ÇOMÜ Zir. Fak. Derg. (COMU J. Agric. Fac.)
2024: 12 (2): 294-305
ISSN: 2147–8384 / e-ISSN: 2564–6826
doi: 10.33202/comuagri.1554675

Research Article

Identification and Expression Profiling of LEA Gene Family in Olive (*Olea europaea* L.)

Fatih Sezer[1*]

[1] Çanakkale Onsekiz Mart University, Faculty of Science, Department of Molecular Biology and Genetics, Çanakkale, Türkiye
*Sorumlu yazar: fatihsezer@comu.edu.tr

Received Date: 23.09.2024 Accepted Date: 04.12.2024

Abstract

The olive tree (*Olea europaea* L.) is a vital crop in Mediterranean agriculture, frequently exposed to harsh drought conditions. Among the molecular mechanisms that confer drought tolerance, Late Embryogenesis Abundant (LEA) proteins play a central role. These hydrophilic proteins function in cellular protection during water deficit, preventing protein denaturation, stabilizing membranes, and scavenging reactive oxygen species. In this study, we performed a comprehensive phylogenetic and gene expression analysis of LEA proteins in olive. We identified LEA genes expressed across different tissues and conducted a differential expression analysis to assess their response to drought stress. A phylogenetic tree was constructed to classify LEA family members, and expression data was mapped onto the tree to link evolutionary conservation with functional responses. LEA proteins were classified into distinct subgroups (LEA_1-5, Dehydrin, SMP) to highlight their functional diversity. Additionally, a tissue-specific expression heatmap was generated to illustrate the spatial dynamics of LEA gene activity. Our results provide valuable insights into the molecular mechanisms of drought tolerance in olive and offer potential targets for genetic improvement to enhance resilience in olive cultivation.
Keywords: Olive, Late Embryogenesis Abundant Gene Family, LEA, Gene Expression

Zeytinde (*Olea europaea L.)* LEA Gen Ailesinin Tanımlanması ve Gen Anlatım Profillerinin Belirlenmesi

Öz

Zeytin ağacı (*Olea europaea L.*), Akdeniz tarımında hayati öneme sahip bir üründür ve sıklıkla şiddetli kuraklık koşullarına maruz kalmaktadır. Kuraklık toleransını sağlayan moleküler mekanizmalar arasında, Late Embryogenesis Abundant (LEA) proteinleri merkezi bir rol oynar. Bu hidrofilik proteinler, su kıtlığı sırasında hücre korumasını üstlenerek protein denatürasyonunu önler, zar bütünlüğünü korur ve reaktif oksijen türlerini ortadan kaldırır. Bu çalışmada, zeytinde LEA proteinlerinin kapsamlı bir şekilde filogenetik ve gen anlatımı profilleri incelenmiştir. LEA genlerinin farklı dokulardaki anlatım profilleri belirlenmiş ve kuraklık stresine karşı verdikleri yanıtların belirlenmesi amacıyla gen anlatım analizleri gerçekleştirilmiştir. LEA ailesi üyelerini filogenetik olarak sınıflandırılmış ve gen anlatımı verileri ilişkilendirilen analizler gerçekleştirilmiştir. LEA proteinleri, işlevsel çeşitliliklerini vurgulayacak şekilde (LEA_1–5, Dehydrin, SMP) gibi farklı alt gruplara ayrılmıştır. Ayrıca, LEA gen aktivitesinin farklı dokulardaki dinamiklerini gözler önüne sermek için dokuya özgü bir ısı haritası oluşturulmuştur. Sonuçlarımız, zeytinde kuraklık toleransının moleküler mekanizmalarına dair önemli bilgiler sunmakta ve zeytin yetiştiriciliğinde dayanıklılığı artırmaya yönelik genetik iyileştirme çalışmalarında potansiyel hedefler ortaya koymaktadır.
Anahtar Kelimeler: Zeytin, Late Embryogenesis Abundant Gen Ailesi, LEA, Gen Anlatımı

Introduction

The Late Embryogenesis Abundant (LEA) gene family comprises a diverse group of hydrophilic proteins crucial for enhancing plant stress tolerance, particularly under abiotic conditions such as drought, salinity, and low temperatures. Characterized by their accumulation during the later stages of seed development, LEA proteins protect cellular structures from desiccation and oxidative damage, thereby playing a significant role in plant survival and resilience during critical growth

phases (Hand et al. 2011; J.-S. Jia et al. 2023; Magwanga et al. 2018). With over 300 identified LEA proteins across various plant species, their functional diversity and evolutionary significance underscore their adaptability to environmental stressors (Hanin et al. 2011; Hundertmark and Hincha 2008; Sun et al. 2021).

The LEA proteins are a diverse group of hydrophilic proteins that play crucial roles in plant stress tolerance, particularly during periods of dehydration. Based on their sequence similarity and structural characteristics, LEA proteins can be classified into eight distinct subgroups: LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6, Dehydrins (DHN), and Seed Maturation Proteins (SMP) (Huang et al. 2022; Hundertmark and Hincha 2008; C. Jia et al. 2022; Li et al. 2021; Lin et al. 2024; Zhang et al. 2022). These proteins stabilize cell membranes and biomolecules, facilitate the refolding of misfolded proteins, and scavenge reactive oxygen species generated during stress, making them integral to plant physiological responses. (Huang et al. 2022; J.-S. Jia et al. 2023; Wang et al. 2024). LEA proteins are predominantly localized in the nuclear regions and cytoplasm of plant cells. They have been identified in various tissues, including roots, leaves, buds, and seedlings, although their primary function is observed in seeds (Hundertmark and Hincha 2008).

The expression of LEA genes is tightly regulated by hormonal pathways, particularly abscisic acid (ABA), indicating their complex involvement in the plant's stress response mechanisms (Hundertmark and Hincha 2008; C. Jia et al. 2022; Wang et al. 2024). The importance of LEA proteins extends beyond basic plant biology; they have significant implications for agriculture and crop improvement strategies. Research shows that transgenic plants overexpressing specific LEA proteins demonstrate enhanced abiotic stress tolerance, suggesting their potential for developing resilient crop varieties in the face of climate change (Brini et al. 2007; Hanin et al. 2011; Peng et al. 2008; RoyChoudhury, Roy, and Sengupta 2007). Despite their benefits, the complexity of LEA gene regulation and the variability in expression across different plant species and environmental conditions present ongoing challenges and opportunities for further research in this area.

The olive tree (*Olea europaea* L.) represents one of the most economically and culturally important crops of the Mediterranean basin, known for its resilience in semi-arid climates. However, olive cultivation is frequently challenged by prolonged drought conditions, a major environmental stressor affecting yield and plant health (Gholami et al. 2022; Rico et al. 2023). In such water-limited environments, the olive tree’s ability to endure and continue its physiological functions is underpinned by an array of stress-responsive molecular mechanisms (Ben Abdallah et al. 2018). One such mechanism, LEA proteins have been recognized for their role in response to various abiotic stresses, particularly dehydration.

In summary, the LEA gene family is pivotal for plant stress tolerance, with a notable role in agricultural applications aimed at enhancing crop resilience. Continued investigations into their functional mechanisms and regulatory pathways will be essential for leveraging their potential in combating the challenges posed by climate variability and ensuring global food security. This study presents a comprehensive analysis of LEA proteins in olive with a focus on their evolutionary relationships and role in stress tolerance, particularly under drought conditions. Using RNA-seq data, we identified LEA genes expressed across different tissues of the olive tree, followed by a differential expression analysis to assess their transcriptional response to drought stress. We constructed a phylogenetic tree to elucidate the evolutionary connections among LEA family members. These findings contribute to a deeper understanding of the molecular basis of drought tolerance in olive and provide a foundation for future genetic improvement strategies aimed at enhancing resilience in olive cultivation.

Materials and Methods

Identification of LEA Gene Family Members in Olive Genome

To identify LEA protein family members in olive, we utilized the genome assembly *Olea europaea* var. sylvestris (wild olive) GCF_002742605.1 (Unver et al. 2017). The protein sequences of the entire genome were analyzed using HMMER version 3.1b1. We employed Hidden Markov Model (HMM) profiles corresponding to LEA domains sourced from the Pfam database: PF03760 (LEA_2), PF03168 (LEA_1), PF03242 (LEA_3), PF02987 (Dehydrin), PF00477 (SMP), PF00257 (LEA_4), and PF04927 (LEA_5). This approach yielded a total of 111 putative LEA proteins. Following the domain-based identification, we retrieved the corresponding transcript sequences and genomic loci for each

protein. These sequences were used in further analyses, including phylogenetic studies and gene expression profiling.

Protein Characterization

The identified LEA proteins from the *Olea europaea* var. sylvestris reference genome (GCF_002742605.1) were subjected to comprehensive characterization, focusing on their structural, functional, and physicochemical properties. A set of bioinformatics tools was employed to gather key attributes for each protein, including domain composition, chromosomal localization, exon count, gene length, mRNA length, protein sequence length, subcellular localization, signal peptide presence, molecular weight, and isoelectric point (pI). The HMMER software version 3.1b1 was used to identify conserved domains within the protein sequences, utilizing PFAM profiles corresponding to the LEA gene family domains (PF03760, PF03168, PF03242, PF02987, PF00477, PF00257, PF04927). This process enabled precise annotation of domain architecture for each protein. Chromosomal positions and exon counts for the identified LEA genes were extracted from the reference genome annotations. Genomic lengths and mRNA lengths were retrieved from the genome annotation data, while protein lengths were calculated based on the amino acid sequences. Subcellular localization predictions were made using DeepLoc 1.0, which classified the proteins into various cellular compartments (Ødum et al. 2024). Molecular weight and isoelectric points (pI) were predicted using the pepstats tool from the EMBOSS suite, based on the amino acid compositions of the LEA proteins (Madeira et al. 2024).

Gene Expression Analysis

The RNA-seq datasets used in this study were sourced from the BioProject databases under accession numbers PRJNA590386 (Ramírez-Tejero et al., 2020) and PRJNA606032 (Tsamir-Rimon et al., 2021). Raw sequence data were downloaded from the Sequence Read Archive (SRA) and preprocessed for quality control. Low-quality bases and adapter sequences were trimmed and the processed reads were then mapped to the *Olea europaea* var. sylvestris reference genome (GCF_002742605.1) using the tools within Geneious 2024 (Kearse et al., 2012). Read count data were generated for each gene using the same software.

Subsequent differential gene expression analysis was performed using the DESeq2 package in R. Genes with a log2 fold change greater than 2 or less than -2, and adjusted p-values below 0.05, were considered significantly differentially expressed. The expression data were compared across multiple tissue types and stress conditions. To visualize the gene expression profiles, pheatmap was used to generate heatmaps of differentially expressed genes.

Phylogenetic Analysis

For the phylogenetic analysis, the protein sequences of the identified LEA proteins were aligned using ClustalW in Geneious 2024 (Kearse et al., 2012). An unweighted pair group method with arithmetic mean (UPGMA) tree was constructed based on the multiple sequence alignment, with 1,000 bootstrap replicates to ensure the robustness of the tree. The tree file was exported in newick format and used in further visualization. To incorporate gene expression and domain data into the phylogenetic tree, further visualization was conducted in R using the ggtree and ggtreeExtra packages. The expression data were integrated into the tree using bar plots to display the log2 fold change between control and drought conditions, while domain architecture was represented as colored tiles to highlight the conserved protein domains.

Results and Discussion

Olive LEA Proteins

In this study, we performed a comprehensive identification of LEA proteins in olive. The analysis of 111 LEA proteins was performed, focusing on their domain classifications, chromosomal locations, exon count, sequence length, subcellular localization, and physicochemical properties such as molecular weight and isoelectric point (pI) (Table 1).

A total of 111 LEA proteins were identified in the olive genome, each characterized based on domain composition, chromosomal distribution, exon counts, sequence lengths, subcellular localizations, and physicochemical properties. Among these proteins, 36 contained LEA2 domains, making it the most common domain, followed by 17 proteins with LEA_3 domains, 14 with LEA_1 domains, 5 with Dehydrin (DHN) domains, and 4 with Seed Maturation Protein (SMP) domains. The chromosomal locations of the LEA proteins were found to be widely distributed across various chromosomes, with no significant clustering that might suggest specific regions of high LEA gene

density. This wide distribution hints at the possibility of their involvement in diverse stress response mechanisms throughout the plant.

Exon counts ranged from 1 to 14, indicating structural diversity in gene organization, which could contribute to functional versatility. Genomic lengths varied from 500 bp to over 10,000 bp, while mRNA lengths spanned from 300 bp to 9,500 bp, indicating differences in transcript complexity. Protein lengths similarly showed considerable variability, ranging from 90 to over 600 amino acids, reflecting the diverse roles these proteins may play under different physiological conditions.

Localization predictions revealed that 48 of the LEA proteins have transmembrane domains, likely indicating their involvement in membrane stability under stress conditions. Beyond transmembrane proteins, 20 proteins were predicted to localize to the mitochondria, suggesting a role in protecting mitochondrial integrity during stress. The remaining proteins were mostly soluble, localizing primarily to the cytosol (24 proteins) or nucleus (10 proteins), indicating their potential regulatory roles in stress response and gene expression. A small number of proteins were associated with the endoplasmic reticulum (3 proteins), possibly indicating involvement in protein folding or secretion under stress.

Physicochemical analysis showed that the molecular weight (MW) of the LEA proteins ranged from 10 kDa to 60 kDa, with most proteins falling in the range of 15 kDa to 40 kDa, suggesting they are relatively small proteins, ideal for quick response to abiotic stress. The isoelectric points (pI) varied from 4.5 to 9.5, reflecting the proteins' wide range of charge properties. This variation in pI suggests that these proteins might function in different cellular environments and under different stress conditions, maintaining stability in both acidic and basic intracellular compartments.

Table 1. Protein characteristics of Olive LEA proteins.

Protein Accession	Domain	Chr	Exon Count	Genomic Length	mRNA Length	Protein Length	Localization	Signal	Molecular Weight (kDa)	pI
XP_022842058.1	LEA_1	17	2	698	610	151	Cyt		33,75	6,35
XP_022842435.1	LEA_2	17	1	993	993	209	C_memb	SP|TD	28,46	5,95
XP_022843012.1	LEA_3	18	2	804	558	84	MD	MTP	15,62	6,58
XP_022843751.1	LEA_6	18	1	1005	1005	85	Cyt|Nucleus		23,59	5,53
XP_022844117.1	LEA_2	18	1	1133	1133	252	C_memb	TD	20,39	6,71
XP_022844300.1	Dehydrin	18	2	1141	1046	239	Cyt|Nucleus	NLS	28,52	6,22
XP_022844537.1	LEA_2	18	1	1229	1229	209	C_memb	SP|TD	20,42	5,71
XP_022844932.1	LEA_2	18	3	4334	1118	228	C_memb|ER|LYSO/VAC	SP|TD	22,02	6,11
XP_022845171.1	LEA_2	18	3	1949	1423	323	C_memb|LYSO/VAC	SP|TD	26,53	6,05
XP_022845238.1	LEA_2	19	2	1234	606	201	C_memb|ER	SP|TD	28,54	5,81
XP_022845239.1	LEA_2	19	2	2461	366	121	C_memb|ER	SP|TD	26,83	5,88
XP_022846305.1	LEA_2	1	3	4051	993	222	ER|LYSO/VAC	SP|TD	26,09	6,45
XP_022850613.1	LEA_2		1	792	792	201	C_memb	SP|TD	25,67	5,92
XP_022853417.1	LEA_3		2	788	540	81	MD	MTP	18,42	6,63
XP_022853418.1	LEA_3		2	788	561	80	MD	MTP	16,72	6,58
XP_022853760.1	LEA_2		1	987	987	209	C_memb	SP|TD	24,5	6,72
XP_022854074.1	LEA_2		1	949	949	199	C_memb	SP|TD	19,86	5,88
XP_022854327.1	LEA_3		2	756	633	103	MD	MTP	14,02	6,12
XP_022854420.1	LEA_2		1	990	990	209	C_memb	SP|TD	24,89	6,64
XP_022854693.1	LEA_3		2	846	744	97	MD	MTP	26,92	6,08
XP_022855095.1	LEA_2		2	2482	1267	203	C_memb	SP|TD	27,01	6,35
XP_022858503.1	Dehydrin		3	496	345	114	Cyt|Nucleus	NLS	14,43	5,98
XP_022858973.1	SMP	3	3	1447	944	239	Cyt		18,73	6,85
XP_022859033.1	Dehydrin		2	1758	843	153	Cyt|Nucleus	NLS	24,53	6,91
XP_022859405.1	LEA_3		2	718	577	92	MD	MTP	20,65	5,92

Table 1 cont.

XP_022859406.1	LEA_3		2	718	577	92	MD	MTP	18,64	5,99
XP_022860591.1	LEA_2		1	1003	1003	223	C_memb	SP|TD	26,71	6,02
XP_022860835.1	LEA_2		1	1475	1475	210	C_memb|LYSO/VAC	SP|TD	22,18	6,54
XP_022861872.1	Dehydrin		1	687	687	148	Cyt	NLS	19,75	6,31
XP_022862923.1	LEA_2		1	976	976	208	C_memb	SP|TD	22,39	6,67
XP_022862928.1	LEA_2		1	420	420	139	Nucleus|C_memb	NLS	25,74	5,97
XP_022862930.1	LEA_2		1	612	612	203	C_memb	SP|TD	20,11	6,51
XP_022863156.1	LEA_4		3	1901	1711	400	Cyt|C_memb	NLS	21,5	6,38
XP_022863204.1	LEA_2		3	1813	1545	309	C_memb|LYSO/VAC	SP|TD	27,22	6,11
XP_022864302.1	LEA_1		2	786	703	152	Cyt|Nucleus		18,87	5,86
XP_022865060.1	LEA_2		1	1078	1078	259	C_memb	TD	22,74	6,12
XP_022865774.1	LEA_2	3	1	897	897	222	C_memb|LYSO/VAC	SP|TD	17,96	5,94
XP_022865782.1	LEA_2	3	1	896	896	222	C_memb|LYSO/VAC	SP|TD	17,57	6,01
XP_022866689.1	LEA_2		1	867	867	209	C_memb|LYSO/VAC	SP|TD	27,88	6,24
XP_022866926.1	LEA_2		1	1047	1047	250	C_memb	TD	25,97	5,98
XP_022867195.1	LEA_2		1	1087	1087	218	C_memb|LYSO/VAC	TD	21,43	6,1
XP_022868236.1	LEA_2		1	992	992	211	C_memb	SP|TD	19,65	5,88
XP_022868427.1	LEA_2		1	741	741	222	C_memb|LYSO/VAC	SP|TD	22,31	6,43
XP_022869332.1	LEA_2		1	836	836	208	C_memb	SP|TD	19,92	6,26
XP_022869337.1	LEA_2		4	4949	1096	250	C_memb	TD	24,22	6,14
XP_022869339.1	LEA_2		4	4949	1076	250	C_memb	TD	19,92	6,26
XP_022869340.1	LEA_2		4	4949	1261	250	C_memb	TD	24,35	6,42
XP_022869341.1	LEA_2		4	4949	947	250	C_memb	TD	23,56	6,22
XP_022869344.1	LEA_2		2	2692	1079	270	C_memb	TD	20,15	6,09
XP_022869391.1	SMP	1	3	1326	1164	263	Cyt|Nucleus		24,89	6,15
XP_022871447.1	LEA_2		2	1848	807	268	C_memb	TD	18,74	6,05
XP_022871461.1	LEA_2		1	933	933	260	C_memb	SP|TD	18,74	6,05
XP_022872520.1	LEA_2	3	3	2430	1123	240	C_memb	TD	22,71	6,27
XP_022873986.1	LEA_2		3	1260	1025	306	C_memb|LYSO/VAC	SP|TD	19,32	6,1
XP_022875541.1	LEA_2	1	1	897	897	209	C_memb	SP|TD	23,56	6,09
XP_022875755.1	LEA_3	1	2	921	706	99	MD	MTP	26,71	6,23
XP_022876757.1	Dehydrin	5	2	958	753	175	Cyt|Nucleus	NLS	26,53	6,11
XP_022876768.1	LEA_1	5	2	1000	696	136	Cyt		21,96	6,4
XP_022878701.1	LEA_2	6	2	2437	934	223	C_memb	SP|TD	21,82	6,23
XP_022878798.1	LEA_2	1	4	4119	1066	234	C_memb	SP|TD	22,34	6,42
XP_022878806.1	LEA_2	1	4	4119	1169	234	C_memb	SP|TD	19,74	6,3
XP_022879111.1	LEA_2	1	4	4831	1317	289	C_memb|ER|LYSO/VAC	SP|TD	17,29	5,97
XP_022879120.1	LEA_2	1	4	4831	1152	220	C_memb|ER|LYSO/VAC	SP|TD	16,91	5,94
XP_022881290.1	LEA_2	1	2	1333	1044	210	C_memb|LYSO/VAC	SP|TD	21,94	6,07
XP_022882247.1	LEA_2	8	1	953	953	222	C_memb|LYSO/VAC	SP|TD	19,97	6,21
XP_022882268.1	LEA_5	8	2	609	536	110	Cyt|Nucleus	NLS|NES	22,81	6,36
XP_022883025.1	LEA_2	9	1	1027	1027	217	C_memb	SP|TD	23,27	6,19
XP_022883611.1	LEA_2	10	1	783	783	260	C_memb	TD	18,62	6,04
XP_022884625.1	LEA_2	10	1	1062	1062	252	C_memb	TD	19,14	6,12
XP_022885703.1	Dehydrin	10	2	1005	887	180	Cyt|Nucleus	NLS	19,73	6,22
XP_022885704.1	Dehydrin	10	2	943	828	148	Cyt|Nucleus	NLS	19,73	6,22
XP_022885755.1	SMP	1	3	1190	1035	250	Cyt|Nucleus	NLS	23,09	6,31
XP_022886511.1	LEA_2	10	1	1923	1923	223	C_memb	SP|TD	23,56	6,25

Table 1 cont.

XP_022887510.1	LEA_2	11	3	2854	1398	324	Cyt	NLS	21,83	6,2
XP_022887511.1	LEA_2	11	3	2854	1392	322	Cyt	NLS	21,83	6,2
XP_022887512.1	LEA_2	11	3	2854	1389	321	Cyt	NLS	21,83	6,2
XP_022887839.1	LEA_2	11	1	1012	1012	208	C_memb	SP|TD	21,42	6,18
XP_022887840.1	LEA_2	11	1	1040	1040	260	C_memb	TD	21,42	6,18
XP_022888469.1	LEA_2	11	1	1060	1060	237	C_memb	TD	24,76	6,31
XP_022888912.1	LEA_2	11	2	899	834	277	C_memb	SP|TD	19,57	6,17
XP_022889217.1	LEA_2	11	2	3083	1148	274	C_memb	SP|TD	22,53	6,27
XP_022889488.1	LEA_1	11	2	742	567	124	Cyt|C_memb	NLS	19,94	6,18
XP_022890720.1	LEA_2	12	1	1137	1137	268	C_memb	TD	21,83	6,23
XP_022890958.1	LEA_5	12	2	630	519	90	Cyt|Nucleus	NES	24,98	6,36
XP_022890959.1	LEA_5	12	2	614	512	90	Cyt|Nucleus	NES	24,98	6,36
XP_022891618.1	LEA_2	12	3	2915	1343	167	C_memb|ER|LYSO/VAC	SP|TD	23,42	6,23
XP_022891747.1	LEA_2	12	1	633	633	210	C_memb	SP|TD	23,31	6,27
XP_022891939.1	LEA_2	12	1	648	648	215	C_memb	SP|TD	21,89	6,18
XP_022892036.1	LEA_5	12	3	2076	507	94	Cyt|Nucleus		24,13	6,3
XP_022892037.1	LEA_5	12	3	2076	493	86	Nucleus	NLS	24,13	6,3
XP_022892074.1	SMP	12	3	1136	963	270	Cyt		22,63	6,22
XP_022892120.1	LEA_5	12	2	659	547	90	Cyt|Nucleus	NES	24,67	6,35
XP_022894413.1	Dehydrin	14	2	1195	1099	242	Cyt|Nucleus	NLS	23,16	6,24
XP_022895516.1	LEA_2	15	2	1532	1304	167	Cyt		22,41	6,29
XP_022895517.1	LEA_2	15	2	1629	878	155	Cyt	PTS	22,41	6,29
XP_022895786.1	LEA_2	15	1	841	841	211	C_memb|LYSO/VAC	SP|TD	21,72	6,21
XP_022896449.1	LEA_2	15	2	25104	546	152	Cyt	NLS	20,38	6,14
XP_022896966.1	LEA_2	2	1	1046	1046	256	C_memb	TD	21,93	6,23
XP_022897316.1	SMP	15	3	2343	991	238	Cyt		23,07	6,25
XP_022897382.1	Dehydrin	15	2	846	749	195	Cyt|Nucleus	NLS	22,58	6,28
XP_022897396.1	LEA_3	15	2	694	412	102	MD	MTP	20,43	6,17
XP_022898650.1	Dehydrin	16	2	1258	910	189	Cyt|Nucleus|C_memb	NLS	22,43	6,24
XP_022898889.1	LEA_2	16	2	3556	2391	250	C_memb	TD	22,14	6,26
XP_022898891.1	LEA_2	16	1	921	921	207	C_memb	SP|TD	22,14	6,26
XP_022898968.1	LEA_2	16	1	1072	1072	258	C_memb	TD	21,72	6,21
XP_022899059.1	LEA_3	16	2	723	620	97	MD	MTP	23,37	6,29
XP_022899240.1	LEA_2	16	1	528	528	175	C_memb	SP|TD	24,08	6,32
XP_022899358.1	LEA_2	2	1	873	873	210	C_memb	SP|TD	23,81	6,3
XP_022899500.1	LEA_2	16	1	923	923	206	C_memb	SP|TD	23,12	6,25
XP_022899501.1	LEA_2	16	1	838	838	207	C_memb	SP|TD	23,12	6,25

The acronyms represent the predicted subcellular localization and features of the LEA proteins. C_memb: Cell membrane localization, SP: Signal peptide, TD: Transmembrane domain, MD: Mitochondrial domain, MTP: Mitochondrial transit peptide, Cyt: Cytoplasmic localization, Nucleus: Nuclear localization, NLS: Nuclear localization signal, NES: Nuclear export signal, ER: Endoplasmic reticulum localization, LYSO: Lysosome localization, VAC: Vacuole localization.

Gene Expression Analysis of LEA Proteins in Various Olive Tissues

The expression patterns of LEA proteins in different tissues provide essential insights into their potential roles in plant development and stress response. To investigate the expression profiles of LEA proteins across multiple tissue types, including leaf, root, stem, meristem, flower, and fruit, RNA-seq data was analyzed, and a heatmap was generated (Figure 1). The heatmap illustrates the differential expression levels of the identified LEA proteins, offering a comprehensive view of tissue-specific expression and clustering patterns.

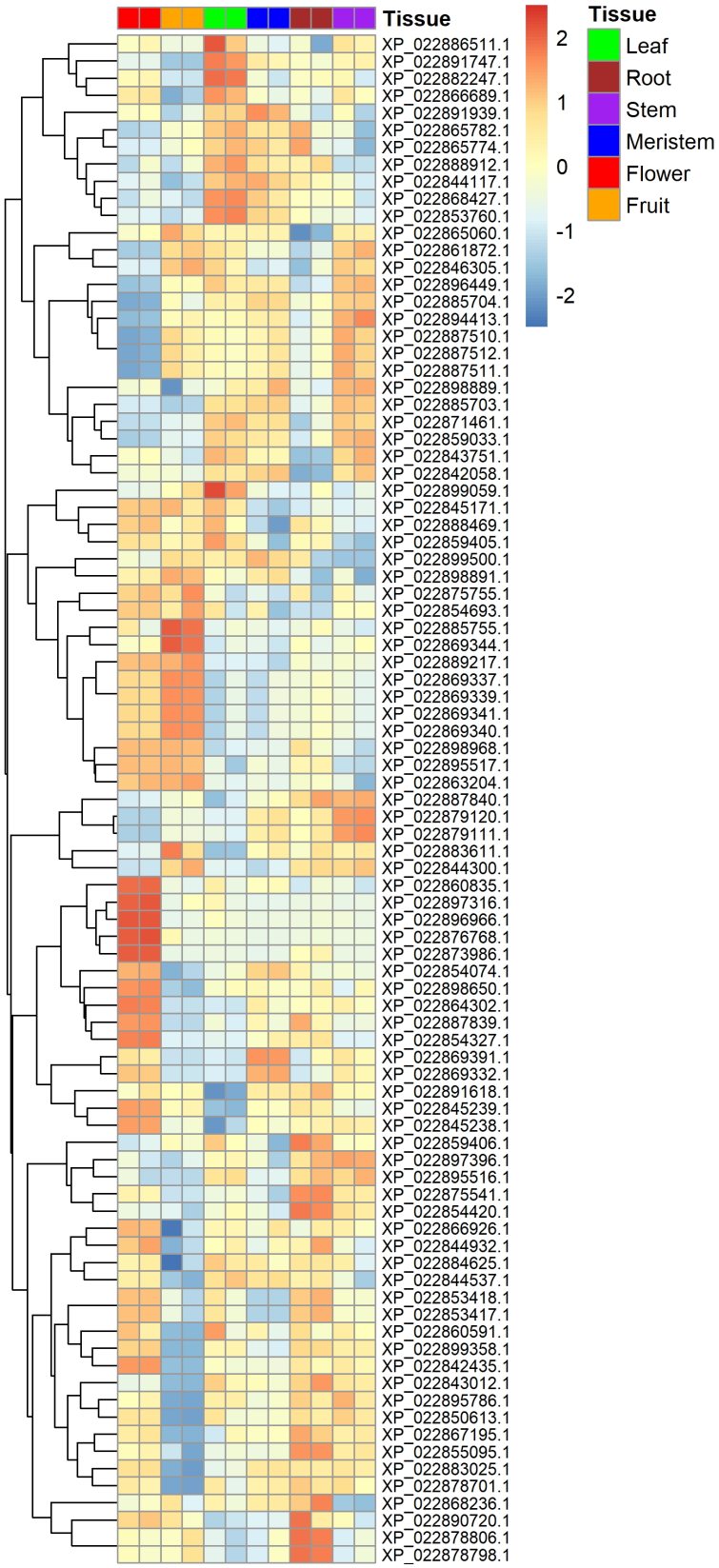

Figure 1. Heatmap showing expression of olive LEA genes in different tissues

In the current study, we analyzed the gene expression profiles of LEA proteins in different tissues of olive, focusing on their expression patterns across fruit, flower, root, stem, leaf, and meristem tissues. The heatmap provides a visual representation of the expression levels, allowing for the identification of tissue-specific or broadly expressed LEA proteins. Several clear expression

patterns emerged, highlighting the functional diversity and potential roles of these LEA proteins in different developmental stages and tissue-specific stress responses.

A subset of proteins was identified as having higher expression in fruit tissue, with notable examples including XP_022885755.1, XP_022869344.1, and XP_022869340.1. These proteins exhibited consistently elevated expression in fruit, suggesting a role in fruit development or stress responses during fruit maturation. Moreover, proteins such as XP_022889217.1 and XP_022898968.1 displayed equal expression in both fruit and flower, indicating a broader role in reproductive tissues.

Several LEA proteins were identified as flower-specific, showing predominant expression in flower tissue, such as XP_022860835.1, XP_022897316.1, and XP_022873986.1. These proteins are likely involved in protecting reproductive organs under stress conditions. Additionally, some proteins, including XP_022854074.1 and XP_022898650.1, were classified as mostly flower-specific, where expression is primarily observed in flower, but detectable to a lesser extent in other tissues.

In contrast, certain LEA proteins showed root-specific expression, such as XP_022878806.1 and XP_022875541.1, with significantly higher expression in root tissues. These proteins may play essential roles in root development and stress adaptation, particularly in drought-prone environments where root function is critical for water uptake.

Another group of proteins exhibited higher expression in stem, including XP_022887840.1 and XP_022879120.1, highlighting their potential involvement in structural integrity and stress protection of supporting tissues. Several proteins were also expressed across multiple tissues, with proteins such as XP_022898889.1 and XP_022871461.1 showing expression in stem, leaf, and meristem. These broadly expressed proteins suggest a more general protective role across various tissues, contributing to the plant's overall stress tolerance mechanisms.

Phylogenetic Analysis and Expression Profiling of LEA Genes

To further investigate the evolutionary relationships and functional diversification of the LEA proteins in olive, we performed a phylogenetic analysis, constructing a circular cladogram based on the sequences of LEA proteins. This analysis provided insights into the clustering of LEA gene family members and revealed their structural conservation across different subgroups. Additionally, we integrated expression data from drought applications vs control, allowing us to examine the correlation between gene evolutionary history and tissue-specific expression patterns (Figure 2).

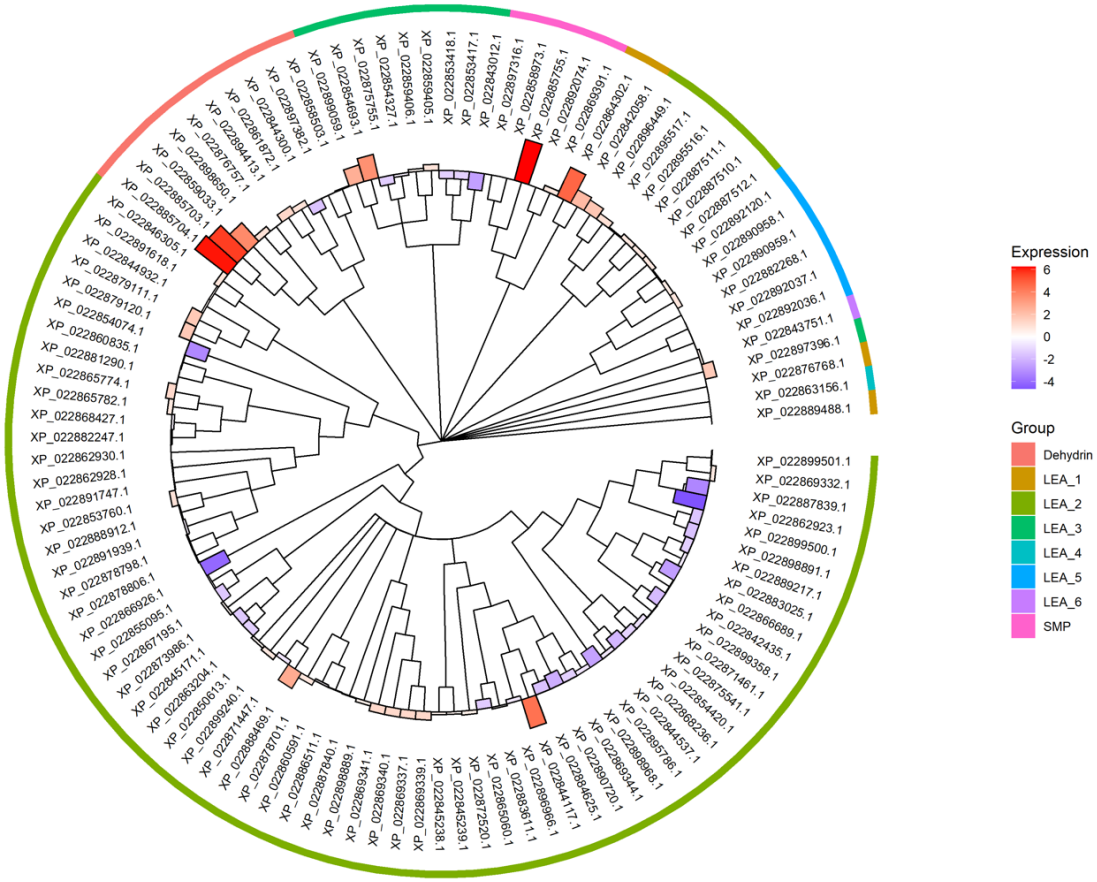

Figure 2. Phylogenetic tree of LEA proteins in olive with integrated differential expression and domain architecture.

The phylogenetic tree in olive presents a comprehensive overview of the Late Embryogenesis Abundant (LEA) gene family, integrating sequence similarity, conserved domain architecture, and gene expression under drought conditions. Based on RNA-seq data, the tree reveals distinct patterns of differential expression in response to drought, providing insights into the functional roles of specific LEA subgroups.

Notably, the Dehydrins, represented by genes such as XP_022885704.1 and XP_022885703.1, exhibit the highest levels of overexpression under drought stress, with log2 fold changes exceeding 6. These findings are consistent with the well-documented role of Dehydrins in protecting cellular structures from dehydration by stabilizing membranes and scavenging reactive oxygen species during abiotic stress. Another overexpressed gene, XP_022885755.1, belongs to the SMP (Seed Maturation Protein) family, further highlighting the involvement of multiple LEA subgroups in drought response. The overexpression of SMP genes, such as XP_022864302.1, suggests a broader protective function, potentially supporting both seed maturation and drought adaptation.

In contrast, certain LEA_2 genes, including XP_022862923.1, demonstrate a trend of lower expression under drought conditions, with fold changes ranging from 0 to -4. This downregulation indicates that not all LEA subgroups contribute equally to the plant's drought stress response. The differential expression patterns of LEA_2 genes could suggest a more specialized role, potentially linked to tissue-specific functions or varying environmental stress responses.

Overall, the phylogenetic clustering aligns with the expression profiles, where specific subgroups such as Dehydrins and SMPs form well-defined clades, exhibiting either high upregulation or more moderate expression levels in drought conditions. The clustering of LEA_2 genes showing different expression patterns further supports their distinct roles within the broader LEA gene family. This analysis underscores the functional diversity within the LEA gene family and highlights key genes that may play pivotal roles in enhancing drought tolerance in olive.

Conclusion

This study presents a comprehensive analysis of the Late Embryogenesis Abundant (LEA) proteins in olive, focusing on their phylogenetic relationships and expression profiles across various tissues and in response to drought stress. Our findings provide new insights into the functional roles of LEA proteins in olive, a species frequently exposed to harsh drought conditions, particularly in Mediterranean climates.

In recent studies, the LEA gene family has been identified in multiple plant species, showcasing a wide range of gene numbers. For instance, a comprehensive analysis in Juglans regia identified 51 LEA members (Ma et al., 2023), while similar studies reported 73 in watermelon (Altunoğlu et al., 2017), 61 in melon (Altunoğlu et al., 2017), and 108 in Brassica napus (Yu et al., 2016). This variability in gene count across species suggests that gene duplication events and evolutionary pressures have shaped the LEA gene family, allowing plants to adapt to their specific environmental conditions. The classification of LEA proteins into various groups based on their conserved motifs and amino acid sequences further emphasizes their functional diversity (Aziz et al., 2021; Battaglia and Covarrubias, 2013).

In olive a total of 111 LEA proteins were identified, with LEA_2 being the most common domain, comprising over 30% of the LEA family in olive. These LEA_2 proteins were distributed across several chromosomes, with no clear clustering, indicating that LEA genes may have evolved independently across the genome to provide widespread protection under diverse stress conditions. The identified proteins also showed significant diversity in their exon count, genomic length, and subcellular localization. The predominance of transmembrane domain-containing proteins (48 proteins) suggests that membrane stabilization plays a crucial role in the olive tree's ability to survive water deficit. Additionally, 20 proteins localized to mitochondria, highlighting the importance of energy homeostasis under stress. These results align with previous reports on LEA proteins' roles in stabilizing cellular structures during dehydration (Hanin et al., 2011; Hundertmark and Hincha, 2008).

Our gene expression analysis revealed distinct tissue-specific expression patterns, which may indicate specialized roles for LEA proteins in different parts of the plant. For instance, several LEA proteins were highly expressed in fruit, including XP_022885755.1 and XP_022869344.1, suggesting a potential role in fruit development and maturation under stress conditions. These findings are consistent with the known protective functions of LEA proteins in reproductive organs (Sun et al., 2021). In contrast, other LEA proteins were predominantly expressed in flowers, such as XP_022860835.1 and XP_022897316.1. The high expression of these proteins in flowers may suggest a protective role during pollination or seed formation, which are critical periods for drought-sensitive crops like olive.

In root tissues, proteins such as XP_022878806.1 and XP_022875541.1 showed the highest expression levels, highlighting their role in root adaptation to water scarcity. Given that roots are the primary organs for water uptake, these proteins are likely involved in protecting root cells from desiccation and ensuring continuous water transport. Similarly, LEA proteins like XP_022887840.1 and XP_022879120.1 were highly expressed in stems, where they may play a role in maintaining structural integrity and facilitating water transport under drought stress.

The phylogenetic analysis revealed evolutionary conservation across different LEA subgroups, with distinct clustering of LEA_1, LEA_2, LEA_3, and Dehydrin proteins. The most significant finding from the phylogenetic tree was the high overexpression of Dehydrins, such as XP_022885704.1 and XP_022885703.1, under drought conditions. Dehydrins are well-known for their role in protecting plants from desiccation, and their strong upregulation in response to drought stress is consistent with previous studies in other species (Brini et al., 2007; Hanin et al., 2011). Interestingly, several SMP proteins, including XP_022885755.1, also showed significant overexpression during drought, suggesting that these proteins may have broader roles beyond seed maturation, possibly contributing to drought adaptation mechanisms in other tissues.

On the other hand, LEA_2 proteins exhibited more variable expression, with some, like XP_022862923.1, showing downregulation under drought conditions. This variability could indicate that LEA_2 proteins are more specialized, potentially fulfilling tissue-specific roles or responding to types of stress. Further functional studies will be required to determine the precise mechanisms through which these proteins contribute to stress tolerance.

In conclusion, this study sheds light on the functional diversity and evolutionary history of the LEA gene family in olive. The integration of phylogenetic analysis with gene expression data provides a clearer understanding of how different LEA subgroups contribute to drought tolerance across various tissues. Future research should focus on the functional characterization of key LEA proteins, particularly those that are highly expressed in drought-sensitive tissues such as roots and reproductive organs. These findings could inform breeding programs aimed at improving the resilience of olive and other drought-prone crops through the targeted manipulation of LEA genes.

Authors' Contributions

Authors declare that they have contributed equally to the article.

Conflicts of Interest Statement

The authors declare that there is no conflict of interest.

References

Altunoğlu, Y. Ç., Baloğlu, M. C., Baloglu, P., Yer, E. N., Kara, S., 2017. Genome-wide identification and comparative expression analysis of lea genes in watermelon and melon genomes. Physiology and Molecular Biology of Plants. 23(1): 5-21.

Aziz, M. A., Sabeem, M., Kutty, M. S., Brini, F., Masmoudi, K., 2021. Plant group ii lea proteins: intrinsically disordered structure for multiple functions in response to environmental stresses. Biomolecules. 11(11): 1662.

Battaglia, M., Covarrubias, A. A., 2013. Late embryogenesis abundant (lea) proteins in legumes. Frontiers in Plant Science. 4: Article 190.

Ben Abdallah, M., Trupiano, D., Polzella, A., De Zio, E., Sassi, M., Scaloni, A., Zarrouk, M., Ben Youssef, N., Stefania Scippa, G., 2018. Unraveling physiological, biochemical and molecular mechanisms involved in olive (Olea europaea L. cv. Chétoui) tolerance to drought and salt stresses. Journal of Plant Physiology. 220: 83–95.

Brini, F., Hanin, M., Lumbreras, V., Amara, I., Khoudi, H., Hassairi, A., Pagès, M., Masmoudi, K., 2007. Overexpression of wheat dehydrin DHN-5 enhances tolerance to salt and osmotic stress in Arabidopsis thaliana. Plant Cell Reports. 26(11): 2017–2026.

Gholami, R., Fahadi Hoveizeh, N., Zahedi, S. M., Gholami, H., Carillo, P., 2022. Effect of three water-regimes on morpho-physiological, biochemical and yield responses of local and foreign olive cultivars under field conditions. BMC Plant Biology. 22(1): 477.

Hand, S. C., Menze, M. A., Toner, M., Boswell, L., Moore, D., 2011. LEA proteins during water stress: not just for plants anymore. Annual Review of Physiology. 73(1): 115–134.

Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., Masmoudi, K., 2011. Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. Plant Signaling & Behavior. 6(10): 1503–1509.

Huang, R., Xiao, D., Wang, X., Zhan, J., Wang, A., He, L., 2022. Genome-wide identification, evolutionary and expression analyses of LEA gene family in peanut (Arachis hypogaea L.). BMC Plant Biology. 22(1): 155.

Hundertmark, M., Hincha, D. K., 2008. LEA (Late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics. 9(1): 118.

Jia, C., Guo, B., Wang, B., Li, X., Yang, T., Li, N., Wang, J., Yu, Q., 2022. 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): 596.

Jia, J. S., Ge, N., Wang, Q. Y., Zhao, L. T., Chen, C., Chen, J. W., 2023. 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): 126.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A., 2012, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12):1647-1649. 10.1093/bioinformatics/bts199. Epub 2012 Apr 27. PMID: 22543367; PMCID: PMC3371832.

Li, Z., Chi, H., Liu, C., Zhang, T., Han, L., Li, L., Pei, X., Long, Y., 2021. Genome-wide identification and functional characterization of LEA genes during seed development process in linseed flax (Linum usitatissimum L.). BMC Plant Biology. 21(1): 193.

Lin, Y., She, M., Zhao, M., Yu, H., Xiao, W., Zhang, Y., Li, M., 2024. Genome-wide analysis and functional validation reveal the role of late embryogenesis abundant genes in strawberry (Fragaria × ananassa) fruit ripening. BMC Genomics. 25(1): 228.

Ma, J., Zuo, D., Ye, H., Yan, Y., Li, M., Zhao, P., 2023. 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).

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. N., Lopez, R., Butcher, S., 2024. The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. Nucleic Acids Research. 52(W1): W521–525.

Magwanga, R. O., Lu, P., Kirungu, J. N., Lu, H., Wang, X., Cai, X., Zhou, Z., 2018. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. BMC Genetics. 19(1): 6.

Ødum, M. T., Teufel, F., Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Winther, O., Nielsen, H., 2024. DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. Nucleic Acids Research. 52(W1): W215–220.

Peng, Y., Reyes, J. L., Wei, H., Yang, Y., Karlson, D., Covarrubias, A. A., Krebs, S. L., Fessehaie, A., Arora, R., 2008. RcDhn5, a cold acclimation-responsive dehydrin from Rhododendron catawbiense rescues enzyme activity from dehydration effects in vitro and enhances freezing tolerance in RcDhn5-overexpressing Arabidopsis plants. Physiologia Plantarum. 134(4): 583–597.

Ramírez-Tejero JA., Jiménez-Ruiz J., Leyva-Pérez MO, Barroso JB., Luque F., 2020. Gene Expression Pattern in Olive Tree Organs (Olea europaea L.). Genes (Basel). 12;11(5):544. 10.3390/genes11050544. PMID: 32408612; PMCID: PMC7291012.

Rico, E. I., Martos De La Fuente, G. C., Ortega Morillas, A., Fernández Ocaña, A. M., 2023. Physiological and biochemical study of the drought tolerance of 14 main olive cultivars in the Mediterranean Basin. Photosynthesis Research.

RoyChoudhury, A., Roy, C., Sengupta, D. N., 2007. Transgenic tobacco plants overexpressing the heterologous lea gene Rab16A from rice during high salt and water deficit display enhanced tolerance to salinity stress. Plant Cell Reports. 26(10): 1839–1859.

Sun, Z., Li, S., Chen, W., Zhang, J., Zhang, L., Sun, W., Wang, Z., 2021. Plant dehydrins: expression, regulatory networks, and protective roles in plants challenged by abiotic stress. International Journal of Molecular Sciences. 22(23): 12619.

Tsamir-Rimon, M., Ben-Dor, S., Feldmesser, E., Oppenhimer-Shaanan, Y., David-Schwartz, R., Samach, A., Klein, T., 2021. Rapid starch degradation in the wood of olive trees under heat and drought is permitted by three stress-specific beta amylases. New Phytol. 229(3):1398-1414. doi: 10.1111/nph.16907. Epub 2020 Oct 1. PMID: 32880972.

Unver, T., Wu, Z., Sterck, L., Turktas, M., Lohaus, R., Li, Z., Yang, M., et al., 2017. Genome of wild olive and the evolution of oil biosynthesis. Proceedings of the National Academy of Sciences. 114(44).

Wang, W., Liu, Y., Kang, Y., Liu, W., Li, S., Wang, Z., Xia, X., et al., 2024. Genome-wide characterization of LEA gene family reveals a positive role of BnaA.LEA6.a in freezing tolerance in rapeseed (Brassica napus L.). BMC Plant Biology. 24(1): 433.

Yu, L., Xiong, Z., Zheng, J., Xu, D., Zhu, Z. W., Xiang, J., … Li, M., 2016. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (lea) gene family formation pattern in Brassica napus. Scientific Reports. 6(1).

Zhang, Y., Fan, N., Wen, W., Liu, S., Mo, X., An, Y., Zhou, P., 2022. Genome-wide identification and analysis of LEA_2 gene family in alfalfa (Medicago sativa L.) under aluminum stress. Frontiers in Plant Science. 13: 976160.