, leaves, flowers, and roots, across a diverse array of plant species such as rice, wheat, potato, maize, barley, rye, sunflower, bean, grape, carrot, apple, Arabidopsis, soybean, tomato and canola [18, 19, 20]. These proteins are mostly located in intracellular organelles like mitochondria, cytoplasm, and chloroplasts [13, 21]. Their presence extends to various ecosystem of prokaryotes, invertebrates, and plants ranging from algae to angiosperms [7]. Different types of higher plants including *P. trichocarpa* L., *O. sativa* L., and *A. thaliana* (L.) Heynh have different members of LEA protein

family due to differences in their amino acid sequences, evolutionary relationships and conserved domains.

In current plant biology, it is significant to be able to examine and evaluate massive, complicated datasets [22]. Thanks to technological strides in genome-wide sequencing, the LEA proteins have now been identified in numerous plant species [19, 20, 23, 24]. The genome-wide characterization and identification studies of commercially important plants like *P. vulgaris* are expected to be valuable. This study aims to identify genes that could enhance plant resistance and reduce crop losses caused by environmental and biological factors in plants. It provides a comprehensive genome-wide analysis along with detailed categorization and identification of *P. vulgaris LEA* genes, a novel contribution to the existing literature. Also, the identification of 80 *PvLEA* genes from the *P. vulgaris* genome was validated by analyzing publicly available RNAseq data which shows transcriptomic changes in response to drought and salt stress in common bean. We anticipate that the information we have collected may contribute to the future studies of the *LEA* gene family in *P. vulgaris*.

## 2. MATERIALS AND METHODS

### 2.1 Identification of *PvLEA* gene family members

LEA protein sequences of *P. vulgaris* were obtained from Phytozome v13 database through keyword search with default parameters using following Pfam codes: [(LEA_1 (PF03760), LEA_2 (PF03168), LEA_3 (PF03242), LEA_4 (PF02987), LEA_5 (PF00477), LEA_6 (PF10714), DHN (PF00257), and SMP (PF04927)] (http://pfam.xfam.org/) [25, 26]. To identify the putative proteins, redundant sequences were manually eliminated, and the remaining PvLEA proteins were then subjected to blastp analysis in NCBI database. Physicochemical properties such as amino acid composition, molecular weights (MWs), isoelectric points (pIs) and grand average hydropathicity values (GRAVY) were determined using the ProtParam tool (https://web.expasy.org/protparam/), while protein domains were identified using HMMER (www.ebi.ac.uk/Tools/hmmer/) [27]. Subcellular localization of the PvLEA proteins in *P.vulgaris* was predicted using the WoLF PSORT: Protein Subcellular Localization Prediction Tool [28].

### 2.2 Phylogenetic analysis and classification of PvLEA members

The alignment of *Arabidopsis thaliana* and *Phaseolus vulgaris* LEA proteins was performed using ClustalW [29]. The phylogenetic tree was created utilizing the neighbor-joining (NJ) method with 1000 bootstrap replicates in MAFFT (https://mafft.cbrc.jp/alignment/software/). The resulting tree was visualized using the iTOL online tool (https://itol.embl.de/) [30, 31].

## 2.3 Chromosomal localization, gene structure, and conserved motif analyses

*P. vulgaris* general feature format (GFF3) was downloaded from JGI Data Portal (https://data.jgi.doe.gov/). The chromosomal distribution of *PvLEA* genes was illustrated using the GFF3 file in TBtools software [32]. MEME tool (Multiple Expectation Maximization for Motif Elicitation) (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) was employed to identify conserved motifs in PvLEA proteins using specified parameters including maximum number of motifs '10'; site distribution, any number of repetitions for site distribution, a minimum motif width of '2' and a maximum motif width '50' [33]. TBtools software was utilized to visualize the exon-intron structure of *PvLEA* genes via GFF3 file [32].

## 2.4 Gene duplication and synteny analysis

GFF3 and genome sequence files of *G.max, A.thaliana,* and *P.vulgaris* species were downloaded through the JGI Data Portal (https://data.jgi.doe.gov/). Gene duplication events were analyzed in the TBtools program using the Multiple Collinearity Scan toolkit (MScanX). Gene duplication analyses were conducted using the Multiple Collinearity Scan toolkit (MScanX) in TBtools program. Synteny maps were generated using the Dual Synteny Plotter software based on orthologous genes between *P. vulgaris* and other species (*A.thaliana* and *G.max* L.) [32]. Collinearity analysis validated paralogous relationships which were then visually represented using the Circos tool in TBtools software [32]. Duplicated gene pairs identified through synteny analysis were examined for selective pressure using the ratio of nonsynonymous (Ka) and synonymous (Ks) substitutions (Ka/Ks) calculated with KaKs_Calculator [32]. The duplication period (MYA-million years ago) and divergence of each *PvLEA* gene were calculated using $T = Ks/2\lambda$ ($\lambda = 6.56E^{-9}$) formula [34].

## 2.5 Cis-Regulatory element analyses in promoter regions of *PvLEA* genes and protein structure homology modeling

To examine the cis-regulatory elements within the promoter region of *PvLEA* genes, the 1500 bp genomic sequence before the initiation codon (ATG) of each gene was obtained from the Phytozome database. Potential cis-elements in the promoter sequence were assessed using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [35]. Additionally, all PvLEA proteins were searched against the Protein Data Bank (PDB) using BLASTP (default settings) to recognize templates with similar sequences and three-dimensional structures [36]. Subsequently, Phyre2 (Protein Homology/AnalogY Recognition Engine; http://www.sbg.bio.ic.ac.uk/phyre2) was employed for protein homology modeling in "intensive" mode to forecast protein configurations [37].

### 2.6 miRNA analysis in the *PvLEA* gene family

The PmiREN database (https://www.pmiren.com/) was utilized to download all known *P.vulgaris* miRNA sequences [38]. The psRNATarget Server was then used to determine the miRNAs targeting *PvLEA* genes (https://www.zhaolab.org/psRNATarget/) [39]. The miRNAs and targeted *PvLEA* genes were displayed using the Cytoscape software [40].

### 2.7 Tissue-specific mRNA levels of *PvLEA* genes

The Phytozome v13 Database was utilized to evaluate the expression levels of *PvLEA* genes across various plant tissues during different developmental stages, including root 10, nodules, root 19, young buds, stem 10, stem 19, green mature buds, leaves, young trifoliates, flower buds, and flowers [26]. In-silico expression levels were quantified using FPKM (fragments per kilobase of transcript sequence per million base pairs sequenced) values, which were log2 transformed for analysis. A heatmap illustrating the expression patterns was created utilizing the TBtools program [32].

### 2.8 Identification of *PvLEA* expression levels in response to salt and drought stresses through RNAseq analysis

The expression levels of *PvLEA* genes were quantified using Illumina RNA-seq data obtained from the Sequence Read Archive (SRA) during exposure to salt and drought stressors. The data accession numbers SRR957667 (control leaf for salt experiment), SRR957668 (salt-treated leaf), SRR8284481 (drought-treated leaf), and SRR8284480 (control leaf for drought experiment) were utilized as previously specified by Büyük et al. (2016) [9, 41, 42]. Expression data underwent conversion to Log2 format, followed by the creation of a heatmap using the TBtools software to visualize the results [32].

## 3. RESULTS AND DISCUSSION

### 3.1 An extensive *LEA* gene family in *P. vulgaris*

In this study, 80 *LEA* genes have been identified in *P. vulgaris* genome (Supplementary Table 1). These genes were named from *PvLEA-01* to *PvLEA-80* according to their chromosomal order. The *PvLEA* genes were found to be classified into eight subfamilies [(LEA_1 (PF03760), LEA_2 (PF03168), LEA_3 (PF03242), LEA_4 (PF02987), LEA_5 (PF00477), LEA_6 (PF10714), SMP (PF04927), and DHN (PF00257)] based on sequence homology. The number of identified *PvLEA* genes in *P. vulgaris* genome was found to be higher than those previously identified in *Solanum tuberosum* L. (n = 74) [7], *Zea mays* L. (n = 32) [43], *Oryza sativa* L. (n = 34) [44] and *Cucumis sativus* L. (n = 79) [45]. However, *Triticum aestivum* L. (n = 179) [46], *Nicotiana tabacum* L. (n = 123) [8], and *Arachis hypogaea* L. (n = 126) genomes were found to have more *LEA* genes than *Phaseolus vulgaris* L. (n = 80) [21].

*PvLEA* genes were mostly found to be clustered in LEA_2 subfamily according to the phylogenetic analysis. Similarly, the majority of *LEA* genes from *Camellia sinensis* L. [23], *Sorghum bicolor* (L.) Moench [47], and *Triticum aestivum* [48] were also found to be clustered in LEA_2 subfamily. LEA_6 subfamily was the cluster with lowest number of genes and contained only single *PvLEA* gene in this study. In previous studies on *Camellia sinensis* [23] and *Solanum lycopersicum* [49], LEA_6 subfamily could not even be detected. This absence might be attributed to the LEA_6 subfamily being considered the most recent addition to the LEA family in plants, emerging only in early angiosperms [50].

The lengths and molecular weights of identified PvLEA proteins ranged from 82 to 468 amino acids and from 8.76 to 50.61 kDa, respectively. The pI (the theoretical isoelectric point) values of PvLEA proteins were between 4.7 to 10.3. It was determined that 80% (64 members) of PvLEA proteins showed basic properties. Calculated grand average of hydropathy index (GRAVY) values of all PvLEAs were between −1.558 and 0.411. A significant number, constituting 71% of PvLEAs (57 members), demonstrated GRAVY values below zero suggesting the hydrophilic nature of most PvLEAs. Similarly, previous studies conducted on *Salvia miltiorrhiza* Bunge [14], *Brassica campestris* L. [51] and *Panax notoginseng* L. [52] also showed that LEA proteins were mostly hydrophilic. The aliphatic index, which shows the amount of aliphatic side chains (leucine, isoleucine, valine, and alanine) in a protein, is known to increase its overall thermostability [53]. The aliphatic index values of PvLEA proteins were between 23.3 and 121.3, indicating the high level of protein thermostability. The instability index values were found to range between -3.16 and 68.04 and it was determined that instability index values of 48 members (60%) were below 40, signifying the stability of the majority of PvLEA proteins. Notably, PvLEA-26 stands out as the protein with the lowest instability index value, suggesting the most stable protein structure. In a previous study on tomato, more than %50 of SlLEA proteins were also found to be stable according to the instability index values [54].

The subcellular localization analysis of PvLEA proteins showed that approximately 27.5% and 32.5% of PvLEA proteins were located in cytoplasm and chloroplast, respectively (Supplementary Table 1). In previous studies, LEA proteins were mostly shown to localize in cytoplasm, mitochondria, and chloroplasts [7, 13, 55]. A study by [43] showed that 56.3% of ZmLEA proteins from maize were specifically localized in the nucleus while the remaining part was predominantly distributed in chloroplast, mitochondria, and cytoplasm. The presence of LEA proteins across various cellular compartments and tissues strongly implies their essential role in cellular activities during stress [14].

### 3.2 Phylogeny, conserved motif and gene structure analysis of PvLEA members

To date, the *LEA* gene family has been identified in across diverse plant species [16]. However, the specific number of genes encoding LEA proteins in *P. vulgaris* remains unknown. Thus, we conducted a phylogenetic analysis to characterize the LEA protein family in *P. vulgaris.* Accordingly, 80 LEA proteins were identified which were found to be categorized into eight main clades. Phylogenetic studies revealed that eight clades could be distinguished based on different structural domains of PvLEA proteins [15]. Among these clades, the LEA_2 subfamily was notable for having the most members with 101 (63.5%). Following the LEA_2 subfamily LEA_1, LEA_3, LEA_4, LEA_5, LEA_6, SMP and DHN, subfamilies contained 8 (5%), 13 (8.1%), 7 (4.4%), 4 (2.5%), 4 (2.5%), 10 (6.2%) and 12 (7.5%) proteins, respectively. Similar to the findings obtained in this study, LEA_2 group was also found to have the most LEA members in *Solanum lycopersicum*, *Camellia sinensis* and *Nicotiana tabacum* genomes in previous studies [8, 12, 54]. The group with fewest PvLEA proteins was LEA_6 with only single member (*PvLEA-20*).

In addition to phylogenetic analysis, motif structure analysis was carried out to assess the distribution and characteristics of motifs found in PvLEA proteins (Figure 2).

In general, the distribution of motifs was found to be different in different PvLEA subgroups shown in phylogenetic tree, while proteins in the same subgroups were found to have similar motif distribution pattern [56]. The number of motifs present in each LEA_2 subgroup member was different. For example, the majority of proteins in LEA_2 subgroup contained 5 motifs however PvLEA-13, PvLEA-15, and PvLEA-37 proteins restrict the content to a singular motif. The LEA_4 subgroup was found to have proteins including all 10 conserved motifs (PvLEA-03 and PvLEA-50 proteins). Due to the presence of diverse conserved motifs in various subgroups, it is probable that these subgroups might be originated from independent ancestors that possessed different conserved motifs [49]. Furthermore, this analysis highlighted that the closely related proteins displayed a characteristic motif composition and showed greater structural similarity, while the genes encoding them having significantly different exon and intron lengths (Figure 2 and 3).

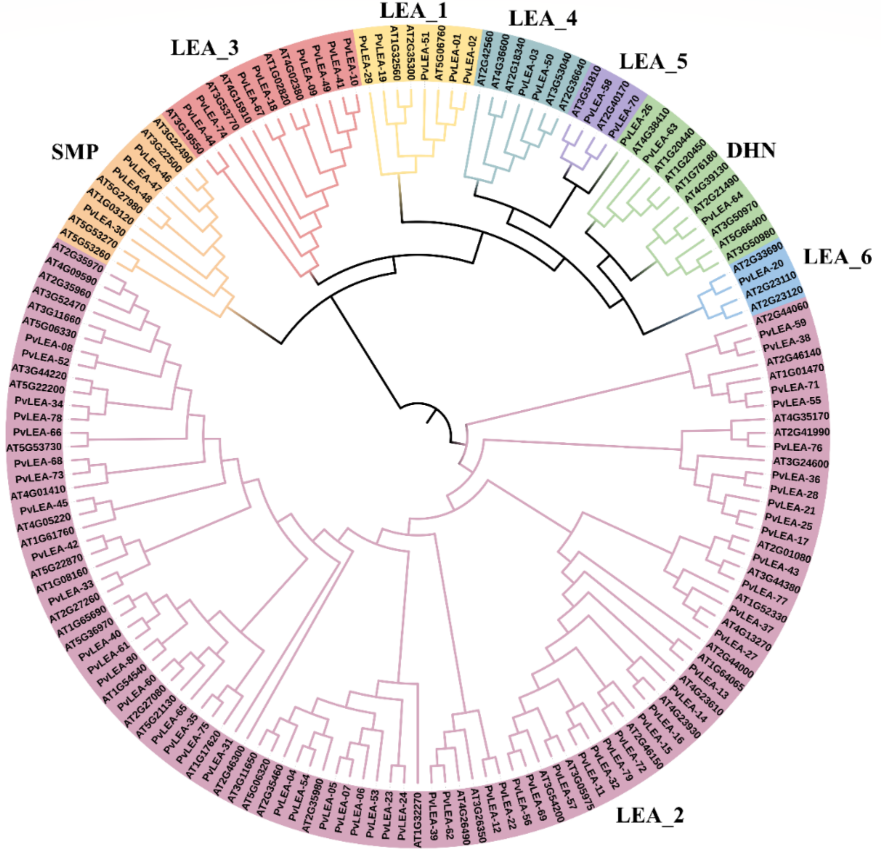



FIGURE 1. Phylogenetic analysis of LEA proteins from *P. vulgaris* and *A. thaliana.*

Further investigation was carried out on the gene structure of the *PvLEA* gene family (Figure 3). According to the findings of gene structure and chromosomal distribution analyses, it is emphasized that tandem duplication events can increase the number of genes in a gene family and duplications are also important for increasing adaptation of plants to different environmental conditions [57, 58]. In this perspective, the organization of exon and intron structures in *PvLEA* genes reflected the diversity in gene structures which could contribute to the functional variations observed in homologous genes. Homologous genes in a group typically have a similar structure, involving the length and number of exons of the genes (Figure 3). Analyses of the 80 *PvLEA* genes demonstrated that 56.25% (45) of these genes lacked introns (intronless), while 43.75% (35) of them had a range of 1 to 3 introns (intron-poor) (Figure 3). In a study on *Nicotiana tabacum*, it was also revealed that among the 123 *NtLEA* genes, two were characterized as intron-rich, while the remaining 121 were identified as intron-poor genes [8]. *PvLEA-48* was found to have the most exons and introns in *PvLEA* gene family bearing 4 exons and 3 introns.

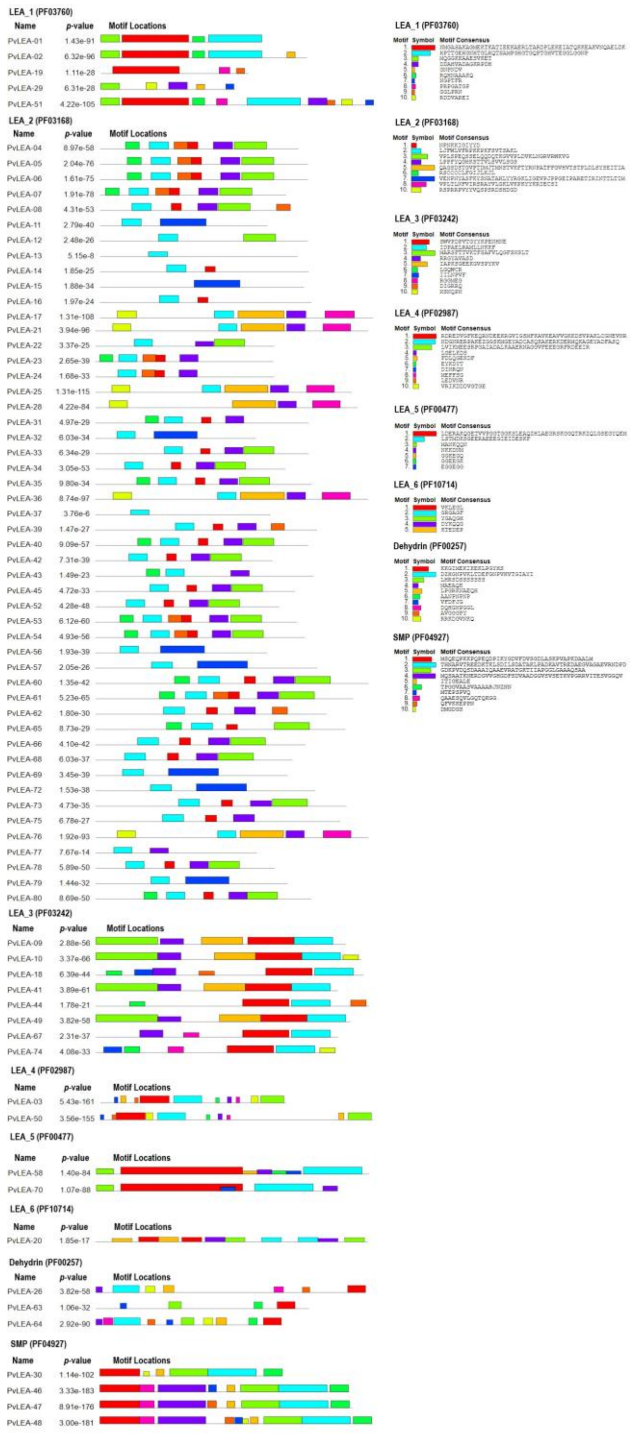



FIGURE 2. Motifs of PvLEA proteins in *P. vulgaris*. Each motif category was shown with a unique color and the numbers (1-10).

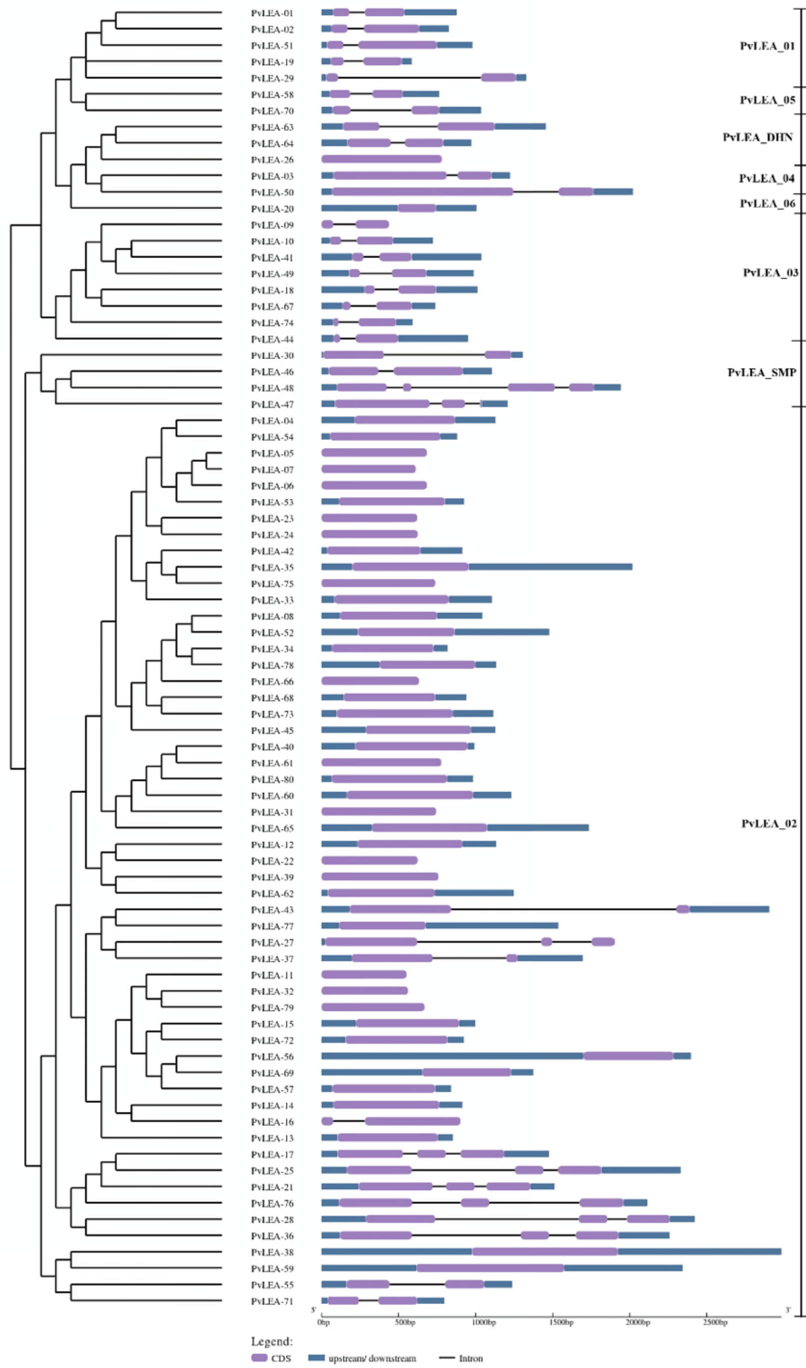



FIGURE 3. The exon and intron arrangement of *PvLEA* genes in *P. vulgaris.* Blue boxes in the diagram denote untranslated regions (UTR), purple boxes signify coding regions (CDS), and lines are used to represent introns.

*PvLEA-17*, *-21*, *-25*, *-27*, *-28*, *-36*, *-47* and *-76* genes are intron and exon (3 exons and 2 introns) rich members after *PvLEA-48*. The *PvLEA-38* gene, which belongs to the LEA_2 subfamily, was found to be longer than all other genes, having a gene structure of around 3000 bp, as well as intronless (0 intron) and including 1 exon. There was also a significant relationship within each phylogenetic group in terms of intron-exon structures (Figures 1 and 3).

### 3.3 Chromosomal localization, gene duplication and syntenic analysis of the *PvLEA* genes

The genomic localization analysis showed the localization of *PvLEA* througout 11 common bean chromosomes. According to Figure 4, the highest number of *PvLEA* genes was found on chromosome 7 (11 genes) and the lowest on chromosome 11 (4 genes). Furthermore, five *PvLEA* genes were found to be located on chromosome 3, with six genes on each of chromosomes 4, 6, and 9. Chromosomes 2 and 10 housed seven *PvLEA* genes each, while chromosome 8 contained eight, chromosome 5 held nine, and chromosome 1 accommodated 10 genes. The *PvLEAs* were distributed unevenly among the 11 chromosomes in *Phaseolus vulgaris*. Moreover, just one gene (*PvLEA-80*) was identified on the unassembled scaffolds (Figure 5). Interestingly, all genes on chromosomes 11 and 2 were from LEA_2 subfamily, and LEA_2 members were detected to be present on all chromosomes. This pattern aligns with the findings obtained from a previous study on *Arachis hypogaea* L., which reported the presence of 126 *AhLEA* genes spreaded out across 20 chromosomes. In their study, LEA_2 members were also found to be present on all chromosomes of *Arachis hypogaea* similar to our findings [21]. Moreover, our analysis identified four sets of tandem duplication events, encompassing eight genes found on chromosomes 1, 2, and 7 (Figure 4).

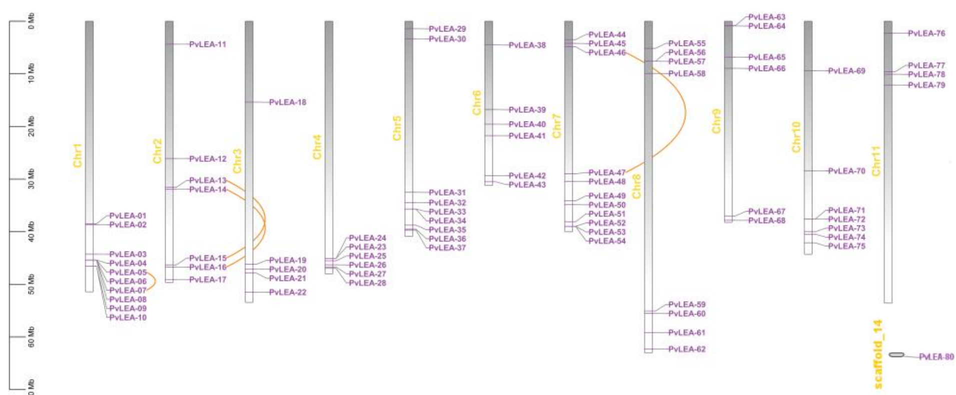



FIGURE 4. The chromosomal positioning of *PvLEA* genes and gene duplication events. The *PvLEA* genes were mapped onto the chromosomes of *P. vulgaris*, and tandemly repeated gene pairs were indicated by orange lines.

Genomic duplications, tandem and segmental duplications, are important events which contribute to the expansion of gene families [59]. In this study, four pairs of tandem duplication and 24 pairs of segmental duplication were detected (Figure 5). In previous studies, different numbers of tandem duplications were also found in *LEA* genes from different plant species including *P. trichocarpa, S. lycopersicum* and *S. pimpinellifolium* [54, 60] [24, 54]. In this study, the origin of tandem duplication events in *PvLEA* genes was found to date back to 0.41 to 25.8 million years ago (MYA) while segmental duplication events were recorded approximately 3.58 to 23.38 MYA. Three of four tandem duplication events were in between *PvLEA* genes from LEA_2 group and the other one was in between *PvLEA* genes from SMP group. The majority (83.3%) of segmentally duplicated genes was from LEA_2 group, followed by LEA_3 (8.3%), LEA_1 (4.1%) and LEA_4 (4.1%) groups. Similarly, in a study on *Arachis hypogaea*, it was discovered that 79.5% of the segmentally duplicated genes belonged to the LEA_2 group [21].

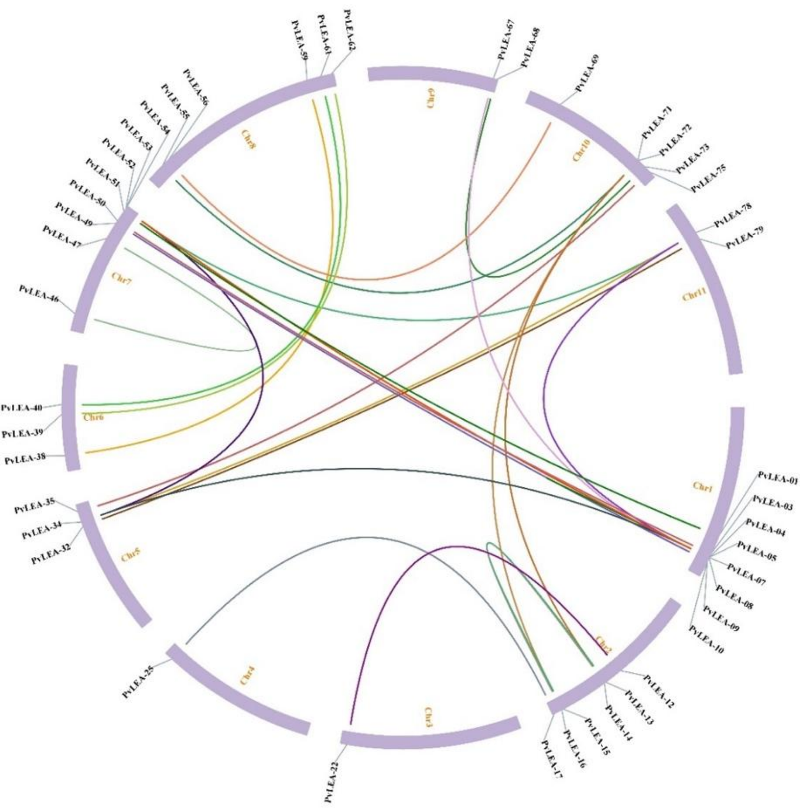



FIGURE 5. Synteny analysis of *PvLEA* genes.

The expansion of the common bean *LEA* gene family might primarily be attributed to the occurrence of segmental duplication events, which outnumber tandem duplication pairs and therefore serve as the main contributing factor [61].

The rate of nonsynonymous substitutions per nonsynonymous site (Ka) and the rate of synonymous substitutions per synonymous site (Ks) commonly utilized to find out how fast protein-coding genes are changing [62]. The Ka/Ks ratio serves as an indicator of the selection pressure and evolutionary rate on a gene. In our analysis for duplicated *PvLEA* genes, the Ka/Ks ratio ranged from 0.649 to 0.114, all falling below 1 (refer to Supplementary Table 2). A Ka/Ks ratio less than 1 signifies purifying selection, where natural selection acts to eliminate deleterious variations, maintaining the functionality of the gene [63, 64]. The constant pattern seen in the ratios indicates that the *LEA* gene family in these species has mostly undergone purifying selection throughout evolution. This observation aligns with findings from the studies on *Cucumis sativus, Populus trichocarpa,* and *Arachis hypogaea L.* species, where genes were similarly found to be affected by purifying selection [21, 45, 65]. It was estimated that tandem duplications occurred 13.81 MYA and segmental duplications occurred 9.10 MYA in this study. These observations indicated that segmental duplications played a more prominent role than tandem duplications for the expansion of *PvLEA* genes. This highlights the essential function of segmental duplication events in stimulating the duplication of *LEA* genes.

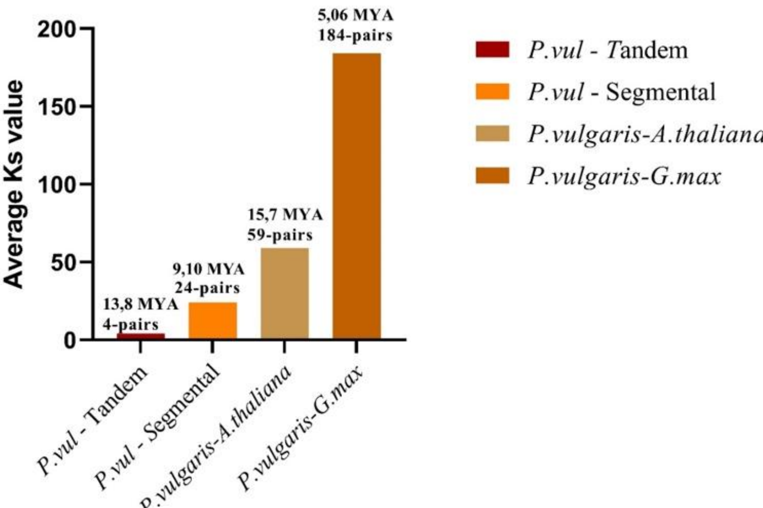



FIGURE 6. The number of ortholog gene pairs (*P. vulgaris* L.-*A. thaliana; P. vulgaris* L.-*G. max*) as well as the tandem and segmental duplication events in *P. vulgaris* L., along with their respective estimated duplication times in million years ago (MYA)

In this study, we explored the orthologous relationships in *LEA* genes among the genomes of *P. vulgaris, A. thaliana*, and *G. max.* The orthologous genes, derived from a shared ancestral gene and exhibiting similar functions in different species, were examined. The genomic comparisons revealed that *P. vulgaris* shares 59 orthologous gene pairs with *A. thaliana* and 184 orthologous gene pairs with *G. max* (Figure 6 and 7, Supplementary Table 2). Notably, an orthologous relationship was identified between *LEA* genes from *G. max* and the *PvLEA-80* gene located on an unassembled scaffold, stemming from the orthologous relationships among *P. vulgaris* and *G. max*. The estimated timeline for these orthologous relationships indicates the occurrence of duplications 5.06 MYA for *P. vulgaris-G. max* and 15.7 MYA for *P. vulgaris-A. thaliana* (Figure 6 and 7, Supplementary Table 2). Our findings revealed that the highest number of *LEA* orthologous gene pairs were observed between *P. vulgaris* and *G. max.* This highest similarity for the *LEA* genes from *P. vulgaris,* and *G. max* is unsurprising given the known genetic resemblance between the two genomes.

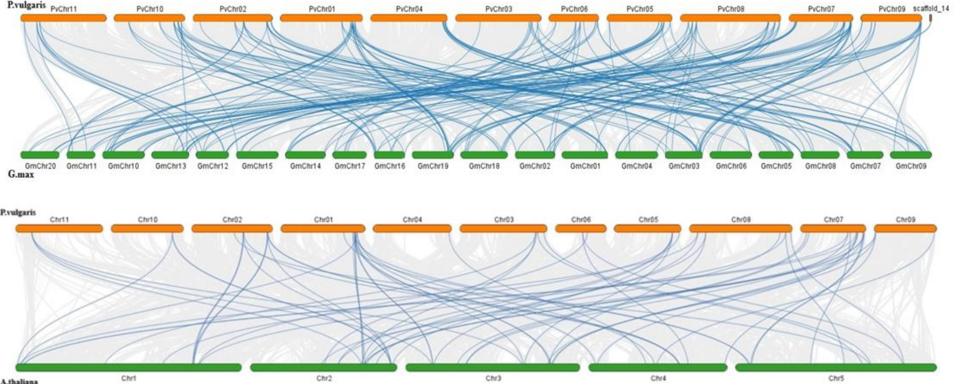



FIGURE 7. The collinearity map illustrates the relationships among *LEA* genes in common bean and two representative plant species, *Glycine max and Arabidopsis thaliana*. In the diagram, blue lines represent collinearity, specifically highlighting synthetic orthologous gene pairs. The background gray lines indicate the collinearity existing among the common bean and the other two species.

### 3.4. Analysis of cis-regulatory elements and homology modeling of *PvLEA* gene family

The presence of a diverse array of cis-regulatory elements (CREs) in gene promoters suggests a range of functions for these genes, encompassing aspects such as plant growth, responses to abiotic stress, and adaptation to environmental factors [66]. Consequently, the cis-acting elements present in *PvLEA* genes were investigated and 110 cis elements were found accordingly (Figure 8, Supplementary Table 3). These predicted CREs were classified into eight classes rely on their functions and these classes were named as CREs associated with (1) responses to stimuli, including environmental stress, (2) light, (3) development,

(4) hormone, (5) promoter, (6) site binding, (7) biotic stress, and (8) CREs with unknown functions (Supplementary Table 3).

The highest number of CRE was in the LEA_2 (66.1%) subfamily, followed by LEA_3 (10.6%), LEA_1 (6.7%), SMP (5.1%), DHN (4.6%), LEA_4 (2.9%), LEA_5 (2.4%) and LEA_6 (1.3%) subfamilies. When CREs were evaluated excluding promoter, site binding, and unknown function groups, the group with the highest CRE was found to be light with 36.01%, followed by hormone (26.54%), environmental stress (19.15%), biotic stress (9.9%) and, development (8.38%) respectively (Figure 8).

Phytohormones are essential for regulating growth and coordinating responses to environmental and biological stresses [67]. In the promoters of *PvLEA* genes, researchers have identified eight CRE groups and some of them were found to be associated with plant hormone responses. These plant hormone related elements were ABREs (26.8%, abscisic acid- related), TGA-element (3.5%, auxin-related), GARE-motif (1.3%, gibberellin-related), P-box (2.1%, gibberellin- related), CGTCA-motif (9.4%, methyl jasmonate-related), TGACG-motif (9.1%, methyl jasmonate- related), TCA element (6.3%, salicylic acid-related) and ERE (28.1%, ethylene-related). Moreover, many elements, including G-Box, ACE, Box-4, Sp1, TCT-Motif and GATA motifs, have previously been associated with drought, salinity, cold and heat-sensitive genes, and were found to have an important role in modifying transcriptional activity of these genes in different plant species [68, 69].

Furthermore, the promoter regions of *PvLEA* genes were found to contain numerous stress- and phytohormone-responsive CREs. These CREs were MYC (37.2%, linked to drought stress), ARE (26.1%, ABA responsive element), WUN-motif (8.5%, sensitive to wound), MBS (7.9%, associated with drought stress) and LTR (2.9%, responsive to low temperature) elements. In a previous study on LEA promoter regions of *A. thaliana,* the presence of ABRE (82%) or LTRE (69%) CREs was already reported similar to our findings in this study [19]. The diversity of CREs in *PvLEA* genes implies potential involvement of these genes in a range of developmental processes in plants. Therefore, our results offer valuable insights into how the *PvLEA* gene family might be regulated, revealing its role in stress tolerance, response to phytohormones, defense mechanisms, and developmental processes.

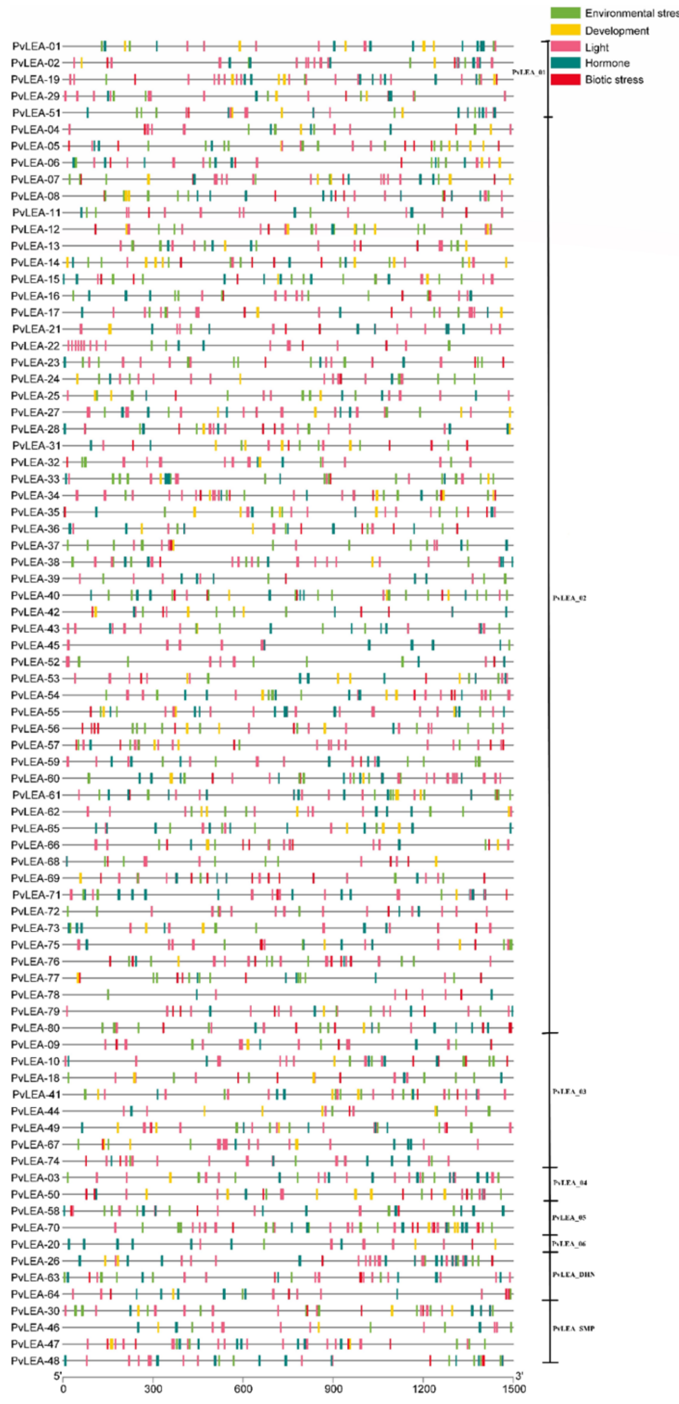



FIGURE 8. The cis-regulatory elements in the promoters of *PvLEA* genes. Different cis-acting elements are represented by colored rectangles.

The 3D structure prediction and homology modeling of 80 PvLEA proteins were conducted using BLASTP against the Protein Data Bank (PDB). Creating 3D models provides valuable insights into protein structure, localization, function, and interaction networks [69]. Differences in 3D topologies were observed in the structure of PvLEA proteins (Figure 9).

Protein models demonstrating high confidence and identity levels surpassing 90% were chosen. A total of four PvLEA proteins [PvLEA-38 (96%), PvLEA-59 (93%), PvLEA-71(99%), and PvLEA-78 (90%)] exhibited high homology with a confidence level higher than 90% in intensive analysis mode (Figure 9). β-sheets were found to be dominant in the secondary structure of PvLEA protein models. Some amino acids are known to disrupt hydrogen bonds during folding due to conflicts among the conformational energy of the side chain and maximal hydrogen bonding [70]. PvLEA-71 (%99) contained the most β sheet structure. Previous studies already showed that LEA_2 group proteins mostly contained β-sheets formation and also some amount of α-helix formation [43, 71]. This structure, like fibronectin Type III domains seen on animal cell surfaces, could function in reducing fluid loss to alleviate the impact of stress or damage on plant tissues [43, 71]. These predicted protein structures are thought to be useful for future molecular biotechnology studies.

FIGURE 9. The three-dimensional structures and binding locations of the PvLEA proteins.

### 3.5. miRNA analysis of *PvLEA* gene family

MiRNAs affect target gene expression in plants under abiotic and biotic stress conditions, thereby contribute to the determination of the role of the targeted genes [70, 72]. Therefore, using data obtained from the PmiREN and psRNATarget databases, the interactions between miRNAs and *PvLEA* genes were depicted in Figure 10. A total of 21 *PvLEA* gene (*PvLEA-07, -09, -13, -14, -23, -24, -31, -35, -42, -43, -46, -49, -50, -53, -58, -62, -63, -64, -66, -69, -79*) were found to be targeted by miRNAs (Figure 10 and Supplementary Table 4). Among the *PvLEA* genes, specifically *PvLEA-14, PvLEA-63*, and *PvLEA-64*, were the primary targets of miRNAs, with miRN2588 and miR164 being the most prevalent. Kavas et al. (2022) investigated the role of R2R3-MYBs, and genes associated with anthocyanin biosynthesis in the development of seed color

in *P. vulgaris*. Interestingly, they discovered that the R2R3-MYB gene was likewise a target of Pvu-miRN2588 [73]. According to the previous findings in the literature, the primary role of miR164 in plants is to regulate cell division, root development, and stress tolerance [74, 75]. Some other studies also demostrated that miR164 expression was strongly affected by mechanical stresses including tension and compression, as well as by drought stress in various plant species including poplar, *Medicago truncatula* L. and *Populus trichocarpa* [76, 77]. Understanding the structure of miRNAs may facilitate the discovery of miRNAs that are particular to certain tissues, hence enhancing our overall comprehension of gene regulatory networks in plants.

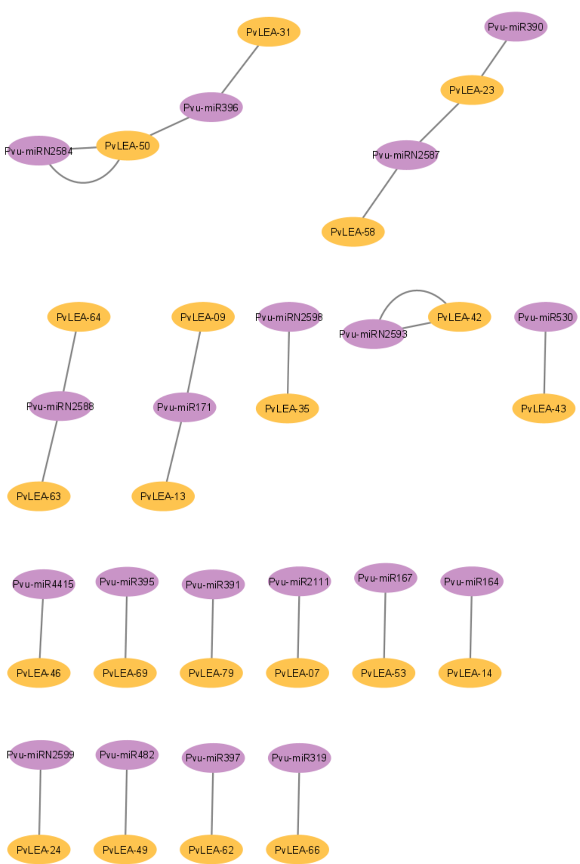



FIGURE 10. The miRNAs targeting the *PvLEA* genes.

### 3.6 Tissue-specific mRNA levels of *PvLEA* genes

Expression profiles provide important insights into the gene activities [78]. To understand the likely functions and expression patterns of *PvLEA* genes throughout various developmental stages, we investigated their tissue-specific expression under normal conditions. Expression data from eleven distinct tissues including flower buds, flowers, green mature pods, leaves, nodules, root 10, root 19, stem 10, stem 19, young pods, and young trifoliates were analyzed. The resulting expression heatmap visually represents the tissue-specific expression patterns of *PvLEA* genes, aiding in our understanding of their roles in various parts of the common bean (Figure 11). Notably, *PvLEA-63* exhibited the highest expression level, followed by *PvLEA-27, PvLEA-35*, *PvLEA-41*, *PvLEA-49*, and *PvLEA-52* genes which are also highly expressed across most tissue types. All *PvLEA* genes, except for *PvLEA-11*, were shown to be expressed in at least one tissue analyzed. Conversely, eight genes (*PvLEA-9, PvLEA-23, PvLEA-24, PvLEA-31, PvLEA-32, PvLEA-46, PvLEA-70,* and *PvLEA-71*) exhibited low expression levels across most tissue types. These findings suggest that *PvLEA* genes exhibit diverse expression patterns across tissues and are implicated in growth and development processes (Figure 11).

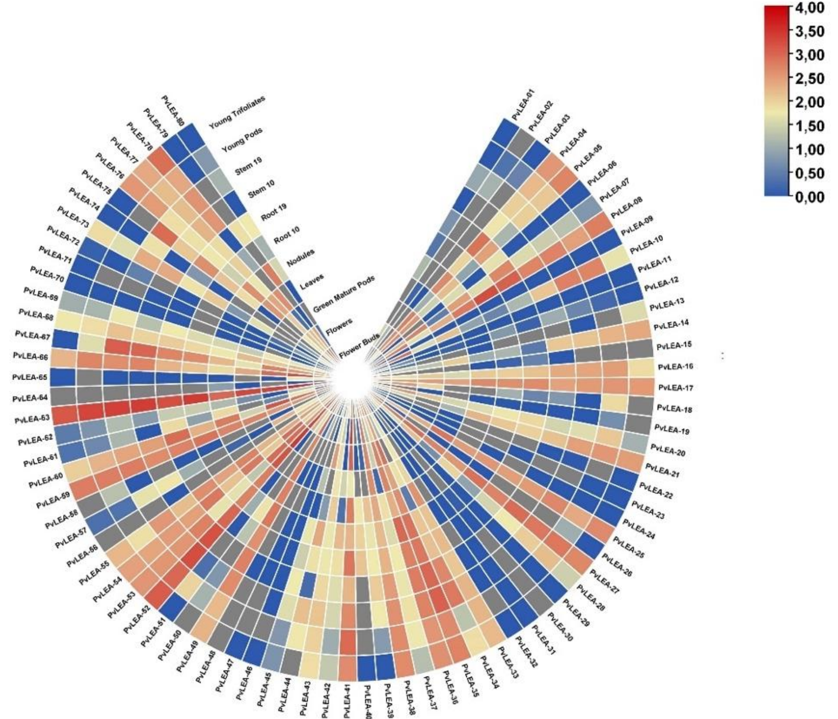



FIGURE 11. Tissue-specific expression level of *PvLEA* genes. Genes exhibiting increased expression levels are shown in red, whereas genes displaying reduced expression levels are represented in blue.

### 3.7. RNAseq analysis of *PvLEA* genes under salt and drought stresses

The expression patterns of *PvLEA* genes under salt and drought stresses were analyzed using RPKM values, revealing differential expression compared to the control (Figure 12). Notably, when comparing the expression values of *PvLEA* genes under salt and drought stresses, a clear contrast emerged: while salt stress led to decreased expression levels of the affected members, drought stress resulted in increased expression levels compared to the control (Figure 12). Notably, *PvLEA-11* and *PvLEA-29* exhibited no change in expression levels under either stress condition compared to the control. Furthermore, a closer examination based on severe expression changes revealed nuanced responses: some genes, such as *PvLEA-08, -21, -40, -42*, and *-75*, showed decreased expression levels under salt stress, whereas others, including *PvLEA-01, -02, -26, -50, -51,* and *-57*, exhibited increased expression levels. Similarly, under drought stress, while certain genes like *PvLEA-02, -20, -50, -51*, and *-79* showed decreased expression, others such as *PvLEA-44* and *-53* displayed increased expression levels. Overall, when evaluating the cumulative expression changes under both salt and drought conditions, the LEA_2 subfamily emerged as particularly responsive, exhibiting the most significant expression changes compared to the control post-stress. In their study, Khodajou-Masouleh et al. (2021) emphasized the presence of late embryogenesis proteins throughout normal plant growth and development. These proteins play a significant role in responding to abiotic challenges, namely drought tolerance [79]. In summary, these findings provide valuable insights for future research endeavors aimed at understanding plant growth, development, and stress tolerance in the context of both abiotic and biotic stresses.

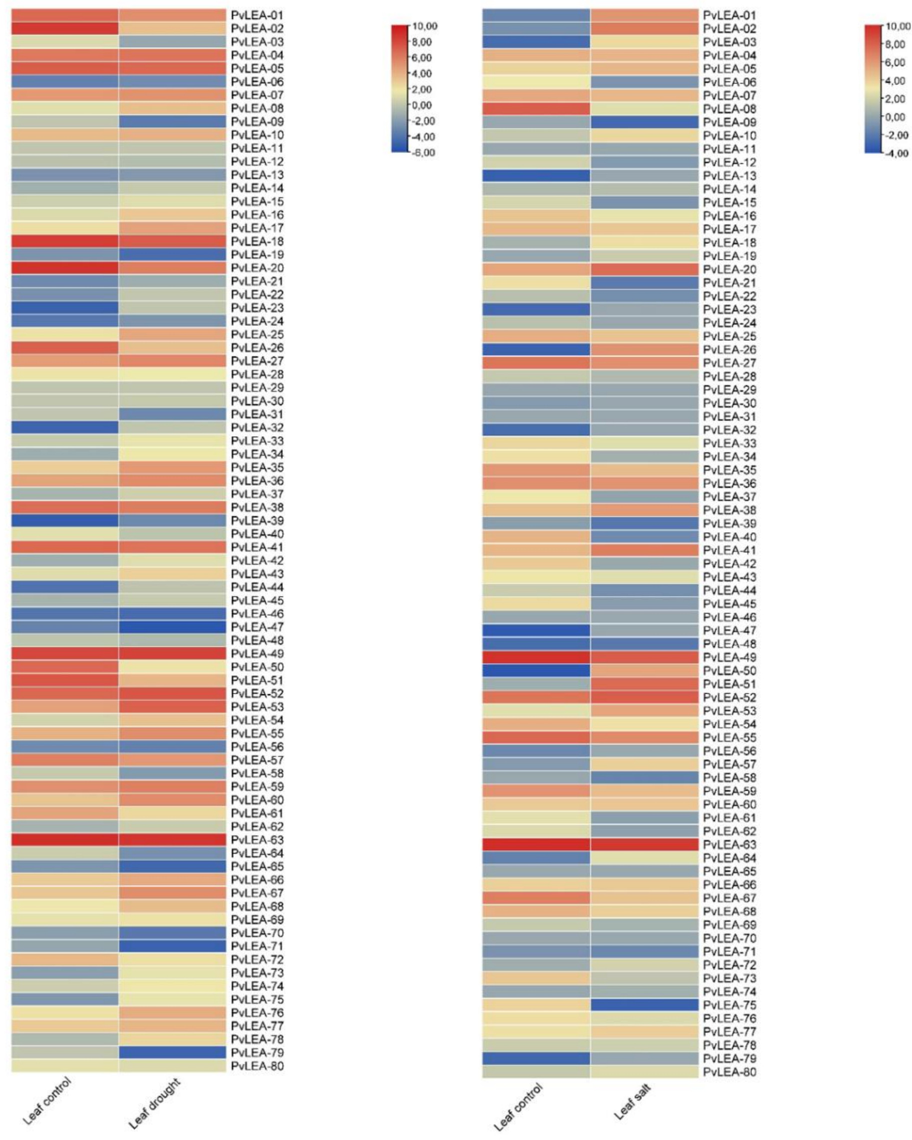



FIGURE 12. A heatmap illustrating the differential expression of *PvLEA* genes in response to drought and salt stress conditions.

## 4. CONCLUSIONS

In this study, we identified 80 *LEA* genes which were classified into eight subfamilies according to the phylogenetic analyzes. The diversed expression patterns across different tissues and the regulation by both cis-regulatory elements and miRNAs highlight the complexity and importance of *LEA* genes in *P. vulgaris*. The functional significance of gene duplication events and the preservation of these genes via purifying selection highlights their importance in evolutionary insights. The LEA_2 subfamily is particularly noteworthy because to its high prevalence and substantial alterations in gene expression in response to salt and drought stress. It might play a crucial role for enhancing resistance to abiotic stress. These findings offer a valuable genetic resource for further functional characterization studies, which could pave the way for genetic improvements in crop resilience to environmental stresses. Overall, this work establishes a robust framework for future investigations into the roles of LEA proteins in *Phaseolus vulgaris* and potentially other legume crops.

**Acknowledgement** SEA was awarded a PhD Scholarship from the Scientific and Technological Research Council of Türkiye (TUBITAK) and Council of Higher Education (YOK).

**Author Contribution Statements** DF- data collection, management and manuscript writing. SEA- project development, manuscript editing. IB- project development, data analysis, manuscript writing and manuscript editing. All authors have read and approved the manuscript.

**Declaration of Competing Interests** The authors declare no conflict of interest.

## REFERENCES

[1] Rodríguez, L., Mendez, D., Montecino, H., Carrasco, B., Arevalo, B., Palomo, I., Fuentes, E., Role of *Phaseolus vulgaris* L. in the prevention of cardiovascular diseases—cardioprotective potential of bioactive compounds, *Plants*, 11 (2) (2022), 186. https://doi.org/10.3390/plants11020186

[2] Nadeem, M. A., Yeken, M. Z., Shahid, M. Q., Habyarimana, E., Yılmaz, H., Alsaleh, A., Hatipoğlu, R., Çilesiz, Y., Khawar, K. M., Ludidi, N., Common bean as a potential crop for future food security: an overview of past, current and future contributions in genomics, transcriptomics, transgenics and proteomics, *Biotechnology & Biotechnological Equipment*, 35 (1) (2021), 759-787. https://doi.org/10.1080/13102818.2021.1920462

[3] Yeken, M. Z., Özer, G., Çiftçi, V., Genome-wide identification and expression analysis of DGK (Diacylglycerol Kinase) genes in common bean, *Journal of Plant Growth Regulation*, 42 (4) (2023), 2558-2569. https://doi.org/10.1007/s00344-022-10726-x

[4] Hutchins, A. M., Winham, D. M., Thompson, S. V., Phaseolus beans: impact on glycaemic response and chronic disease risk in human subjects, *British Journal of Nutrition*, 108 (S1) (2012), S52-S65. https://doi.org/10.1017/S0007114512000761

[5] Büyük, İ., Okay, A., Gorska, M., Ilhan, E., Aras, E. S., Identification and characterization of the Pvul-*GASA* gene family in the *Phaseolus vulgaris* and expression patterns under salt stress, *Turkish Journal of Botany*, 45 (7) (2021), 655-670. https://doi.org/10.3906/bot-2101-13

[6] Büyük, İ., Aras, S., Screening of *PvLEA3* gene mRNA expression levels with qRT-PCR in different bean varieties (*Phaseolus vulgaris* L.) subjected to salt and drought stress, *Turkish Journal of Botany*, 39 (6) (2015), 1014-1020. https://doi.org/10.3906/bot-1502-4

[7] Chen, Y., Li, C., Zhang, B., Yi, J., Yang, Y., Kong, C., Lei, C., Gong, M., 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*, 10 (2) (2019), 148. https://doi.org/10.3390/genes10020148

[8] Geng, W., Wang, Y., Zhang, J., Liu, Z., Chen, X., Qin, L., Yang, L., Tang, H., Genome-wide identification and expression analyses of late embryogenesis abundant (*LEA*) gene family in tobacco (*Nicotiana tabacum* L.) reveal their function in abiotic stress responses, *Gene*, 836 (2022), 146665. https://doi.org/10.1016/j.gene.2022.146665

[9] Hiz, M. C., Canher, B., Niron, H., Turet, M., Transcriptome analysis of salt tolerant common bean (*Phaseolus vulgaris* L.) under saline conditions, *PloS One*, 9 (3) (2014), e92598. https://doi.org/10.1371/journal.pone.0092598

[10] Yamaguchi-Shinozaki, K., Shinozaki, K., Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses, *Annual Review of Plant Biology*, 57 (2006), 781-803. https://doi.org/10.1146/annurev.arplant.57.032905.105444

[11] Gao, J., Lan, T., Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabuliformis* (Pinaceae) in *Escherichia coli*, *Scientific Reports*, 6 (1) (2016), 19467. https://doi.org/10.1038/srep19467

[12] Jin, X., Cao, D., Wang, Z., Ma, L., Tian, K., Liu, Y., Gong, Z., Zhu, X., Jiang, C., Li, Y., Genome-wide identification and expression analyses of the LEA protein gene family in tea plant reveal their involvement in seed development and abiotic stress responses, *Scientific Reports*, 9 (1) (2019), 14123. https://doi.org/10.1038/s41598-019-50645-8

[13] Magwanga, R. O., Lu, P., Kirungu, J. N., Lu, H., Wang, X., Cai, X., Zhou, Z., Zhang, Z., Salih, H., Wang, K., Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton, *BMC Genetics*, 19 (1) (2018), 1-31. https://doi.org/10.1186/s12863-017-0596-1

[14] Chen, J., Li, N., Wang, X., Meng, X., Cui, X., Chen, Z., Ren, H., Ma, J., Liu, H., Late embryogenesis abundant (*LEA*) gene family in *Salvia miltiorrhiza*: identification, expression analysis, and response to drought stress, *Plant Signaling & Behavior*, 16 (5) (2021), 1891769. https://doi.org/10.1080/15592324.2021.1891769

[15] Ma, J., Zuo, D., Ye, H., Yan, Y., Li, M., Zhao, P., Genome-wide identification, characterization, and expression pattern of the late embryogenesis abundant (*LEA*) gene family in *Juglans regia* and its wild relatives *J. mandshurica*, *BMC Plant Biology*, 23 (1) (2023), 80. https://doi.org/10.1186/s12870-023-04096-z.

[16] Leprince, O., Buitink, J., Desiccation tolerance: from genomics to the field, *Plant Science*, 179 (6) (2010), 554-564. https://doi.org/10.1016/j.plantsci.2010.02.011

[17] Dure III, L., Greenway, S. C., Galau, G. A., Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis, *Biochemistry*, 20 (14) (1981), 4162-4168. https://doi.org/10.1021/bi00517a033

[18] Hong-Bo, S., Zong-Suo, L., Ming-An, S., LEA proteins in higher plants: structure, function, gene expression and regulation, *Colloids and Surfaces B: Biointerfaces*, 45 (3-4) (2005), 131-135. https://doi.org/10.1016/j.colsurfb.2005.07.017

[19] Hundertmark, M., Hincha, D. K., LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*, *BMC Genomics*, 9 (2008), 1-22. https://doi.org/10.1186/1471-2164-9-118

[20] Liang, Y., Xiong, Z., Zheng, J., Xu, D., Zhu, Z., Xiang, J., Gan, J., Raboanatahiry, N., Yin, Y., Li, M., Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (*LEA*) gene family formation pattern in *Brassica napus*, *Scientific Reports*, 6 (1) (2016), 24265. https://doi.org/10.1038/srep24265

[21] Huang, R., Xiao, D., Wang, X., Zhan, J., Wang, A., He, L., Genome-wide identification, evolutionary and expression analyses of *LEA* gene family in peanut (*Arachis hypogaea* L.), *BMC Plant Biology*, 22 (1) (2022), 155. https://doi.org/10.1186/s12870-022-03462-7

[22] Akbulut, S. E., Okay, A., Aksoy, T., Aras, E. S., Büyük, İ., The genome-wide characterization of *WOX* gene family in *Phaseolus vulgaris* L. during salt stress, *Physiology and Molecular Biology of Plants*, 28 (6) (2022), 1297-1309. https://doi.org/10.1007/s12298-022-01208-1

[23] Wang, W., Gao, T., Chen, J., Yang, J., Huang, H., Yu, Y., The late embryogenesis abundant gene family in tea plant (*Camellia sinensis*): Genome-wide characterization and expression analysis in response to cold and dehydration stress, *Plant Physiology and Biochemistry*, 135 (2019), 277-286. https://doi.org/10.1016/j.plaphy.2018.12.009

[24] Chen, L., Xin, J., Song, H., Xu, F., Yang, H., Sun, H., Yang, M., Genome-wide study and functional characterization elucidates the potential association of late embryogenesis abundant (*LEA*) genes with lotus seed development, *International Journal of Biological Macromolecules*, 226 (2023), 1-13. https://doi.org/10.1016/j.ijbiomac.2022.11.301.

[25] Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., The Pfam protein families database, *Nucleic Acids Research*, 32 (1) (2004), D138-D141. https://doi.org/10.1093/nar/gkh121

[26] Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Phytozome: a comparative platform for green plant genomics, *Nucleic Acids Research*, 40 (D1) (2012), D1178-D1186. https://doi.org/10.1093/nar/gkr944

[27] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker, J.M. Editor. *The Proteomics Protocols Handbook*. Springer, (2005), 571-607. https://doi.org/10.1385/1-59259-890-0:571

[28] Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C., Nakai, K., WoLF PSORT: protein localization predictor, *Nucleic Acids Research*, 35 (2) (2007), W585-W587. https://doi.org/10.1093/nar/gkm259

[29] Thompson, J. D., Higgins, D. G., Gibson, T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Research*, 22 (22) (1994), 4673-4680. https://doi.org/10.1093/nar/22.22.4673

[30] Katoh, K., Standley, D. M., MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Molecular Biology and Evolution*, 30 (4) (2013), 772-780. https://doi.org/10.1093/molbev/mst010

[31] Letunic, I., Bork, P., Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Research*, 49 (W1) (2021), W293-W296. https://doi.org/10.1093/nar/gkab301

[32] Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., Xia, R., TBtools: an integrative toolkit developed for interactive analyses of big biological data, *Molecular Plant*, 13 (8) (2020), 1194-1202. https://doi.org/10.1016/j.molp.2020.06.009

[33] Bailey, T. L., Johnson, J., Grant, C. E., Noble, W. S., The MEME suite, *Nucleic Acids Research*, 43 (W1) (2015), W39-W49. https://doi.org/10.1093/nar/gkv416

[34] Yang, Z., Nielsen, R, Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models, *Molecular Biology and Evolution*, 17 (1) (2000), 32-43. https://doi.org/10.1093/oxfordjournals.molbev.a026236

[35] Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., Rombauts, S., PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences, *Nucleic Acids Research*, 30 (1) (2002), 325-327. https://doi.org/10.1093/nar/30.1.325

[36] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., The protein data bank, *Nucleic Acids Research*, 28 (1) (2000), 235-242. https://doi.org/10.1093/nar/28.1.235

[37] Kelley, L. A., Sternberg, M. J., Protein structure prediction on the Web: a case study using the Phyre server, *Nature Protocols*, 4 (3) (2009), 363-371. https://doi.org/10.1038/nprot.2009.2

[38] Guo, Z., Kuang, Z., Wang, Y., Zhao, Y., Tao, Y., Cheng, C., Yang, J., Lu, X., Hao, C., Wang, T., PmiREN: a comprehensive encyclopedia of plant miRNAs, *Nucleic Acids Research*, 48 (D1) (2020), D1114-D1121. https://doi.org/10.1093/nar/gkz894

[39] Dai, X., & Zhao, P. X., psRNATarget: a plant small RNA target analysis server, *Nucleic Acids Research*, 39 (suppl_2) (2011), W155-W159. https://doi.org/10.1093/nar/gkr319

[40] Kohl, M., Wiese, S., Warscheid, B., Cytoscape: software for visualization and analysis of biological networks, *Data Mining in Proteomics: From Standards to Applications*, (2011), 291-303. https://doi.org/10.1007/978-1-60761-987-1_18

[41] Büyük, İ., Inal, B., Ilhan, E., Tanriseven, M., Aras, S., Erayman, M., Genome-wide identification of salinity responsive HSP70 s in common bean, *Molecular Biology Reports*, 43 (2016), 1251-1266. https://doi.org/10.1007/s11033-016-4057-0

[42] Jorge, J. G., Villalobos-López, M. A., Chavarría-Alvarado, K. L., Ríos-Meléndez, S., López-Meyer, M., Arroyo-Becerra, A., Genome-wide transcriptional changes triggered by water deficit on a drought-tolerant common bean cultivar, *BMC Plant Biology*, 20 (1) (2020), 1-20. https://doi.org/10.1186/s12870-020-02664-1

[43] Li, X., Cao, J., Late embryogenesis abundant (LEA) gene family in maize: identification, evolution, and expression profiles, *Plant Molecular Biology Reporter*, 34 (2016), 15-28. https://doi.org/10.1007/s11105-015-0901-y

[44] Wang, X.S., Zhu, H.B., Jin, G.L., Liu, H.L., Wu, W.R., Zhu, J., Genome-scale identification and analysis of *LEA* genes in rice (*Oryza sativa* L.), *Plant Science*, 172 (2) (2007), 414-420. https://doi.org/10.1016/j.plantsci.2006.10.004

[45] Celik Altunoglu, Y., Baloglu, P., Yer, E. N., Pekol, S., Baloglu, M. C., Identification and expression analysis of *LEA* gene family members in cucumber genome, *Plant Growth Regulation*, 80 (2016), 225-241. https://doi.org/10.1007/s10725-016-0160-4

[46] Liu, H., Xing, M., Yang, W., Mu, X., Wang, X., Lu, F., Wang, Y., Zhang, L., Genome-wide identification of and functional insights into the late embryogenesis abundant (*LEA*) gene family in bread wheat (*Triticum aestivum*), *Scientific Reports*, 9 (1) (2019), 13375. https://doi.org/10.1038/s41598-019-49759-w

[47] Nagaraju, M., Kumar, S. A., Reddy, P. S., Kumar, A., Rao, D. M., Kavi Kishor, P., 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) (2019), e0209980. https://doi.org/10.1371/journal.pone.0209980

[48] 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*, 736 (2020), 144422. https://doi.org/10.1016/j.gene.2020.144422

[49] Cao, J., Li, X., Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*), *Planta*, 241 (2015), 757-772. https://doi.org/10.1007/s00425-014-2215-y

[50] Artur, M. A. S., Zhao, T., Ligterink, W., Schranz, E., Hilhorst, H. W., Dissecting the genomic diversification of late embryogenesis abundant (LEA) protein gene families in plants, *Genome Biology and Evolution*, 11 (2) (2019), 459-471. https://doi.org/10.1093/gbe/evy248

[51] Jiang, Y., Zhang, S., Xu, H., Tian, H., Zhang, M., Zhu, S., Wang, C., Hou, J., Chen, G., Tang, X., Identification of the *BcLEA* gene family and functional analysis of the *BcLEA73* gene in Wucai (*Brassica campestris* L.), *Genes*, 14 (2) (2023), 415. https://doi.org/10.3390/genes14020415

[52] Jia, J.S., Ge, N., Wang, Q.Y., Zhao, L.T., Chen, C., Chen, J.W., Genome-wide identification and characterization of members of the *LEA* gene family in *Panax notoginseng* and their transcriptional responses to dehydration of recalcitrant seeds, *BMC Genomics*, 24 (1) (2023), 126. https://doi.org/10.1186/s12864-023-09229-0.

[53] Aybüke, O., Sümer, A., Büyük, İ., Detailed characterization of lectin genes in common bean using bioinformatic tools, *Communications Faculty of Sciences University of Ankara Series C Biology*, 31 (1) (2022), 1-25. https://doi.org/10.53447/communc.1056557

[54] Jia, C., Guo, B., Wang, B., Li, X., Yang, T., Li, N., Wang, J., Yu, Q., The *LEA* gene family in tomato and its wild relatives: genome-wide identification, structural characterization, expression profiling, and role of SlLEA6 in drought stress, *BMC Plant Biology*, 22 (1) (2022), 596. https://doi.org/10.1186/s12870-022-03953-7.

[55] Huang, Z., Zhong, X.J., He, J., Jin, S.H., Guo, H.D., Yu, X.F., Zhou, Y.J., Li, X., Ma, M.D., Chen, Q.B., Genome-wide identification, characterization, and stress-responsive expression profiling of genes encoding LEA (late embryogenesis abundant) proteins in Moso bamboo (*Phyllostachys edulis*), *PloS One*, 11 (11) (2016), e0165953. https://doi.org/10.1371/journal.pone.0165953

[56] Olvera-Carrillo, Y., Luis Reyes, J., Covarrubias, A. A., Late embryogenesis abundant proteins: versatile players in the plant adaptation to water limiting environments, *Plant Signaling & Behavior*, 6 (4) (2011), 586-589. https://doi.org/10.4161/psb.6.4.15042

[57] Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S., Melzer, R., Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization, *New Phytologist*, 225 (1) (2020), 511-529. https://doi.org/10.1111/nph.16122

[58] Chen, X., Wang, P., Gu, M., Lin, X., Hou, B., Zheng, Y., Sun, Y., Jin, S., Ye, N., R2R3-MYB transcription factor family in tea plant (*Camellia sinensis*): Genome-wide characterization, phylogeny, chromosome location, structure and expression patterns. *Genomics*, 113 (3) (2021), 1565-1578. https://doi.org/10.1016/j.ygeno.2021.03.033

[59] Mehan, M. R., Freimer, N. B., Ophoff, R. A., A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture, *Human Genomics*, 1 (5) (2004), 1-10. https://doi.org/10.1186/1479-7364-1-5-335

[60] Cheng, Z., Zhang, X., Yao, W., Zhao, K., Liu, L., Fan, G., Zhou, B., Jiang, T., Genome-wide search and structural and functional analyses for late embryogenesis-abundant (*LEA*) gene family in poplar, *BMC Plant Biology*, 21 (1) (2021), 1-17. https://doi.org/10.1186/s12870-021-02872-3

[61] Jiang, S.-Y., González, J. M., Ramachandran, S., Comparative genomic and transcriptomic analysis of tandemly and segmentally duplicated genes in rice, *PloS One*, 8 (5) (2013), e63551. https://doi.org/10.1371/journal.pone.0063551

[62] Hurst, L. D., The Ka/Ks ratio: diagnosing the form of sequence evolution, *Trends in Genetics*, 18 (9) (2002), 486-487. https://doi.org/10.1016/S0168-9525(02)02722-1

[63] Koonin, E. V., Rogozin, I. B., Getting positive about selection, *BioMed Central*, (2003). https://doi.org/10.1186/gb-2003-4-8-331

[64] Lohani, N., Golicz, A. A., Singh, M. B., Bhalla, P. L., Genome-wide analysis of the *Hsf* gene family in *Brassica oleracea* and a comparative analysis of the *Hsf* gene family in *B. oleracea*, *B. rapa* and *B. napus*, *Functional & Integrative Genomics*, 19 (2019), 515-531. https://doi.org/10.1007/s10142-018-0649-1

[65] Lan, T., Gao, J., Zeng, Q.-Y., Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in *Populus trichocarpa*, *Tree Genetics & Genomes*, 9 (2013), 253-264. https://doi.org/10.1007/s11295-012-0551-2

[66] Yamaguchi-Shinozaki, K., Shinozaki, K., Organization of cis-acting regulatory elements in osmotic-and cold-stress-responsive promoters, *Trends in Plant Science*, 10 (2) (2005), 88-94. https://doi.org/10.1016/j.tplants.2004.12.012

[67] Niu, L., Chu, H. D., Tran, C. D., Nguyen, K. H., Pham, H. X., Le, D. T., Li, W., Wang, W., Le, T. D., Tran, L.-S. P., The GATA gene family in chickpea: structure analysis and transcriptional responses to abscisic acid and dehydration treatments revealed potential genes involved in drought adaptation, *Journal of Plant Growth Regulation*, 39 (2020), 1647-1660. https://doi.org/10.1007/s00344-020-10201-5

[68] Shariatipour, N., Heidari, B., Meta-analysis of expression of the stress tolerance associated genes and uncover their-regulatory elements in rice (L.), *The Open Bioinformatics Journal*, 13 (1) (2020). https://doi.org/10.2174/1875036202013010039

[69] Ain-Ali, Q.U., Mushtaq, N., Amir, R., Gul, A., Tahir, M., Munir, F., Genome-wide promoter analysis, homology modeling and protein interaction network of dehydration responsive element binding (DREB) gene family in *Solanum tuberosum*, *PloS One*, 16 (12) (2021), e0261215. https://doi.org/10.1371/journal.pone.0261215

[70] Büyük, İ., Okay, A., Aras, S., Identification and characterization of *SRS* genes in *Phaseolus vulgaris* genome and their responses under salt stress, *Biochemical Genetics*, (2022), 1-22. https://doi.org/10.1007/s10528-021-10108-0

[71] Singh, S., Cornilescu, C. C., Tyler, R. C., Cornilescu, G., Tonelli, M., Lee, M. S., Markley, J. L. Solution structure of a late embryogenesis abundant protein (LEA14) from *Arabidopsis thaliana*, a cellular stress-related protein, *Protein Science*, 14 (10) (2005), 2601-2609. https://doi.org/10.1110/ps.051579205

[72] Akbulut, S. E., Şafak, Z., Okay, A., Amirina, K., Büyük, İ., EIN3/EIL (Ethylene insensitive3/Ethylene insensitive3 like) protein family in *Phaseolus vulgaris*: Identification, evolution and expression analysis within the genome, *Turkish Journal of Agriculture-Food Science and Technology*, 11 (2) (2023), 215-226. https://doi.org/10.24925/turjaf.v11i2.215-226.5172

[73] Kavas, M., Abdulla, M. F., Mostafa, K., Seçgin, Z., Yerlikaya, B. A., Otur, Ç., Gökdemir, G., Kurt Kızıldoğan, A., Al-Khayri, J. M., Jain, S. M., Investigation and expression analysis of R2R3-MYBs and anthocyanin biosynthesis-related genes during seed color development of common bean (*Phaseolus vulgaris*), *Plants*, 11 (23) (2022), 3386. https://doi.org/10.3390/plants11233386

[74] Guo, H.S., Xie, Q., Fei, J.F., & Chua, N.H., MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for Arabidopsis lateral root development, *The Plant Cell*, 17 (5) (2005), 1376-1386. https://doi.org/10.1105/tpc.105.030841

[75] Kim, Y.S., Kim, S.G., Park, J.E., Park, H.Y., Lim, M.H., Chua, N.H., Park, C.M., A membrane-bound NAC transcription factor regulates cell division in Arabidopsis, *The Plant Cell*, 18 (11) (2006), 3132-3144. https://doi.org/10.1105/tpc.106.043018

[76] Lu, S., Sun, Y. H., Chiang, V. L., Stress-responsive microRNAs in Populus, *The Plant Journal*, 55 (1) (2008), 131-151. https://doi.org/10.1111/j.1365-313X.2008.03497.x

[77] Wang, T., Chen, L., Zhao, M., Tian, Q., Zhang, W.H., Identification of drought-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing, *BMC Genomics*, 12 (1) (2011), 1-11. https://doi.org/10.1186/1471-2164-12-367

[78] Chen, Y., Cao, J., Comparative genomic analysis of the *Sm* gene family in rice and maize, *Gene*, 539 (2) (2014), 238-249. https://doi.org/10.1016/j.gene.2014.02.006.

[79] Khodajou-Masouleh, H., Shahangian, S. S., Attar, F., H. Sajedi, R., Rasti, B., Characteristics, dynamics and mechanisms of actions of some major stress-induced biomacromolecules; addressing Artemia as an excellent biological model, *Journal of Biomolecular Structure and Dynamics*, 39 (15) (2021), 5619-5637. https://doi.org/10.1080/07391102.2020.1796793