Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi
Journal of Agricultural Faculty of Gaziosmanpasa University
https://dergipark.org.tr/tr/pub/gopzfd

Araştırma Makalesi/*Research Article*

JAFAG
ISSN: 1300-2910
E-ISSN: 2147-8848
(2024) 41(2) 86-97
doi:10.55507/gopzfd. 1517870

Bioinformatic Analysis of LEA Genes in Stout Camphor Tree (*Cinnamomum micranthum* f. *Kanehirae*)

Tevfik H. CAN[1]* Tamer KUŞAKSIZ[1]

Manisa Celal Bayar University Alaşehir Vocational School, Department of Plant and Animal Production, Alaşehir, Manisa
***Corresponding author's email:** tevfik.can@cbu.edu.tr**

Alındığı tarih (Received): 18.07.2024 Kabul tarihi (Accepted): 01.08.2024

Abstract: LEA proteins have an important role in the response of plants to abiotic stresses. *Cinnamomum micranthum* f. *kanehirae*, a medicinal and aromatic plant belonging to the *Lauraceae* family. The genome sequence of the Kanehirae or Stout Camphor tree was recently completed. Although there are studies on its genome, there are no studies on LEA genes.

57 LEA genes (CmiLEA) were identified in the Stout Camphor genome. CmiLEA was divided into 8 distinct clusters based on phylogenetic analysis. When the subcellular localizations of CmiLEA were examined, they were found to be localized mostly in the cytoplasm. A total of 13 genes targeting only one miRNA were identified. In CmiLEA, a total of 23 genes were found to have only exon regions and no introns. In total, 35 conserved motifs were identified, while there was only one conserved motif in CmiLEA-42. Consistent with the 3D structure results, CmiLEA-21, CmiLEA-31, CmiLEA-44, CmiLEA-45, and CmiLEA-57 from the LEA-2 subfamily showed over 90% accuracy.

The present study was the first *in-silico* analysis of LEA genes in *Cinnamomum micranthum* f. *Kanehirae*. It is thought that it may form a base for advanced functional analysis in *Cinnamomum* in future.

Keywords: Bioinformatics, LEA genes, LEA proteins, miRNA, Stout Camphor Tree

Stout Kafur Ağacında LEA Genlerinin Biyoinformatik Analizi (*Cinnamomum micranthum f. Kanehirae*)

Öz: LEA proteinleri bitkilerin abiyotik streslere karşı tepkilerinde önemli bir role sahiptir. *Lauraceae* ailesine ait tıbbi ve aromatik bir bitki olan *Cinnamomum micranthum* f. *Kanehirae* veya Stout Kafur ağacının genom dizisi yakın zamanda tamamlanmıştır. Stout Kafur genomunda çalışmalar olmasına rağmen LEA genleri ile alakalı herhangi bir çalışma bulunmamaktadır. Bu nedenle bu çalışmada biyoinformatik araçlar kullanılarak Stout Kafur genomunda yer alan LEA genlerinin genom çapında analizinin yapılması amaçlanmıştır.

Stout Kafur genomunda 57 LEA geni (CmiLEA) tanımlandı. CmiLEA filogenetik analize göre 8 ayrı kümeye ayrılmıştır. CmiLEA'nın subselüler lokalizasyonları incelendiğinde daha çok sitoplazmada lokalize oldukları ve sadece bir miRNA hedefleyen toplam 13 gen tanımlanmıştır. CmiLEA' da toplam 23 genin yalnızca ekzon bölgelerine sahip olduğu ve intronsuz olduğu tespit edilmiştir. Toplamda 35 korunmuş motif belirlenirken, CmiLEA-42'de yalnızca bir korunmuş motif bulunmuştur. 3B yapı sonuçlarına uygun olarak LEA_2 alt ailesinden CmiLEA-21, CmiLEA-31, CmiLEA-44, CmiLEA-45 ve CmiLEA-57 %90'ın üzerinde doğruluk göstermiştir.

Bu çalışma, *Cinnamomum micranthum* f. *kanehirae* bitkisinde LEA genleri ile ilgili yapılmış ilk biyoinformatik çalışma olup, *Cinnamomum* cinsinde gelecekte ileri fonksiyonel analizler için bir temel oluşturabileceği düşünülmektedir.

Anahtar Kelimeler: Biyoinformatik, LEA genleri, LEA proteinleri, miRNA, Stout Kafur Ağacı

Introduction

Lauraceae is a family of tropical plants consisting of approximately 2500 species of trees and shrubs in 55 genera. Since members of the *Cinnamomum* are rich in essential oils, they are used in perfume making, spices production and compound for alternative medicine around the world. Studies have shown that they are rich in terpenoids and phenylpropanoids (Dong Wang, 2022). *Cinnamomum micranthum* f. *Kanehirae* or Stout Camphor tree is a medicinal and aromatic plant and is also a valuable tree for forestry due to its rot-resistant trunk. It is a plant of ecological, agricultural and economic importance that grows in the Far East and is endemic in Taiwan. Stout camphor tree is also the only host of *Taiwanofungus camphoratus* which is used in traditional medicine. Studies in high-fat-fed mice reported the anti-inflammatory, anti-obesinogenic and antidiabetic effects of this rare fungus (Chung and Hsieh, 2023; Chang et al., 2018).

Although studies have been carried out to determine the composition of essential oils and to determine genetic relationships in the *Cinnamomum* by sequencing genes such as chloroplast genes, studies at the whole genome level are rare (Dong Wang, 2022). Li et al. (2023) constructed a high-quality reference genome in *Cinnamomum camphora* by whole genome resequencing and made a genomic comparison with *C. Kanehirae*. They identified phenylpropanoid metabolism genes related to cold stress and terpene synthases (TPSs) genes related to defense response. In genomic studies conducted in *Cinnamomum camphora*, it was determined that there are 36,411-24,883 proteins, number of functional annotations are 97.06%-82.71%, and the number of TPS genes is ranged between 72-85 (Shen et al. 2022; Sun et al. 2022; Jiang et al. 2022; Wang et al., 2022; Li et al., 2023).

Late embryogenesis abundant (LEA) proteins were first found in cotton seeds during dehydration and maturation period (Cheng et al., 2021). Then, LEA proteins were detected in many plants such as *Arabidopsis*, poplar, peanut, tobacco, watermelon and melon. These proteins were also identified in organisms such as mosses and fungi (Hundertmark and Hintcha, 2008; Bies-Etheve et al., 2008; Altunoglu et al., 2017; Cheng et al., 2021; Huang et al., 2022; Geng et al., 2022). It has been determined that LEA proteins, which are involved in the response to abiotic stress factors, are found in many different parts of the cell, including the inner and outer membranes, cytoplasm and organelles (Altunoglu et al., 2017). It has been observed that LEA proteins, which are divided into 8 subfamilies ((LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6, DHN (Dehydrin) and SMP (seed maturation protein)) according to their conserved domains, are rapidly expressed in plant tissues in the face of stress factors such as drought, saline, or cold stress (Bies-Etheve et al., 2008). The studies show that LEA proteins are not tissue specific and are produced at different expression levels in tissues throughout developmental processes. These proteins are also known to have a highly hydrophilic structure and are thought to be intrinsically disordered proteins under normal physiological conditions, for example. It is estimated that these ordered structures serve as molecular chaperones and have an important role in ensuring cellular homoestasis by binding to molecules such as enzymes, ions, ROS, etc. (Hong-Bo, et al., 2005; Lin et al., 2021).

Abiotic stress factors like drought and salinity are important environmental factors affecting crop production. Examining the response to these stress factors plays an essential role in organizing the necessary breeding studies to develop resistant plants. At the same time, abiotic stress factors are important in the protection of endemic or endangered plants due to these stress factors's environmental impact. In this study, it was aimed to examine LEA genes and LEA proteins in the stout camphor tree with bioinformatic tools, classify them and determine their predicted functions.

Material and Methods

Material

The data related with Stout camphor tree retrieved from NCBI (The National Center for Biotechnology Information) database (NCBI, 2024).

Methods

Conserved domains in PFAM database were found by using CLC Genomic Workbench 21 (Qiagen, 2022). The sequences screened with the BLASTP tool (NCBI, 2024), and LEA proteins were identified and named. The characteristics of LEA proteins (isoelectric point, protein length, physical position in chromosomes, instability, etc.) were found through the ExPasy ProtPARAm tool (Gasteiger et al., 2005) Exon-intron regions of LEA protein genes were determined and visualized using Gene Structure Displayer Server (GSDS) 2.0 (Hu et al., 2015). The 3D structures of the detected proteins were determined using the Hidden Markov Model (HMM) algorithm in Protein Fold Recognition Server 2 (Phyre2), and 3D protein modeling was performed (Kelley et al., 2015). Sequences of LEA proteins were aligned using the ClustalW in the MEGA 11 program using default options (Tamura and Kumar, 2021). Conserved motifs in amino acid sequences and 3-dimensional structures of proteins were determined with the MEME Suite program (Bailey et al., 2021). Molecular function, subcellular localization and biological processes (Gene ontology) analyzes of LEA proteins were performed by using the Blast2Go (Conesa et al., 2005).

Micro RNA (miRNA) targeting LEA transcripts data retrieved from miRBase (*Arabidopsis thaliana*) (Kozomara and Griffith-Jones, 2013) and LEA genes in stout camphor tree were evaluated using the Plant Small RNA Target Analysis Server, psRNATarget (Dai et al. 2018).

Table 1. Characteristics of *Cinnamomum micranthum* f. *Kanerihae* LEA proteins
Tablo 1. *Cinnamomum micranthumf. KanerihaeLEA proteinlerine ait özellikler*

ID	Cinnamomum micranthum Genomics Database Identfier	Startposition (bp)	EndPosition (bp)	Proteinlength (aa)	pI	Molecular weight (Da)	Instability index	Stable or unstable	Subfamily	GRAVY	Hydropathy	Alifatic Index	Subcellular Localization
CmiLEA-01	RWR97808.1	14,925	20,172	228	9.75	24891.08	33.02	Stable	LEA_2	0.143	Hydrophobic	102.94	InnerMembrane
CmiLEA-02	RWR98064.1	24,801	30,831	165	5.87	17495.89	46.51	Unstable	LEA_5	-1.332	Hydrophilic	33.82	Extracellular, Cytoplasmic, Periplasmic
CmiLEA-03	RWR97823.1	40,810	41,605	209	9.57	22964.80	29.12	Stable	LEA_2	0.076	Hydrophobic	97.32	Cytoplasmic, Periplasmic
CmiLEA-04	RWR97684.1	158,790	160,389	107	5.90	11724.83	42.16	Unstable	LEA_6	-1.036	Hydrophilic	44.77	Extracellular
CmiLEA-05	RWR97695.1	329,333	357,397	188	4.74	20294.58	29.68	Stable	LEA_6	-0,469	Hydrophilic	78.40	Cytoplasmic
CmiLEA-06	RWR93166.1	2,196,823	2,198,729	232	9.23	25384.70	43.40	Unstable	LEA_2	0,033	Hydrophobic	102.50	OuterMembrane, Cytoplasmic
CmiLEA-07	RWR94634.1	3,169,570	3,170,099	95	9.86	10458.11	53.68	Unstable	LEA_3	-0,215	Hydrophilic	83.05	Cytoplasmic, Periplasmic
CmiLEA-08	RWR90147.1	4,913,657	4,914,889	225	9.49	24814.96	40.21	Unstable	LEA_2	0,142	Hydrophobic	101.24	InnerMembrane, Periplasmic
CmiLEA-09	RWR78722.1	5,396,142	5,397,052	252	9.97	27999.78	39.51	Stable	LEA_2	-0,1	Hydrophilic	101.27	OuterMembrane
CmiLEA-10	RWR96356.1	6,195,381	6,196,184	144	5.29	15485.97	25.87	Stable	LEA_4	-1,165	Hydrophilic	44.44	Periplasmic
CmiLEA-11	RWR78857.1	7,768,911	7,769,440	79	9.96	8501.28	51.88	Unstable	LEA_5	-1,358	Hydrophilic	38.35	Cytoplasmic
CmiLEA-12	RWR90377.1	8,575,238	8,586,288	445	5.91	50415.56	57.70	Unstable	LEA_2	-0,527	Hydrophilic	85.84	Cytoplasmic
CmiLEA-13	RWR84490.1	9,743,858	9,744,653	187	9.28	20810.37	31.46	Stable	LEA_2	-0,136	Hydrophilic	106.26	Cytoplasmic
CmiLEA-14	RWR82192.1	10,105,899	10,106,296	108	6.06	11361.74	31.84	Stable	SMP	-0,253	Hydrophilic	75.19	Periplasmic, Extracellular, Cytoplasmic
CmiLEA-15	RWR82193.1	10,108,755	10,117,797	313	4.72	32423.09	42.50	Unstable	SMP	-0,413	Hydrophilic	71.50	Periplasmic
CmiLEA-16	RWR72447.1	11,010,992	11,015,343	306	9.54	33583.29	59.07	Unstable	LEA_2	-0,344	Hydrophilic	75.13	OuterMembrane
CmiLEA-17	RWR82287.1	11,800,766	11,801,561	199	9.51	22673.34	55.08	Unstable	LEA_2	0,102	Hydrophobic	108.69	InnerMembrane
CmiLEA-18	RWR82288.1	11,808,328	11,815,810	209	9.67	23744.66	46.42	Unstable	LEA_2	-0,028	Hydrophilic	85.89	InnerMembrane
CmiLEA-19	RWR88431.1	12,126,011	12,126,806	187	8.55	20550.31	25.88	Stable	LEA_2	0,444	Hydrophobic	113.10	InnerMembrane, Cytoplasmic
CmiLEA-20	RWR82316.1	12,177,214	12,179,889	182	9.30	20054.09	50.70	Unstable	LEA_2	-0,005	Hydrophilic	101.92	Cytoplasmic
CmiLEA-21	RWR96700.1	12,212,253	12,230,276	398	5.42	44448.15	30.01	Stable	LEA_2	-0,309	Hydrophilic	97.66	Cytoplasmic
CmiLEA-22	RWR82451.1	14,709,244	14,710,366	132	9.43	13804.19	31.68	Stable	Dehydrin	-1,049	Hydrophilic	51.74	Cytoplasmic, Periplasmic, Extracellular
CmiLEA-23	RWR93426.1	18,311,369	18,312,164	185	9.08	20332.85	36.67	Stable	LEA_2	0,401	Hydrophobic	110.05	OuterMembrane, Extracellular
CmiLEA-24	RWR93427.1	18,337,243	18,338,038	185	9.08	20332.85	36.67	Stable	LEA_2	0,401	Hydrophobic	110.05	OuterMembrane, InnerMembrane, Extracellular
CmiLEA-25	RWR93452.1	19,621,530	19,622,346	226	9.39	26175.93	39.05	Stable	LEA_2	-0,28	Hydrophilic	75.09	Cytoplasmic, InnerMembrane
CmiLEA-26	RWR82748.1	19,647,065	19,648,003	260	10.51	28235.88	53.68	Unstable	LEA_2	-0,196	Hydrophilic	84.00	OuterMembrane, Extracellular, Periplasmic
CmiLEA-27	RWR93456.1	19,880,530	19,881,325	211	9.49	23938.68	33.16	Stable	LEA_2	-0,146	Hydrophilic	77.16	Periplasmic, Extracellular, OuterMembrane
CmiLEA-28	RWR82951.1	23,165,577	23,165,974	104	4.79	10912.80	53.26	Unstable	LEA_6	-1,004	Hydrophilic	46.92	Extracellular, Periplasmic, Cytoplasmic
CmiLEA-29	RWR97091.1	24,384,561	24,385,374	225	9.24	25852.89	33.88	Stable	LEA_2	-0,264	Hydrophilic	86.22	Cytoplasmic, InnerMembrane

Table 1. Characteristics of *Cinnamomum micranthum* f. *Kanerihae* LEA proteins

Tablo 1. *(Devam) Cinnamomum micranthumf. KanerihaeLEA proteinlerine ait özellikler*

ID	Cinnamomum micranthum Genomics Database Identfier	Startposition (bp)	EndPosition (bp)	Proteinlength (aa)	pI	Molecular weight (Da)	Instability index	Stable or unstable	Subfamily	GRAVY	Hydropathy	Alifatic Index	Subcellular Localization
CmiLEA-30	RWR97092.1	24,459,404	24,460,199	205	9.30	22663.15	55.80	Unstable	LEA_2	0,108	Hydrophobic	100.78	OuterMembrane, InnerMembrane
CmiLEA-31	RWR83165.1	27,151,221	27,154,379	314	4.94	34561.31	16.38	Stable	LEA_2	-0.279	Hydrophilic	94.30	Cytoplasmic
CmiLEA-32	RWR92255.1	28,749,904	28,753,959	435	4.94	47980.19	41.69	Unstable	LEA_4	-1.371	Hydrophilic	42.46	Cytoplasmic, Extracellular, Periplasmic
CmiLEA-33	RWR92404.1	33,140,885	33,143,809	222	5.14	24864.70	59.50	Unstable	Dehydrin	-1.364	Hydrophilic	51.80	Cytoplasmic
CmiLEA-34	RWR95843.1	36,270,638	36,271,721	236	9.71	26828.53	35.39	Stable	LEA_2	0.149	Hydrophobic	99.87	OuterMembrane, InnerMembrane
CmiLEA-35	RWR94184.1	36,359,405	36,360,200	102	7.90	10846.32	44.12	Unstable	LEA_3	-0.331	Hydrophilic	82.25	Periplasmic
CmiLEA-36	RWR92640.1	37,837,684	37,838,591	251	9.83	28205.10	46.78	Unstable	LEA_2	-0.067	Hydrophilic	98.61	OuterMembrane
CmiLEA-37	RWR85244.1	39,510,036	39,510,849	225	9.24	25221.13	35.03	Stable	LEA_2	-0.095	Hydrophilic	90.49	OuterMembrane
CmiLEA-38	RWR85246.1	39,547,897	39,548,976	198	9.94	22718.71	49.87	Unstable	LEA_2	-0.013	Hydrophilic	109.19	InnerMembrane
CmiLEA-39	RWR94357.1	39,965,933	39,967,980	193	5.17	20953.01	48.76	Unstable	LEA_6	-0.891	Hydrophilic	51.09	Extracellular, Periplasmic
CmiLEA-40	RWR94359.1	40,152,225	40,152,622	95	5.27	10302.32	35.69	Stable	LEA_6	-1.125	Hydrophilic	41.16	Periplasmic, Cytoplasmic
CmiLEA-41	RWR89543.1	43,806,867	43,807,662	210	9.73	23303.20	45.63	Unstable	LEA_2	0.061	Hydrophobic	104.38	Periplasmic, OuterMembrane, InnerMembrane
CmiLEA-42	RWR76501.1	45,645,122	45,652,333	131	9.91	14070.92	33.56	Stable	LEA_3	-0.256	Hydrophilic	87.71	Periplasmic, Cytoplasmic
CmiLEA-43	RWR87982.1	48,225,122	48,225,917	209	9.53	23079.04	27.63	Stable	LEA_2	0.097	Hydrophobic	99.19	Cytoplasmic
CmiLEA-44	RWR80213.1	50,510,906	50,530,057	151	4.81	16361.70	12.66	Stable	LEA_2	0.007	Hydrophobic	101.99	Cytoplasmic
CmiLEA-45	RWR80217.1	50,671,714	50,689,053	151	4.81	16373.76	11.84	Stable	LEA_2	0.041	Hydrophobic	104.57	Cytoplasmic, Periplasmic
CmiLEA-46	RWR80282.1	53,670,451	53,671,246	193	9.77	21062.14	40.20	Unstable	LEA_2	0.160	Hydrophobic	80.26	Extracellular, InnerMembrane, OuterMembrane
CmiLEA-47	RWR80488.1	58,855,002	58,861,298	191	10.71	21715.28	57.16	Unstable	LEA_5	-1.142	Hydrophilic	48.48	Cytoplasmic, Extracellular, Periplasmic
CmiLEA-48	RWR80547.1	59,952,002	59,952,797	210	9.84	23089.12	44.40	Unstable	LEA_2	0.187	Hydrophobic	101.57	InnerMembrane
CmiLEA-49	RWR80548.1	59,955,049	59,963,042	240	10.33	26599.21	39.60	Stable	LEA_2	-0.124	Hydrophilic	99.83	OuterMembrane
CmiLEA-50	RWR73823.1	61,242,699	61,243,494	129	9.22	13693.15	26.97	Stable	LEA_1	-0.999	Hydrophilic	49.30	Periplasmic
CmiLEA-51	RWR80844.1	65,778,793	65,779,588	203	9.42	22242.25	42.94	Unstable	LEA_2	0.499	Hydrophobic	103.99	InnerMembrane
CmiLEA-52	RWR80845.1	65,791,299	65,792,094	210	9.63	23339.98	39.22	Stable	LEA_2	-0.024	Hydrophilic	90.95	OuterMembrane, Extracellular, Periplasmic
CmiLEA-53	RWR77559.1	66,813,066	66,823,071	225	7.73	25799.60	45.25	Unstable	LEA_2	-0.116	Hydrophilic	99.16	Cytoplasmic
CmiLEA-54	RWR81329.1	74,132,580	74,133,109	98	9.15	10514.45	47.81	Unstable	LEA_5	-1.393	Hydrophilic	33.98	Cytoplasmic
CmiLEA-55	RWR78006.1	74,571,090	74,574,233	108	9.05	12415.03	34.82	Stable	LEA_3	-0.807	Hydrophilic	70.46	Cytoplasmic
CmiLEA-56	RWR78102.1	76,142,313	76,143,259	262	10.19	28821.31	45.90	Unstable	LEA_2	-0.227	Hydrophilic	85.92	OuterMembrane
CmiLEA-57	RWR81477.1	76,528,885	76,529,931	290	5.09	31166.17	23.32	Stable	LEA_2	0.338	Hydrophobic	107.55	OuterMembrane

Results

LEA genes in *Cinnamomum micranthum* f. *Kanehirae* have been named, and their characteristics such as starting and ending positions of genes in base pairs, isoelectric points, protein lengths, molecular weights, stability, hydropathy have been determined and are given in Table 1. When the results were evaluated, 57 LEA genes were found and named. The chromosome locations of all of these genes were found as scaffolds. LEA genes are divided into 8 subfamilies according to their sequence homologies and conserved domains in the PFAM database. It is classified as LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6, dehydrin and SMP. While it was found that there was 1 gene belonging to the LEA_1 subfamily and 37 genes in the LEA_2 subfamily, 2 genes were detected in the LEA_4, SMP and Dehydrin subfamilies, 3 genes in the LEA_3 gene family, 4 genes in LEA_5 and 5 genes in the LEA_6 gene family, respectively. Since the chromosomal distribution was determined as scaffolds, the exact location of LEA genes on chromosomes has not been revealed. The largest protein length was detected as 445 aa in CmiLEA-12, the shortest protein length was determined as 79 aa in CmiLEA-11, similarly the highest molecular weight (50415 Da) and the lowest molecular weight (8501 da) were calculated in CmiLEA-12 and CmiLEA-11. According to the isoelectric points of the proteins, the lowest isoelectric point was observed in CmiLEA-15 with 4.72, and the highest isoelectric point was observed in CmiLEA-47 with 10.70. The instability index values showed 27 of proteins were stable and 30 of proteins were unstable. The results of hydropathy properties revealed 40 of proteins were hydrophilic and 17 of proteins were hydrophobic. In the Aliphatic index, which is suggested to increase the thermostability of globular proteins, the lowest value was found in CmiLEA-2 with 33.82, and the highest value was found in CmiLEA-19 with the value of 113.10.

The longest upstream/downstream region was found in CmiLEA-40 as the gene was also the longest gene, and total of 23 genes has only exon regions and they were found to be intronless. A total of 35 conserved motifs were determined while CmiLEA-42 had only one conserved motif. Although there are common motifs in general, the motifs showed differences in LEA subfamilies. The highest number of motifs was detected in CmiLEA-43 and CmiLEA-03, the constructed dendrogram shows consistency with the motif patterns. The distribution of exons and introns, conserved motifs and the phylogenetic tree of CmiLEAs are given in Figure 1, Figure 2 and Figure 3, respectively.

In the phylogenetic tree of CmLEA genes, different subfamilies are shown in different colors. It has been observed that it is divided into eight subfamilies. While all families are grouped among themselves; CmLEA 42 and CmLEA-55 genes, which are belonged to the LEA_3 subfamily, are placed in the LEA_2 group. It was determined that the LEA_2 subfamily formed a separate cluster in the phylogenetic tree, LEA_3 was a separate group with a single branch, and the remaining gene subfamilies grouped in a different cluster.

When subcellular localizations were examined, total of 13 proteins localized in cytoplasm (CmiLEA-33, CmiLEA- 12, CmiLEA-13, CmiLEA-20, CmiLEA-21, CmiLEA-31, CmiLEA-43, CmiLEA-44, CmiLEA-53, CmiLEA-55, and CmiLEA-11). Only CmiLEA-04 from the LEA_6 family was found to localize extracellularly. CmiLEA-01, CmiLEA-17, CmiLEA-18, CmiLEA-38, CmiLEA-48, and CmiLEA-51 observed in the inner membrane and all of them were members of the LEA_2 family. It has been found that there are 7 members of the LEA_2 subfamily in the outer membrane (CmiLEA-09, CmiLEA-16, CmiLEA-36, CmiLEA-37, CmiLEA-49, CmiLEA-56, CmiLEA-57), 6 of which are hydrophilic and 1 is hydrophobic. CmiLEA-50, CmiLEA-35, CmiLEA-10, CmiLEA-15 from LEA_1, LEA_3, LEA_5 and SMP subfamilies were found to be located in periplasm. To accordance with structure results, CmiLEA-21, CmiLEA-31, CmiLEA-44, CmiLEA-45 and CmiLEA-57 from the LEA_2 subfamily showed accuracy above 90%. The 3-dimensional structure of these proteins were given in Figure 4.

In total, 220 miRNAs were associated with 52 CmiLEA genes. A total of 13 genes were identified with only one miRNA (CmiLEA-5, CmiLEA-6, CmiLEA-9, CmiLEA-14, CmiLEA-15, CmiLEA-17, CmiLEA-22, CmiLEA-23, CmiLEA-28, CmiLEA-35, CmiLEA-41, CmiLEA-51 CmiLEA-56; ath-miR842, ath-miR5632-5p, ath-miR447c-5p, ath-miR5652, ath-miR773a, ath-miR5658, ath-miR472-3p, ath-miR863-5p, ath-miR399c-5p, ath, respectively. -miR414, ath-miR8168, ath-miR156c-3p and ath-miR5657, respectively). There is no association with any miRNAs for 5 CmiLEA genes (CmiLEA-4, CmiLEA-20, CmiLEA-40, CmiLEA-49, CmiLEA-50). The association between CmiLEAs and miRNAs are given in Figure 5.

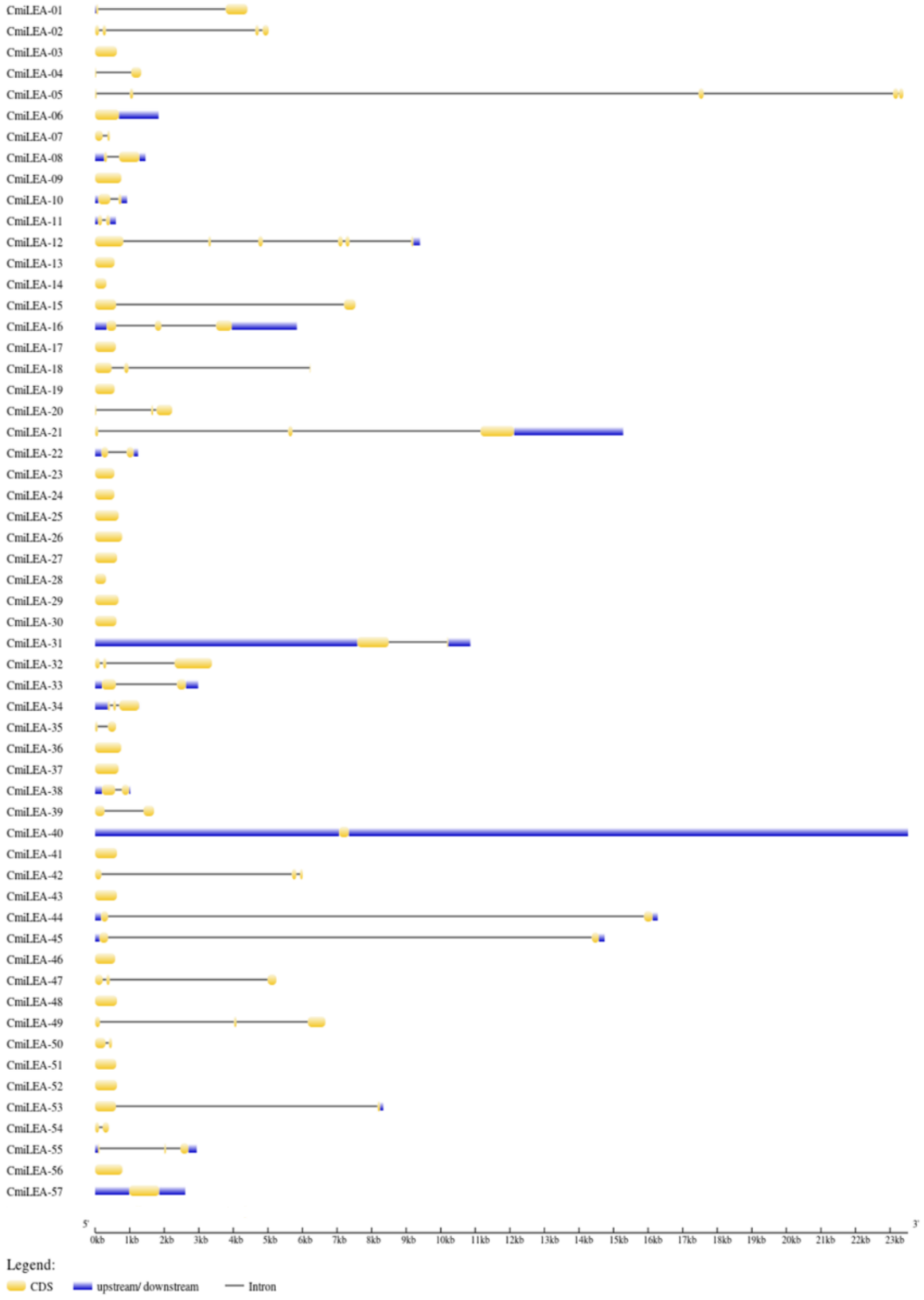

Figure 1. Exon and intron distributions of CmiLEA genes
Şekil 1. *CmiLEA genlerinde ekzon intron dağılımları*

Discussion

LEA proteins are linked with seed development and abiotic stress response. It has been reported that overexpression of LEA genes in transgenic plants shows an increased resistance to abiotic stress factors (Bies-Etheve et al., 2008). LEA genes are divided into 8 groups in plants according to their conservative PFAM

domains, these are LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6, DHN (Dehydrin) and SMP (seed maturation protein) (Geng et al., 2022). In our study, 57 CmiLEA genes divided in 8 subfamilies. However, the classification made according to repeated conserved domains, 9 subfamilies were detected in *Arabidopsis thaliana*, while 8 subfamilies were found in many other studies. In linseed flax (*Linum usitatissimum* L.) fifty LEA genes (LuLEA) were determined and these genes were divided into 8 classes (Li et al., 2021).

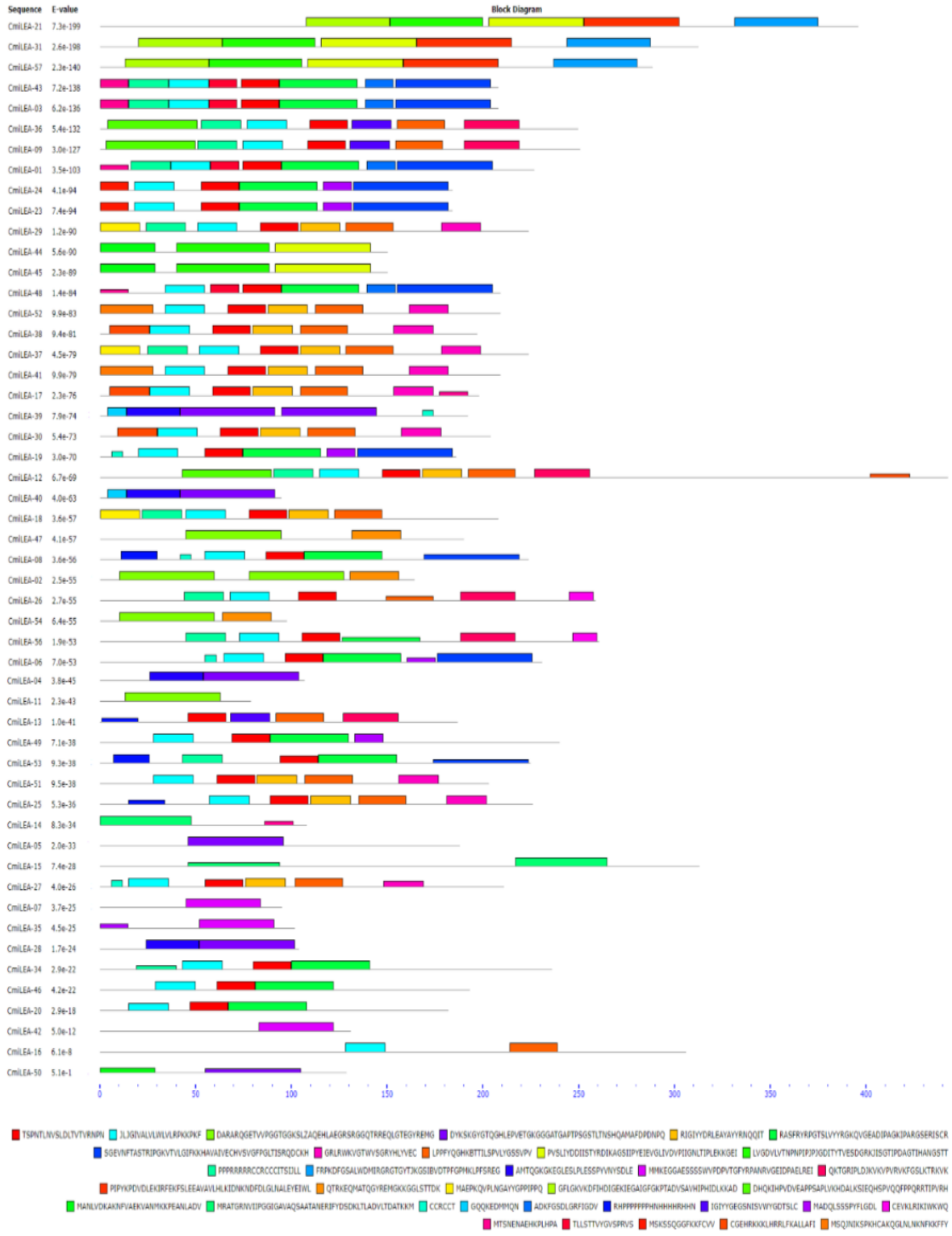

Figure 2. Conserved motifs of CmiLEA genes
Şekil 2. *CmiLEA genlerinde korunmuş proteinler*

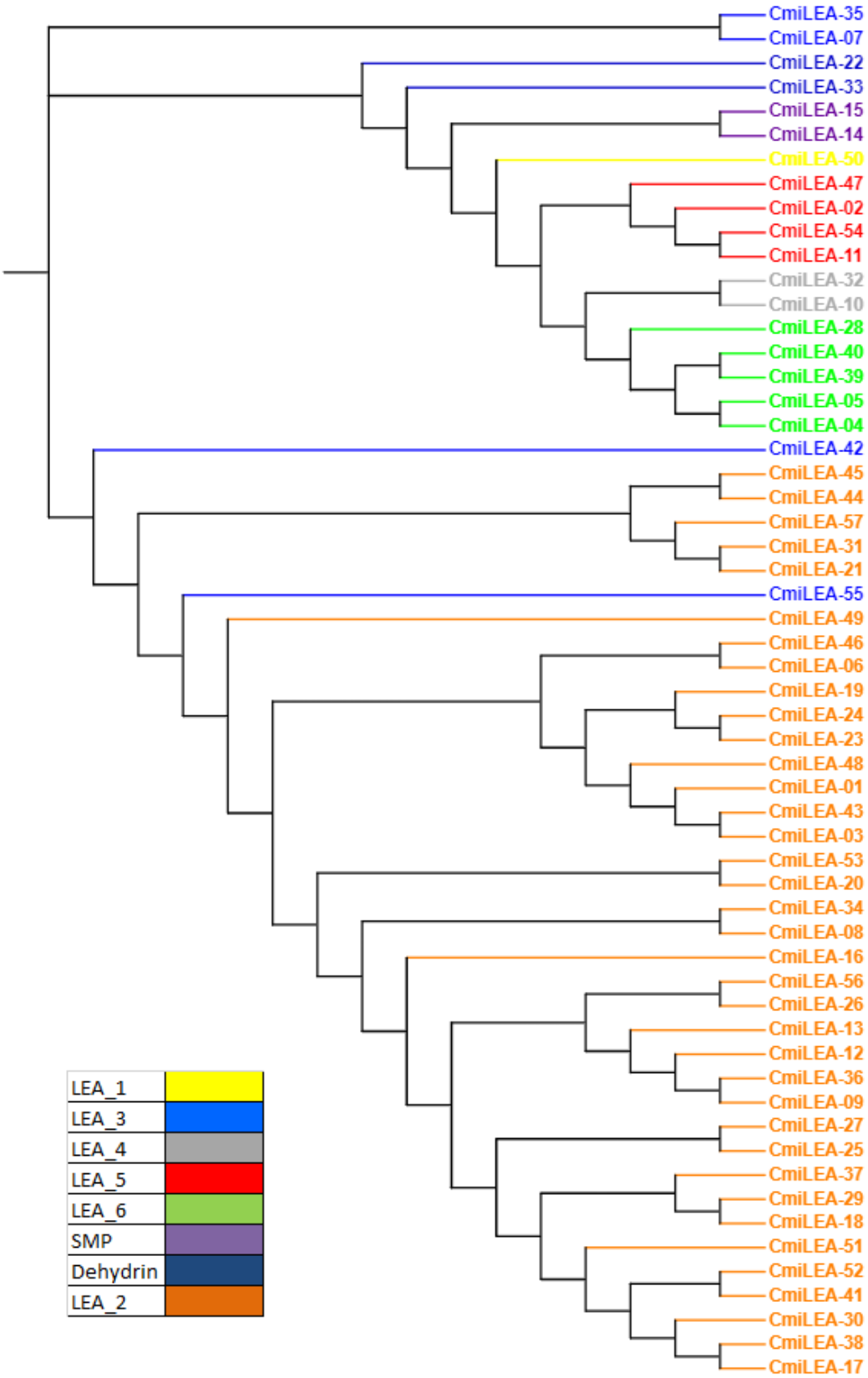

Figure 3. Phylogenetic tree of CmiLEA genes
Şekil 3. *CmiLEA genlerine ait filogenetik ağaç*

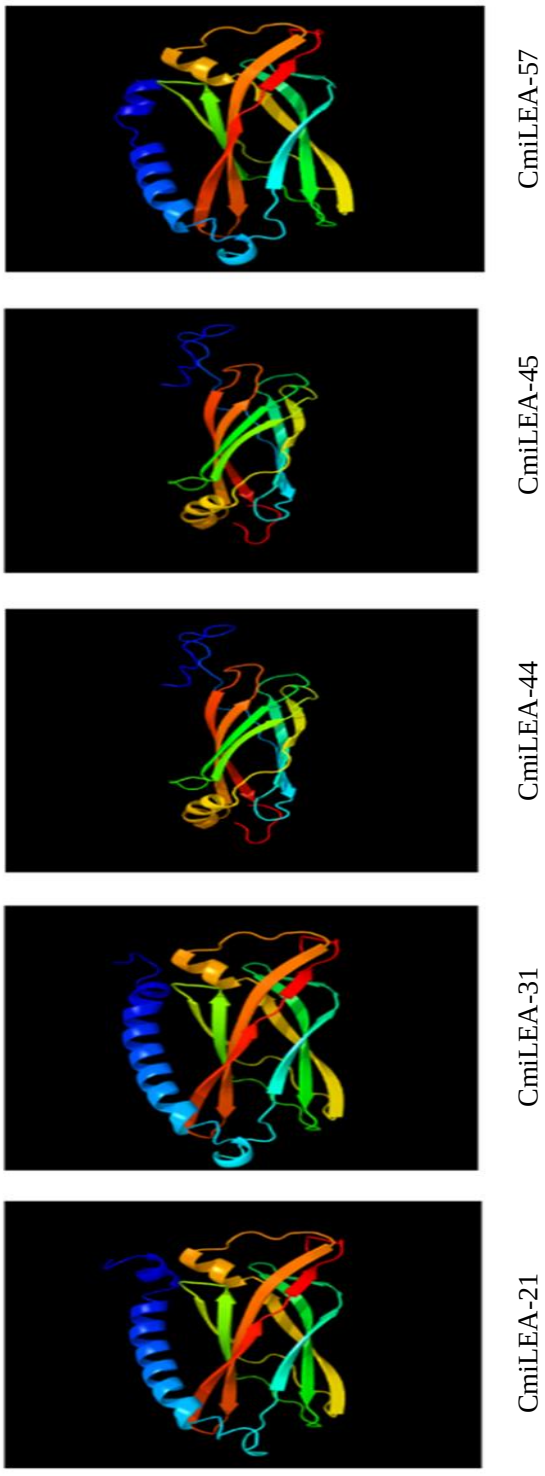

Figure 4. Three dimensional structure of some CmiLEA proteins
Şekil 4. *Bazı CmiLEA proteinlerinin 3 boyutlu yapısı*

Among these, the highest number of genes (10) were detected in the dehydrin subfamily. Lin et al. (2021) identified 84 LEA genes (CrLEA) in *Canavalia rosea*. They found 60 genes in LEA_2 and the fewest genes were found to be in LEA_4, LEA_5, LEA_6 subfamilies with 2 genes. Similarly, Li et al. (2023) reported 79 LEA genes were detected in sweetgum hybrids (*Liquidambar styraciflua* × *Liquidambar formosana*) and the most abundant in 8 subfamilies was the LEA_2 (57 genes). As a result of their study on tobacco, Geng et al. (2022) observed 123 NtLEA genes in *Nicotiana tabacum* L.and reported that the LEA_2 was the most abundant group. In consistent with previous studies, the present study identified the most abundant genes as the members of LEA_2 subfamily

In general, the results showed LEA protein lengths are relatively short, they are basic in terms of isoelectric points, and their molecular weights are low. In our study, it was revealed that 49 LEA proteins weighed less than 30 kDa, 38 of them were basic in character and their lengths were 79-455 aa long. When the average hydropathy values of LEA proteins were examined, it was found that the values were generally below 0 and the proteins mostly had hydrophilic character. These results are also similar with the findings of linseed flax, tobacco, lotus, and *Canavalia rosae* (Li et al., 2021; Lin et al., 2021; Geng et al., 2022; Chen et al., 2023). According to the hydropathy (GRAVY) values, we observed 18 proteins from the LEA_2 subfamily were hydrophilic and 19 proteins were hydrophobic, and all hydrophobic proteins were in LEA_2 subfamily.

By gene structure analysis, it was determined that there are very few introns (≤ 2) or no introns in LEA genes in lotus (Chen et al., 2023). It has been reported that the majority of genes have either no introns or 1 intron in *Canavelia rosae*. Li et al. (2021) stated that there are genes without introns in the LuLEA_2, LuLEA_3 and LuLEA_4 gene families in linseed flax, and the majority of genes have 1 or more introns. Li et al. (2023) were found there were very few introns in the LsfLEA genes in hybrid sweetgum. Our findings are compatible with other studies, and 12 genes in LEA_2 subfamily were determined to be intronless.

Among the target genes of related miRNAs of LEA_2 genes were detected as 2-phosphoglycerate kinase, GRAS, ERF, C2H2 transcription factors, phosphofructokinase family protein, squamosa promoter binding protein-like (SPL), and their possible functions were included nutrient deficiency response, carbohydrate metabolism, and drought response. It was found that there were in leaf and root development processes. For example, it was found that ath-miR447c-5p, associated with CmiLEA-09, was induced under C deficiency conditions but was suppressed under N deficiency conditions (Breakfield et al., 2012; Vidal et al., 2013; Shao et al., 2013; Liang et al., 2015; Thatcher et al., 2015; Rakhmetullina et al., 2021).

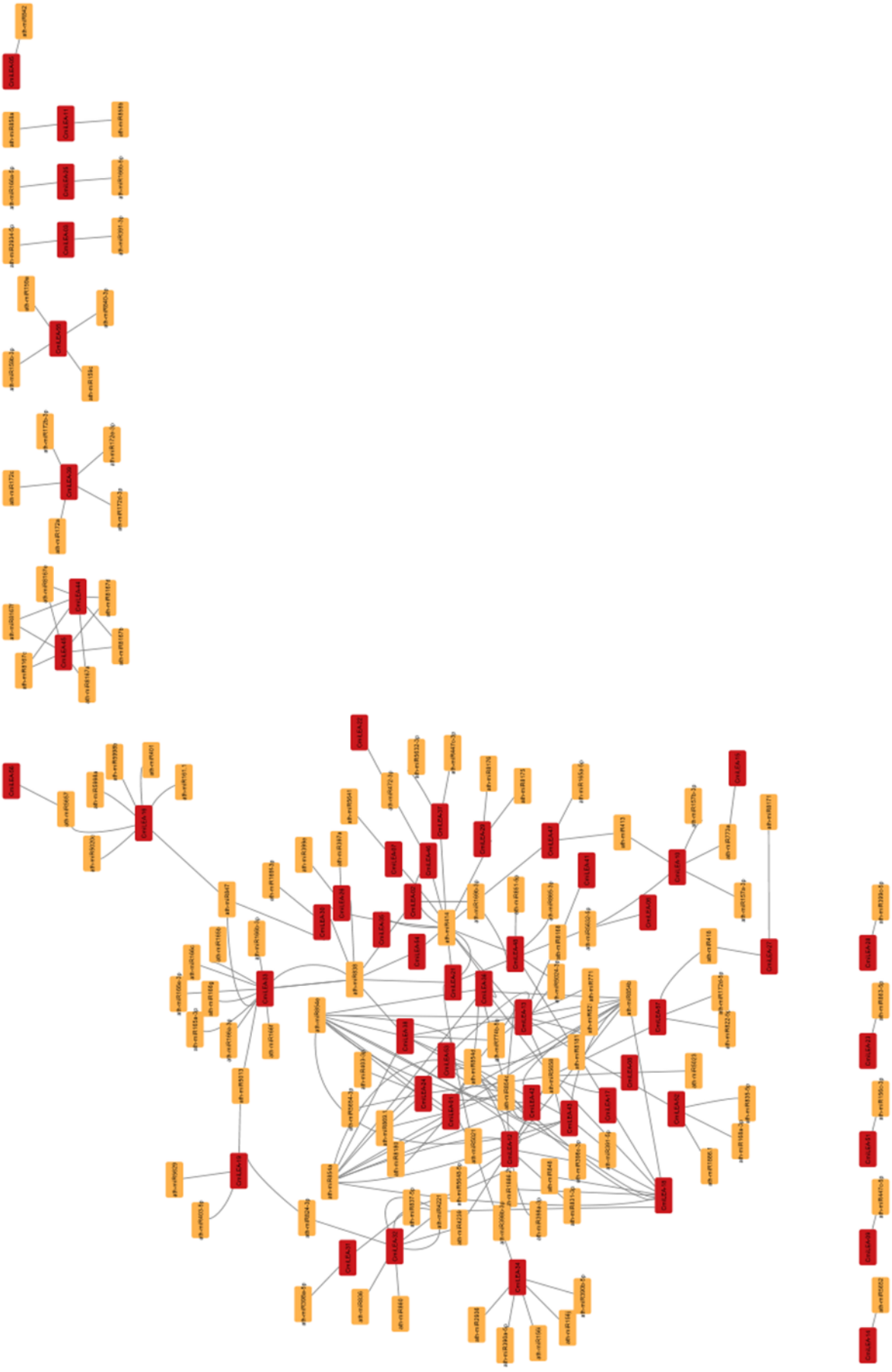

Figure 5. CmiLEA-miRNA interactions
Şekil 5. *CmiLEA-miRNA etkileşimleri*

The target genes of ath-miR842 and ath-miR399c-5p miRNAs, which are associated with CmiLEA-05 and CmiLEA-28 from LEA_6 subfamily, were found as Jacalin lectin family protein, copper superoxide dismutases, Ubiquitin conjugating enzyme (UCE) and vesicle-associated membrane protein and its possible functions can be nitrogen deficiency response and oxidative stress response (Jones-Rhoades et al., 2004; Sunkar and Zhu, 2004; Liang et al., 2012; Liang et al., 2015).

The lack of introns or the low number of introns in genes suggested the expression of LEA proeins are rapid under abiotic stress conditions. Thus, it shows that members of the LEA_2 subfamily are rapidly expressed in response to abiotic stress and may have important roles in the abiotic stress response. Previous studies also viewed that LEA_2 proteins have an ability to perform as molecular chaperones and they are involved in different stress responses such as ROS scavenging, membrane protection or preserving molecule structures under abiotic stress conditions (Aziz et al., 2023). Considering the possible target genes and functions of the CmiLEA proteins and miRNAs we found in present study, it can be suggested that some CmiLEA members from the LEA_2 subfamily are related to the stress response caused by nutrient deficiency.

Conclusion

In conclusion, the present study revealed LEA genes of Stout camphor tree (*Cinnamomum micranthum* f *Kanerihae*). for the first time. Total of 57 genes were identified and classified in subfamilies. Gene structure and miRNA analysis suggested the important functions of LEA genes in abiotic stress response. These findings suggest that it would be useful to analyze stress-sensitive expression patterns of these genes in *Cinnamomum* genus, which shows differences in tolerance to abiotic stress factors. The future studies can be revealed the relationship between gene induction and stress tolerance. Manipulating the survival of these genes in *Cinnamomum* may also lead to increased stress tolerance.

Referances

Aziz, M. A., Sabeem, M., Kutty, M. S., Rahman, S., Alneyadi, M. K., Alkaabi, A. B., & Masmoudi, K. (2023). Enzyme stabilization and thermotolerance function of the intrinsically disordered LEA2 proteins from date palm. *Scientific reports*, *13*(1), 11878.

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME suite. *Nucleic acids research*, *43*(W1), W39-W49.

Bies-Etheve, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., Cooke, & Delseny, M. (2008). Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant molecular biology*, *67*, 107-124.

Breakfield, N. W., Corcoran, D. L., Petricka, J. J., Shen, J., Sae-Seaw, J., Rubio-Somoza, I., R., Weigel, D., Ohler, U.& Benfey, P. N. (2012). High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in *Arabidopsis*. *Genome research*, *22*(1), 163-176.

Celik Altunoglu, Y., Baloglu, 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*, 5-21.

Chang, C. J., Lu, C. C., Lin, C. S., Martel, J., Ko, Y. F., Ojcius, D. M. & Young, J. D. (2018). *Antrodia cinnamomea* reduces obesity and modulates the gut microbiota in high-fat diet-fed mice. *International journal of obesity*, *42*(2), 231-243.

Chen, L., Xin, J., Song, H., Xu, F., Yang, H., Sun, H., & Yang, M. (2023). 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*, 1-13.

Cheng, Z., Zhang, X., Yao, W., Zhao, K., Liu, L., Fan, G., ... & Jiang, T. (2021). Genome-wide search and structural and functional analyses for late embryogenesis-abundant (LEA) gene family in poplar. *BMC Plant Biology*, *21*, 1-17.

Chung, K. F., & Hsieh, C. L. (2023). Synopsis of Camphora (*Cinnamomeae*, *Lauraceae*) of Taiwan, with two new combinations and one new synonym. *Taiwania*, *68*(3).

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674-3676.

Dai, X., Zhuang, Z., & Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic acids research*, *46*(W1), W49-W54.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *Protein identification and analysis tools on the ExPASy server* (pp. 571-607). Humana press.

Geng, W., Wang, Y., Zhang, J., Liu, Z., Chen, X., Qin, L., Yang, L. & Tang, H. (2022). 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*, 146665.

Hong-Bo, S., Zong-Suo, L., & Ming-An, S. (2005). LEA proteins in higher plants: structure, function, gene expression and regulation. *Colloids and surfaces B: Biointerfaces*, *45*(3-4), 131-135.

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., & Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, *31*(8), 1296-1297.

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*, 1-22.

Jiang, R., Chen, X., Liao, X., Peng, D., Han, X., Zhu, C., ... & Li, C. (2022). A chromosome-level genome of the camphor tree and the underlying genetic and climatic factors for its top-geoherbalism. *Frontiers in Plant Science*, *13*, 827890.

Jones-Rhoades, M. W., & Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular cell*, *14*(6), 787-799.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg,

M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, *10*(6), 845-858.

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic acids research*, *47*(D1), D155-D162.

Li, D., Lin, H. Y., Wang, X., Bi, B., Gao, Y., Shao, L. & Zhang, L. (2023). Genome and whole-genome resequencing of *Cinnamomum camphora* elucidate its dominance in subtropical urban landscapes. *BMC biology*, *21*(1), 192.

Li, Y., Qi, S., Chen, S., Li, H., Zhang, T., Bao, F. & Zhao, J. (2023). Genome-wide identification and expression analysis of late embryogenesis abundant (LEA) genes reveal their potential roles in somatic embryogenesis in hybrid sweetgum (*Liquidambar styraciflua*× *Liquidambar formosana*). *Forestry Research*, *3*(1).

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*, 1-13.

Liang, G., Ai, Q., & Yu, D. (2015). Uncovering miRNAs involved in crosstalk between nutrient deficiencies in *Arabidopsis*. *Scientific reports*, *5*(1), 11813.

Liang, G., He, H., & Yu, D. (2012). Identification of nitrogen starvation-responsive microRNAs in *Arabidopsis thaliana. PloS one*, *7*(11), e48951.

Lin, R., Zou, T., Mei, Q., Wang, Z., Zhang, M., & Jian, S. (2021). Genome-wide analysis of the late embryogenesis abundant (LEA) and abscisic acid-, stress-, and ripening-induced (ASR) gene superfamily from Canavalia rosea and their roles in salinity/alkaline and drought tolerance. *International Journal of Molecular Sciences*, *22*(9), 4554.

Lin, R., Zou, T., Mei, Q., Wang, Z., Zhang, M., & Jian, S. (2021). Genome-wide analysis of the late embryogenesis abundant (LEA) and abscisic acid-, stress-, and ripening-induced (ASR) gene superfamily from *Canavalia rosea* and their roles in salinity/alkaline and drought tolerance. *International Journal of Molecular Sciences*, *22*(9), 4554.

NCBI. (2024). https://www.ncbi.nlm.nih.gov/.

Qiagen, 2022. CLC Genomic Workbench 21. https://digitalinsights.qiagen.com/.

Rakhmetullina, A., Zielenkiewicz, P., Pyrkova, A., Uteulin, K., & Ivashchenko, A. (2021). Prediction of characteristics of interactions of miRNA with mRNA of GRAS, ERF, C2H2 genes of A. thaliana, O. sativa and Z. mays. *Current Plant Biology*, *28*, 100224.

Shao, C., Wu, Q., Qiu, J., Jin, S., Zhang, B., Qian, J., Chen, M. & Meng, Y. (2013). Identification of novel microRNA-like-coding sites on the long-stem microRNA precursors in *Arabidopsis*. *Gene*, *527*(2), 477-483.

Shen, T., Qi, H., Luan, X., Xu, W., Yu, F., Zhong, Y., & Xu, M. (2022). The chromosome-level genome sequence of the camphor tree provides insights into *Lauraceae* evolution and terpene biosynthesis. *Plant Biotechnology Journal*, *20*(2), 244.

Sun, W. H., Xiang, S., Zhang, Q. G., Xiao, L., Zhang, D. Y., Zhang, P. L. & Zou, S. Q. (2022). The camphor tree genome enhances the understanding of magnoliid evolution.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, *38* (7), 3022-3027.

Thatcher, S. R., Burd, S., Wright, C., Lers, A., & Green, P. J. (2015). Differential expression of miRNAs and their target genes in senescing leaves and siliques: insights from deep sequencing of small RNAs and cleaved target RNAs. *Plant, cell & environment*, *38*(1), 188-200.

Vidal, E. A., Moyano, T. C., Krouk, G., Katari, M. S., Tanurdzic, M., McCombie, W. R. & Gutiérrez, R. A. (2013). Integrated RNA-seq and sRNA-seq analysis identifies novel nitrate-responsive genes in *Arabidopsis thaliana* roots. *BMC genomics*, *14*, 1-15.

Wang, X. D., Xu, C. Y., Zheng, Y. J., Wu, Y. F., Zhang, Y. T., Zhang, T. & Jiang, X. M. (2022). Chromosome-level genome assembly and resequencing of camphor tree (*Cinnamomum camphora*) provides insight into phylogeny and diversification of terpenoid and triglyceride biosynthesis of *Cinnamomum*. *Horticulture Research*, *9*, uhac216.