

宋梦瑶, 李娜, 秦家范, 等. 谷子 *LEA* 基因家族全基因组鉴定及表达分析[J]. 山西农业科学, 2025, 53(3): 16-26.

SONG M Y, LI N, QIN J F, et al. Genome-wide identification and expression analysis of *LEA* gene family in foxtail millet[J]. Journal of Shanxi Agricultural Sciences, 2025, 53(3): 16-26.

doi:10.3969/j.issn.1002-2481.2025.03.03

## 谷子 *LEA* 基因家族全基因组鉴定及表达分析

宋梦瑶<sup>1</sup>, 李娜<sup>1</sup>, 秦家范<sup>3</sup>, 郭冰<sup>1</sup>, 李春霞<sup>1</sup>, 李君霞<sup>2</sup>, 马小倩<sup>1</sup>

(1. 河南科技大学农学院, 河南 洛阳 471023; 2. 河南省农业科学院粮食作物研究所, 河南 郑州 450002;  
3. 洛阳市农林科学院甘薯谷子研究所, 河南 洛阳 471023)

**摘要:** 为探究谷子 *LEA* 基因家族的特征和功能, 基于已发表的拟南芥 *LEA* 蛋白序列, 通过生物信息学方法对谷子 *LEA* 家族成员进行全基因组鉴定, 分析 *SiLEAs* 编码蛋白的理化性质、保守基序、系统发育树、顺式调控元件、共线性和组织表达模式等。结果表明, 谷子 *LEA* 家族成员中共鉴定出 33 个 *SiLEAs* 成员, *SiLEAs* 不均匀地分布在 9 条染色体上; 亚细胞预测结果发现, *SiLEAs* 在细胞核、细胞质、叶绿体和线粒体中均有存在。通过系统发育树分析发现, 不同亚家族均分为 3 组, 同一亚家族的 *SiLEAs* 成员大多表现出相似的基因结构和保守基序成分特征; 顺式调控元件分析表明, 在 *SiLEAs* 的启动子区存在许多与植物激素和逆境胁迫相关的反应元件; 物种间共线性分析发现, 谷子与单子叶植物共线基因对远大于双子叶植物的共线基因对; 通过对 *SiLEAs* 基因表达模式分析发现, *SiLEA3-6*、*SiLEA3-7*、*SiDHN-4*、*SiDHN-1* 和 *SiLEA3-8* 在根、茎、叶、开花前和开花后 7 d 中高表达, 并且这 5 个基因也均含有与生长发育相关的元件, 暗示 *SiLEA3-6*、*SiLEA3-7*、*SiDHN-4*、*SiDHN-1* 和 *SiLEA3-8* 可能是谷子 *LEAs* 基因家族响应生长发育的关键基因。

**关键词:** 谷子; *LEA* 基因家族; 生物信息学; 系统发育树; 顺式调控元件; 共线性

中图分类号: S515 文献标识码: A 文章编号: 1002-2481(2025)03-0016-11

## Genome-Wide Identification and Expression Analysis of *LEA* Gene Family in Foxtail Millet

SONG Mengyao<sup>1</sup>, LI Na<sup>1</sup>, QIN Jiafan<sup>3</sup>, GUO Bing<sup>1</sup>, LI Chunxia<sup>1</sup>, LI Junxia<sup>2</sup>, MA Xiaoqian<sup>1</sup>

(1. College of Agriculture, Henan University of Science and Technology, Luoyang 471023, China; 2. Institute of Grain Crops, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China; 3. Institute of Sweet Potato and Millet, Luoyang Academy of Agricultural and Forestry Sciences, Luoyang 471023, China)

**Abstract:** In order to explore the characteristics and function of the foxtail millet *LEA* gene family, based on the published *Arabidopsis thaliana* *LEA* protein sequence, *LEA* family members were genome-widely identified using bioinformatics methods. Subsequent analysis were conducted on the physical and chemical properties, conserved motifs, phylogenetic tree, cis-regulatory elements, collinearity, and expression patterns of the proteins encoded by *SiLEAs*. These results showed that a total of 33 *SiLEA* members were identified in the foxtail millet *LEA* family, and *SiLEAs* were unevenly distributed on 9 chromosomes. Subcellular prediction found that *SiLEAs* were located in the nucleus, cytoplasm, chloroplast, and mitochondria. Through phylogenetic tree analysis, different subfamilies were divided into three groups, and most of those *SiLEAs* gene in the same subfamily showed similar gene structure and conserved motif component characteristics. Cis-regulatory element analysis showed that there were many response elements related to plant hormones and stress in the promoter region of *SiLEAs*. Inter-species collinearity analysis found that the number of gene pairs, which were collinear with

收稿日期: 2024-07-01

基金项目: 河南省重点研发专项(231111110300); 国家自然科学基金(32301841); 河南省自然科学基金(242300421320); 河南科技大学 A 类博士人才项目(13480103)

作者简介: 宋梦瑶, 在读硕士, 研究方向: 谷子遗传育种, E-mail: m17852585966\_1@163.com

通信作者: 李君霞, 研究员, 硕士生导师, 主要从事杂粮作物遗传育种研究, E-mail: lijunxia@126.com

马小倩, 讲师, 博士, 主要从事谷子遗传育种研究, E-mail: maxq20210812@163.com

monocotyledon, were much higher than those with dicotyledon. Analysis of the gene expression patterns of *SiLEAs* revealed that *SiLEA3-6*, *SiLEA3-7*, *SiDHN-4*, *SiDHN-1*, and *SiLEA3-8* were highly expressed in roots, stems, leaves, and before flowering and seven days after flowering, and these five genes also all contained the elements related to growth and development, suggesting that *SiLEA3-6*, *SiLEA3-7*, *SiDHN-4*, *SiDHN-1*, and *SiLEA3-8* might be the key genes of *LEA* gene family in response to growth and development in foxtail millet.

**Keywords:** foxtail millet; *LEA* gene family; bioinformatics; phylogenetic tree; *cis*-regulatory elements; collinearity

胚胎发育晚期丰富蛋白(LEA)是种子发育后期大量积累的一类蛋白,广泛存在于植物中,当植物遇到低温、干旱、脱水等不利环境条件时,它们会迅速合成并积累,使其能够应对不利的环境,在各类非生物胁迫中发挥重要作用<sup>[1]</sup>。1981年,DURE等<sup>[2]</sup>在棉花中首次发现了LEA蛋白,并表明该蛋白在种子脱水和成熟过程中大量积累,保护种子免受伤害。LEA蛋白相继在多种植物中被发现,如拟南芥<sup>[3]</sup>、玉米<sup>[4]</sup>、小麦<sup>[5]</sup>、水稻<sup>[6]</sup>、番茄<sup>[7]</sup>、甘蓝型油菜<sup>[8]</sup>、葡萄<sup>[9]</sup>等。根据氨基酸序列和保守基序的相似性,LEA蛋白分为8个亚家族:LEA\_1、LEA\_2、LEA\_3、LEA\_4、LEA\_5、LEA\_6、脱氢蛋白(DHN)和种子成熟蛋白(SMP)<sup>[2]</sup>。

LEA蛋白具有高亲水性和结构无序性2个显著特征,高亲水性是由于LEA蛋白含有高度带电的氨基酸(如丙氨酸),富含甘氨酸、谷氨酸、赖氨酸和苏氨酸,但缺乏半胱氨酸和色氨酸残基。无序性是由于在自然条件下,LEA蛋白不能折叠为明确唯一的空间结构,但在一些干旱和极端温度下,LEA蛋白能具有稳定的三维结构。以上2种特征使LEA蛋白可以增加植物细胞的保水能力,在缺水期间也可以维持细胞功能<sup>[10-12]</sup>,从而提高植物细胞对多种非生物胁迫(干旱、高盐、渗透胁迫、低温、高温等)的耐受性。XIAO等<sup>[13]</sup>对水稻 *OsLEA3-1* 的抗旱性进行检测,发现过表达的 *OsLEA3-1* 基因能显著提高水稻的耐旱性。LIU等<sup>[14]</sup>在对玉米 *ZmLEA3* 进行低温和渗透胁迫以及脱落酸(ABA)和  $H_2O_2$  处理后,发现 *ZmLEA3* 蛋白在低温下可以保护乳酸脱氢酶(LDH)的活性。SHI等<sup>[15]</sup>通过对苜蓿 *MjLEA3* 的研究发现, *MjLEA3* 的组成性表达增强了转基因烟草对寒冷、干旱和强光胁迫的耐受性。刘浩<sup>[5]</sup>通过对苗期小麦进行PEG6000和低温胁迫处理的研究发现, *TaLEA\_3-3*、*TaLEA\_4-1*、*TaLEA\_6-2* 和 *TaDHN43* 能够显著提高大肠杆菌和酵母细胞对高盐和高温胁迫的耐受性。

谷子(*Setaria italica*)广泛种植于我国北方干旱和半干旱地区,籽粒富含蛋白质、维生素等营养

物质,有抗逆性、适应性强等特点,是重要的杂粮作物。谷子作为  $C_4$  植物,其  $CO_2$  同化率、光合效率、生物量、产量和光合利用效率均优于  $C_3$  植物,此外,谷子还具有很强的非生物抗逆性,所以,谷子常被用作胁迫生物学研究的重要模式种<sup>[16-18]</sup>。谷子基因组测序的完成,有助于推动谷子在功能基因组学方面的研究<sup>[19]</sup>。然而,目前谷子 *LEA* 基因家族全基因组的鉴定和表达分析还未见报道。

为挖掘谷子 *LEA* 基因的特征与功能,本研究对谷子的 *LEA* 基因家族进行成员鉴定,并对 *SiLEAs* 基因和其编码蛋白进行生物信息学分析,包括预测亚细胞定位、基因结构、顺式调控元件、共线性、基因表达模式等,旨在为谷子生长发育和逆境响应分析提供一定的理论依据。

## 1 材料和方法

### 1.1 谷子 *LEAs* 基因家族成员鉴定和蛋白理化性质分析

从谷子综合数据库 *Setaria-DB*<sup>[20]</sup>(<http://www.setariadb.com/millet>)获取豫谷1号(YG1)基因组序列。从TAIR(<https://www.arabidopsis.org/>)数据库中获取51条拟南芥 *LEAs* 家族蛋白序列<sup>[3]</sup>,通过Blast方法对谷子 *LEA* 基因家族成员进行鉴定。同时,从Pfam数据库(<http://pfam.xfam.org/>)中下载LEA蛋白的隐马尔可夫模型(Hidden Markov Model,HMM):DHN(PF00257)、LEA\_1(PF03760)、LEA\_2(PF03168)、LEA\_3(PF03242)、LEA\_4(PF02987)、LEA\_5(PF00477)、LEA\_6(PF10714)和SMP(PF04927),并利用TBtools软件<sup>[21]</sup>(版本:1.118)中的Two Sequence Files和Simple HMM Search功能,在谷子基因组序列中进行筛选。将2种方法得到的候选基因在Pfam(<http://pfam-legacy.xfam.org/>)、SMART(<http://smart.embl-heidelberg.de/#>)和NCBI CDD database(<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)数据库进行结构域确认。通过ProtParam工具(<http://web.expasy.org/protparam/>)分析 *SiLEAs* 蛋白的生理生

化参数,使用 WoLF PSORT 工具(<https://wolf-psort.hgc.jp/>)预测 SiLEAs 蛋白的亚细胞定位。

## 1.2 谷子 LEAs 基因的系统发育树、基因结构及保守基序分析

使用 MEGA11<sup>[22]</sup> 中 Neighbor-Joining (NJ) 方法生成系统进化树, Bootstrap 重复值设为 1 000, 使用 iTOL(<https://itol.embl.de/>) 在线工具对生成的系统进化树进行美化。使用 TBtools 软件对谷子 LEAs 基因的保守基序和基因结构进行鉴定, 搜索 Motif 的数量为 20 个。最后使用 TBtools 软件对保守基序和基因结构可视化分析。

## 1.3 谷子 LEAs 基因的顺式调控元件分析

利用 TBtools 软件获取 SiLEAs 基因起始密码子上游 2 000 bp 的序列, 然后将序列提交至 Plant CARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)。软件预测 SiLEAs 基因上游 2 000 bp 的顺式调控元件, 最后使用 TBtools 软件对结果进行可视化处理。

## 1.4 谷子 LEAs 基因的染色体分布和共线性分析

利用 TBtools 软件进行谷子 LEAs 基因染色体位置分析, 基因密度设置为 1 000 000, 并使用 TBtools 软件可视化。利用 TBtools 软件获取染色体长度、基因密度和共线性文件, 共线性文件采用默认值, 使用 TBtools 软件中 Advanced Circos 功能实现谷子 LEAs 基因家族物种内共线性可视化。通过

Ensembl(<https://plants.ensembl.org/index.html>) 下载大豆、水稻、玉米的 DNA 以及 Gff 注释文件, TAIR 数据库中获取 51 条拟南芥 LEAs 家族的蛋白序列<sup>[3]</sup>。利用 TBtools 软件获得谷子与拟南芥、大豆、水稻、玉米的共线性文件, 得到谷子与水稻、玉米、大豆、拟南芥的共线基因对。

## 1.5 谷子 LEAs 基因的编码蛋白互作和基因表达模式分析

利用 STRING(<https://string-db.org/>) 的 Proteins by sequences 对谷子 LEAs 基因家族成员进行蛋白互作分析。根据下载的豫谷 1 号转录组数据<sup>[23]</sup>, 选取萌发种子、灌浆种子、成熟种子及苗期、抽穗期、孕穗期的根、茎、叶及开花期的转录组数据, 对谷子 LEAs 基因表达模式进行分析, 并使用 TBtools 软件可视化。

## 2 结果与分析

### 2.1 谷子中 LEAs 基因家族成员鉴定和蛋白理化性质分析

通过 Blast 和 HMM 这 2 种方法, 从谷子基因组中共鉴定出 33 个 LEAs 蛋白。根据蛋白的保守结构域和 Pfam 命名法, 将 SiLEAs 蛋白划分为 7 个亚家族, 分别为 LEA\_1、LEA\_2、LEA\_3、LEA\_5、LEA\_6、SMP 和 DHN, 其中, LEA\_3 亚家族最多, 共有 8 个成员(表 1)。

表 1 谷子 LEAs 基因家族成员编码蛋白的理化性质分析

Tab.1 Analysis of physical and chemical properties of amino acids of LEA genes family members in foxtail millet

基因 Gene	基因编号 Gene ID	基因亚家族 Gene subfamily	氨基酸长度 Amino acid length	分子量/ku Molecular weight	等电点 pI	总平均亲水系数 GRAVY	亚细胞定位 Subcellular localization
SiSMP-1	Seita.1G212500.1	SMP	167	16.156 72	5.15	-0.139	细胞质
SiSMP-2	Seita.1G212800.1	SMP	143	14.271 86	5.89	-0.176	细胞质
SiDHN-1	Seita.1G267200.1	DHN	290	31.644 05	5.76	-1.246	细胞核
SiLEA2-1	Seita.2G102400.1	LEA_2	331	36.741 76	4.97	-0.334	细胞质
SiLEA1-1	Seita.2G159600.1	LEA_1	168	16.539 32	9.39	-0.489	细胞质
SiLEA6-1	Seita.2G249300.1	LEA_6	117	12.315 53	6.41	-0.925	细胞核
SiLEA2-2	Seita.3G130700.1	LEA_2	170	18.768 62	5.56	-0.155	细胞核
SiLEA3-1	Seita.3G203500.1	LEA_3	78	7.689 58	9.51	-0.062	叶绿体
SiLEA3-2	Seita.3G272300.1	LEA_3	97	10.393 88	5.45	-0.218	叶绿体
SiLEA5-1	Seita.3G277600.1	LEA_5	153	16.419 67	5.74	-1.380	细胞核
SiLEA5-2	Seita.3G277700.1	LEA_5	93	9.799 50	5.48	-1.280	细胞核
SiLEA1-2	Seita.4G007900.1	LEA_1	125	13.494 34	9.84	-0.889	线粒体
SiLEA1-3	Seita.4G140300.1	LEA_1	78	8.543 56	5.91	-1.190	细胞核
SiSMP-3	Seita.4G159500.1	SMP	288	28.979 78	4.63	-0.307	细胞质
SiLEA2-3	Seita.5G014300.1	LEA_2	151	16.302 82	6.11	-0.028	细胞质
SiLEA3-3	Seita.5G021300.1	LEA_3	94	9.762 18	9.77	-0.279	叶绿体

续表1 谷子LEAs基因家族成员氨基酸的理化性质分析

Tab.1(Continued) Analysis of physical and chemical properties of amino acids of LEA genes family members in foxtail millet

基因 Gene	基因编号 Gene ID	基因亚家族 Gene subfamily	氨基酸长度 Amino acid length	分子质量/ku Molecular weight	等电点 pI	总平均亲水系数 GRAVY	亚细胞定位 Subcellular localization
<i>SiLEA3-4</i>	Seita.5G021400.1	LEA_3	96	10.204 83	10.08	-0.132	叶绿体
<i>SiLEA3-5</i>	Seita.5G057300.1	LEA_3	100	9.935 12	10.13	0	叶绿体
<i>SiLEA5-3</i>	Seita.5G113500.1	LEA_5	94	10.075 05	5.34	-1.135	线粒体
<i>SiDHN-2</i>	Seita.5G166200.1	DHN	138	14.126 34	9.19	-1.178	细胞核
<i>SiDHN-3</i>	Seita.5G166300.1	DHN	138	14.101 29	8.90	-1.192	细胞核
<i>SiLEA2-4</i>	Seita.5G232600.1	LEA_2	175	18.999 70	5.01	-0.086	细胞质
<i>SiLEA1-4</i>	Seita.6G114000.1	LEA_1	153	15.3427 2	9.52	-0.722	细胞核
<i>SiSMP-4</i>	Seita.6G123500.1	SMP	181	17.602 27	4.84	-0.317	细胞质
<i>SiLEA3-6</i>	Seita.6G169100.1	LEA_3	96	10.2768 2	9.21	-0.341	叶绿体
<i>SiLEA3-7</i>	Seita.7G182400.1	LEA_3	100	11.016 61	10.05	-0.126	叶绿体
<i>SiLEA1-5</i>	Seita.7G217700.1	LEA_1	116	11.981 34	9.33	-0.928	线粒体
<i>SiDHN-4</i>	Seita.8G115100.1	DHN	192	19.613 58	9.68	-0.985	叶绿体
<i>SiDHN-5</i>	Seita.8G115200.1	DHN	347	33.741 27	8.99	-0.804	细胞核
<i>SiDHN-6</i>	Seita.8G115400.1	DHN	169	16.914 33	8.81	-1.050	细胞核
<i>SiSMP-5</i>	Seita.9G093900.1	SMP	266	26.580 57	4.98	-0.221	细胞质
<i>SiLEA3-8</i>	Seita.9G372100.1	LEA_3	133	14.625 47	9.62	-0.418	叶绿体
<i>SiSMP-6</i>	Seita.9G533500.1	SMP	280	27.752 43	4.65	-0.195	细胞质

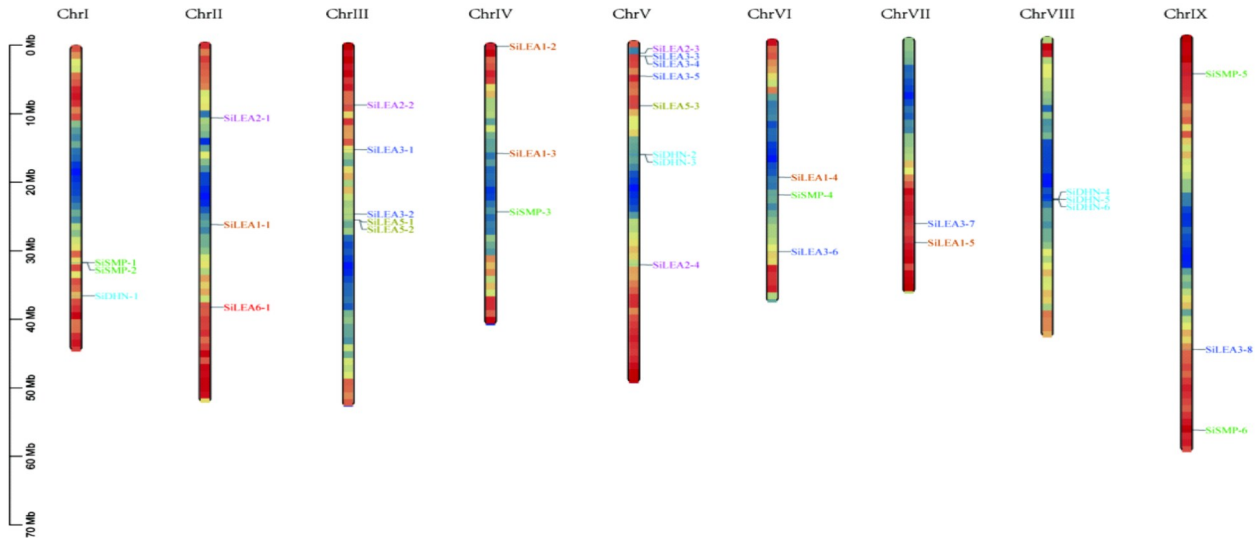
由表1可知,DHN、SMP和LEA\_1亚家族成员分别有6、6、5个。LEA\_2亚家族成员有4个,LEA\_5亚家族成员有3个,LEA\_6亚家族成员只有1个。在谷子中未检测到LEA\_4亚家族。预测的33个SiLEAs蛋白编码氨基酸数量在78~347个,预测分子质量范围为7.69~36.74 ku,等电点(pI)范围为4.63~10.13,平均值为7.39。17个SiLEAs蛋白的pI值小于7,表明它们是酸性蛋白,而其余的SiLEAs蛋白的pI值均大于7,表明它们是碱性蛋白。SiLEAs蛋白的总平均亲水系数(GRAVY)范围为-1.38~0。在SiLEAs蛋白中97%的SiLEAs蛋白的总平均亲水系数小于0,表明大多数SiLEAs蛋白具有亲水性。亚细胞定位预测显示,亚家族SMP的全部成员、LEA\_2的3个成员(SiLEA2-1、SiLEA2-3、SiLEA2-4)、LEA\_1的1个成员(SiLEA1-1)定位在细胞质。位置在细胞核内的有LEA\_6亚家族的全部成员、DHN亚家族的5个成员(SiDHN-1、SiDHN-2、SiDHN-3、SiDHN-5、SiDHN-6)、LEA\_1亚家族的2个成员(SiLEA1-3、SiLEA1-4)、LEA\_2亚家族的1个成员(SiLEA2-2)和LEA\_5亚家族的2个成员(SiLEA5-1、SiLEA5-2)。LEA\_3的全部成员位于叶绿体内,DHN亚家族的SiDHN-4位于叶绿体内。在线粒体中有LEA\_1亚家族的2位成员

(SiLEA1-2、SiLEA1-5)和LEA\_5亚家族的SiLEA5-3。

对谷子LEAs基因家族成员进行染色体定位分析发现(图1),33个成员在9条染色体上均有分布且呈不均匀分布状态,SiLEAs基因大部分位于5号染色体上,包含有8个SiLEAs基因,3号染色体含有5个SiLEAs基因,7号染色体仅含有2个SiLEAs基因。1号、2号、4号、6号、8号、9号染色体各含有3个SiLEAs基因。

## 2.2 谷子LEAs基因的系统发育树分析

为了进一步研究SiLEAs蛋白的系统发育关系,本研究利用33个SiLEAs、51个AtLEAs<sup>[3]</sup>、52个ZmLEAs<sup>[4]</sup>、179个TaLEAs<sup>[5]</sup>和35个SbLEAs<sup>[24]</sup>的蛋白序列构建LEA蛋白不同亚家族系统发育树(图2)。不同亚家族均分为3组,根据其亚家族不同分别定义为I群、II群和III群。LEA\_1亚家族中LEA\_1-I包含2个成员,LEA\_1-III包含3个成员;LEA\_2亚家族的全部成员分布在LEA\_2-III组;LEA\_3亚家族中LEA\_3-I包含6个成员,LEA\_3-III包含2个成员;LEA\_5亚家族的3个成员分别分布在3个组中;LEA\_6亚家族中SiLEA6-1分布在LEA\_6-III;SMP亚家族中SMP-I包含4个成员,SMP-III中包含2个成员;DHN亚家族中DHN-I包含5个成员,DHN-III中包含1个成员。



染色体内部以基因密度填充,颜色由蓝到红代表基因密度由小到大

The interior of the chromosome was filled with gene density, and the color from blue to red indicated that the gene density went from small to large

图 1 谷子 *LEAs* 基因的染色体定位

Fig.1 Chromosome mapping of *LEAs* gene in foxtail millet

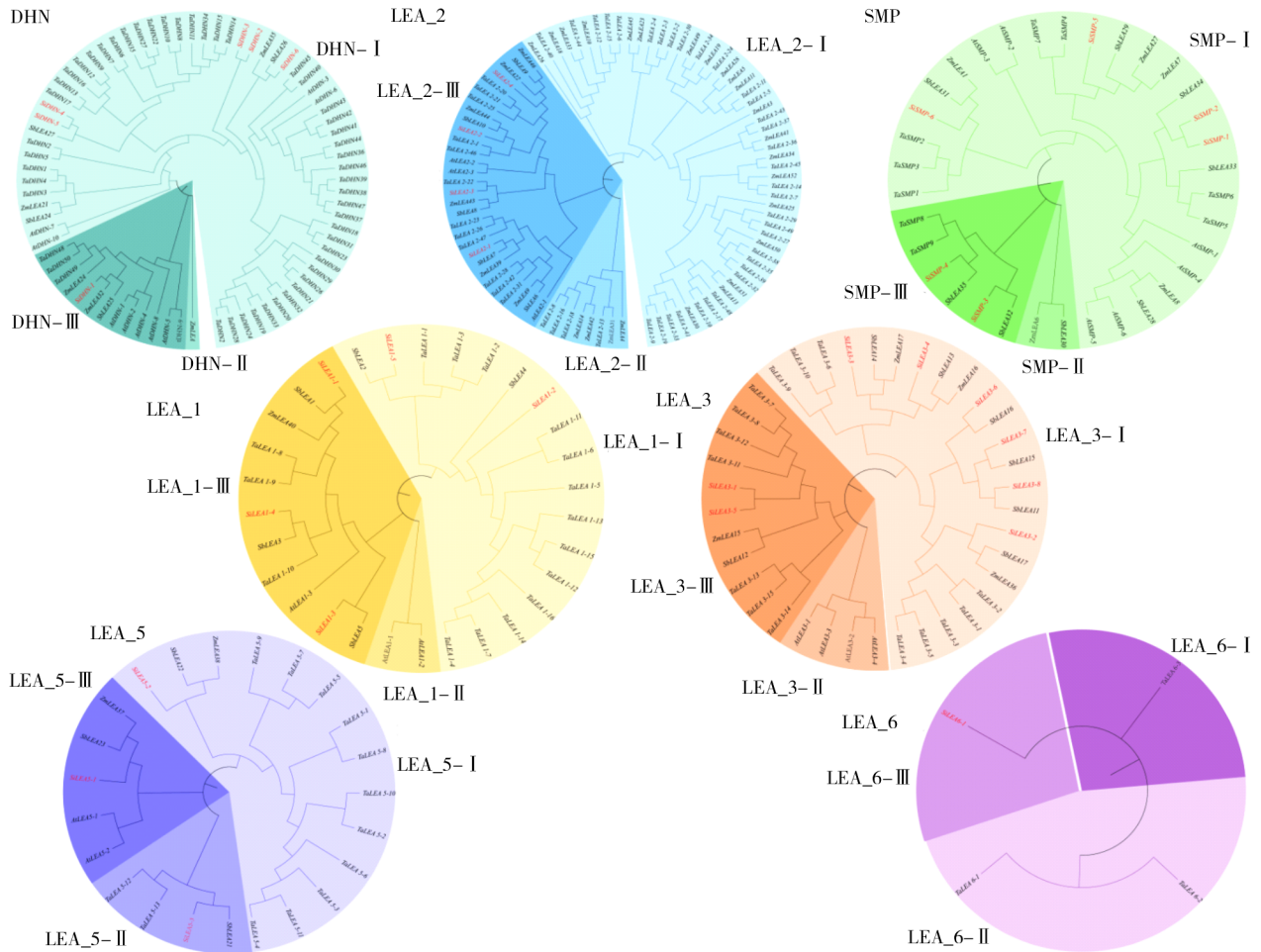


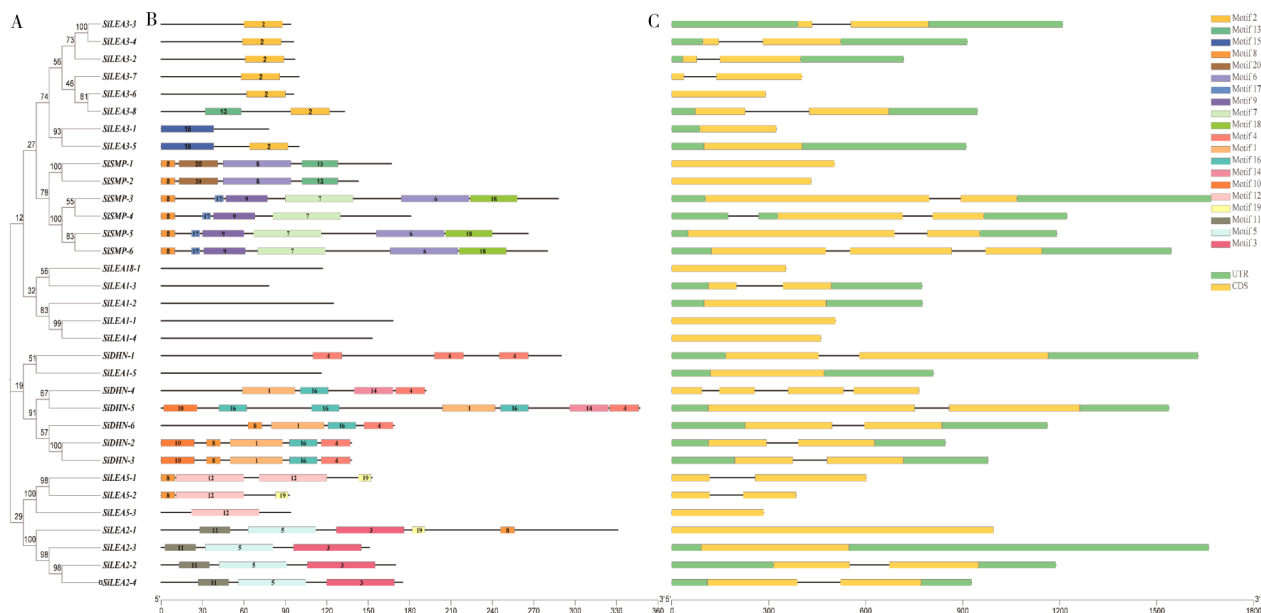
图 2 谷子、拟南芥、玉米、小麦和高粱 *LEAs* 基因亚家族的系统发育树

Fig.2 Phylogenetic trees of *LEAs* gene subfamilies in foxtail millet, *Arabidopsis*, maize, wheat, and sorghum

### 2.3 谷子LEAs基因的保守基序、保守结构域和基因结构分析

分析保守基序、保守结构域和外显子-内含子结构的分布,探索*SiLEAs*基因的结构多样性,共鉴定出20个保守基序(Motif 1~Motif 20)(图3-A、B)。同一亚家族含有的Motif相似。在LEA\_3亚家族中,只有*SiLEA3-1*不含Motif 2,但其含有Motif 15。*SiLEA3-5*除包含Motif 2之外,还包含Motif 15。SMP亚家族成员中*SiSMP-1*、*SiSMP-2*、*SiSMP-3*、*SiSMP-5*、*SiSMP-6*都含有Motif 8与Motif 6,而*SiSMP-4*只含有Motif 8。*SiSMP-1*和*SiSMP-2*还含有Motif 13与Motif 20。*SiSMP-3*、*SiSMP-4*、*SiSMP-5*、*SiSMP-6*还含有Motif 7、Motif 9、Motif 17、Motif 18。LEA\_6与LEA\_1亚家族成员不包括任何Motif。DHN亚家族成员都含有

Motif 4。在DHN亚家族中,除*SiDHN-1*之外,其他成员还含有Motif 1和Motif 16。LEA\_5亚家族的成员都含有Motif 12。*SiLEA5-1*和*SiLEA5-2*还含有Motif 8和Motif 19。LEA\_2亚家族成员都含有Motif 11、Motif 5和Motif 3,成员*SiLEA2-1*还含有Motif 8和Motif 19。以上结果表明,不同亚家族的成员结构差异很大,说明LEA基因功能的复杂性和群体特异性。虽然在各亚家族中Motif类型存在差异,但同一组成员往往表现出相似的基序模式,表明它们之间的功能相似性。33个*SiLEAs*基因的外显子数目不同,1~4个不等(图3-C)。其中,有17个*SiLEAs*基因中含有1个内含子,13个*SiLEAs*基因中不含有内含子,*SiSMP-4*和*SiSMP-6*含有2个内含子,只有*SiDHN-4*含有3个内含子。



A. 系统发育树采用MGEA11邻接法,Bootstrap重复值设为1 000;B.*SiLEAs*基因的保守基序分布,Motif 1~Motif 20的用不同颜色的方框显示;C.*SiLEAs*基因的基因结构,未翻译区、外显子和内含子分别用彩色方框和黑线表示,蛋白质的长度用底部的刻度表示

A. MGEA11 adjacency method was adopted for the phylogenetic tree, and the Bootstrap repeat value was set to 1 000; B. Distribution of conserved motif of *SiLEA* genes. Motif 1~Motif 20 was shown in different colored boxes; C. Genetic structure of *SiLEAs* gene. Untranslated regions, exons, and introns were represented by colored boxes and black lines, respectively. The length of the protein was indicated by the scale at the bottom

图3 谷子LEAs基因的系统发育关系、保守基序、结构域和外显子-内含子结构

Fig.3 Phylogenetic relationship, conserved motif, domain, and exon-intron structure of LEAs gene in foxtail millet

### 2.4 谷子LEAs基因的顺式调控元件分析

进一步对33个*SiLEAs*基因启动子区域(起始密码上游2 000 bp)的顺式调控元件进行分析,结果表明,顺式调控元件主要分为3类,即激素响应类元件、胁迫响应类元件和植物生长发育类元件。

谷子*SiLEAs*基因成员中包括5类激素响应元件,主要有脱落酸(ABA)、生长素(IAA)、赤霉素(GA)、茉莉酸甲酯(MeJA)和水杨酸(SA)(图4)。

由图4可知,几乎所有的*SiLEAs*基因在启动子区都发现了光响应元件(G-box元件,33个成

员)、脱落酸响应元件(ABRE元件,32个成员)、茉莉酸甲酯(MeJA)响应元件(CGTCA-motif和TGACG-motif元件,29个成员)。26个*SiLEAs*基因成员的启动子区含有3~4个激素响应元件。*SiLEA1-1*、*SiLEA2-3*、*SiLEA2-4*、*SiLEA3-4*、*SiLEA5-1*、*SiSMP-3*和*SiSMP-5*成员包含5个激素响应元件。在5个激素响应元件中,脱落酸(ABA)响应元件数量最多,其次是茉莉酸甲酯(MeJA)响应元件、赤霉素(GA)响应元件。除激素响应元件之外,*SiLEAs*基因中胁迫响应元件和植物生长发育元件包括:玉米蛋白代谢调控( $O_2$ -site)元件、MYB结合位点(MBS和MER)元件、低

温响应(LTR)元件、胚乳表达(GCN4\_motif)元件、分生组织表达(CAT-box)元件、种子特异性调控(RY-element)元件。*SiLEA3-3*不含有任何胁迫和植物生长发育因子的元件。含有MYB结合位点(MBS)元件基因数量最多,其次是分生组织表达(CAT-box)元件、低温响应(LTR)元件和玉米蛋白代谢调控( $O_2$ -site)元件。*SiLEAs*基因可能参与了激素响应过程。此外,一些顺式元件可能在发育过程中调控不同组织(种子、胚乳、分生组织)的表达。这些结果表明,*SiLEAs*基因不仅参与植物生长发育过程,还可能响应各种非生物胁迫。

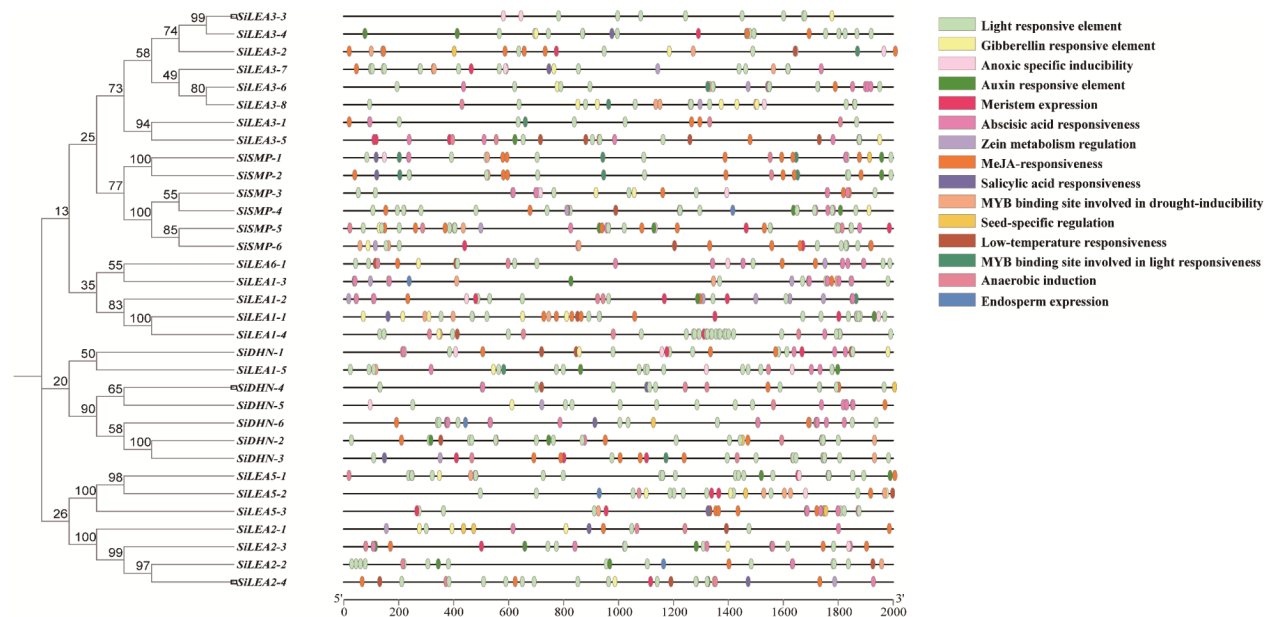


图4 谷子*LEAs*基因成员的顺式调控元件分析

Fig.4 Analysis of *cis*-regulatory elements of *LEAs* gene members in foxtail millet

## 2.5 谷子*LEAs*基因的共线性分析

为研究*SiLEAs*基因的重复基因和共线关系,本研究对*SiLEAs*基因进行了物种内共线性分析(图5),对谷子与其他5个代表性作物(3个单子叶作物:玉米、高粱、水稻;2个双子叶作物:拟南芥、大豆)物种间进行共线性分析(图6)。结果表明,在*SiLEAs*中发现了3对片段重复基因(*SiLEA1-4*与*SiLEA1-1*、*SiLEA2-4*与*SiLEA2-2*和*SiLEA3-3*与*SiLEA3-2*)。这些共线基因分布在第2、3、5、6号染色体上(图1),3对片段重复基因分别属于同一亚家族,这表明染色体片段在进化过程中可能发生了复制事件,并未完全分化,在功能上可能会存在冗余性。在共线关系中,总共发现有

25个*SiLEAs*基因与拟南芥(1对)、玉米(17对)、大豆(10对)、水稻(22对)和高粱(24对)存在共线基因对。共线性图谱显示,*SiLEAs*与高粱共线性最高,其次是水稻、玉米、大豆和拟南芥。进一步研究发现,一些*SiLEAs*基因被发现存在于5个物种的多个共线基因对中,例如,*SiSMP-5*与*Zm00001eb057120\_T001/Os03t0747400-00/AT3G22490.1/KRH32031/KRG91319/EER90879*,表明这些共线基因可能在祖先分化之前就已经存在。此外,部分*SiLEAs*基因(7个基因:*SiDHN-1*、*SiLEA2-1*、*SiLEA2-4*、*SiLEA3-2*、*SiLEA3-3*、*SiLEA3-5*、*SiLEA6-1*)仅与玉米、高粱和水稻(单子叶植物)存在共线性,且谷子与单子叶植物共线性的基因

数量远高于双子叶植物的基因数量,表明单子叶植物之间具有更亲密的亲缘关系。

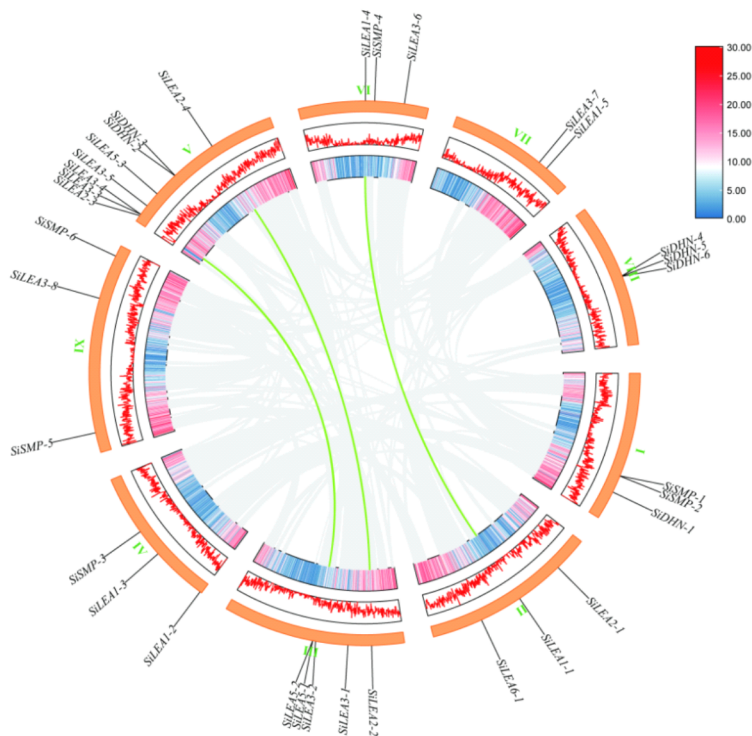


图5 谷子 *LEAs* 基因成员的共线性分析

Fig.5 Collinearity analysis of *LEAs* gene members in foxtail millet

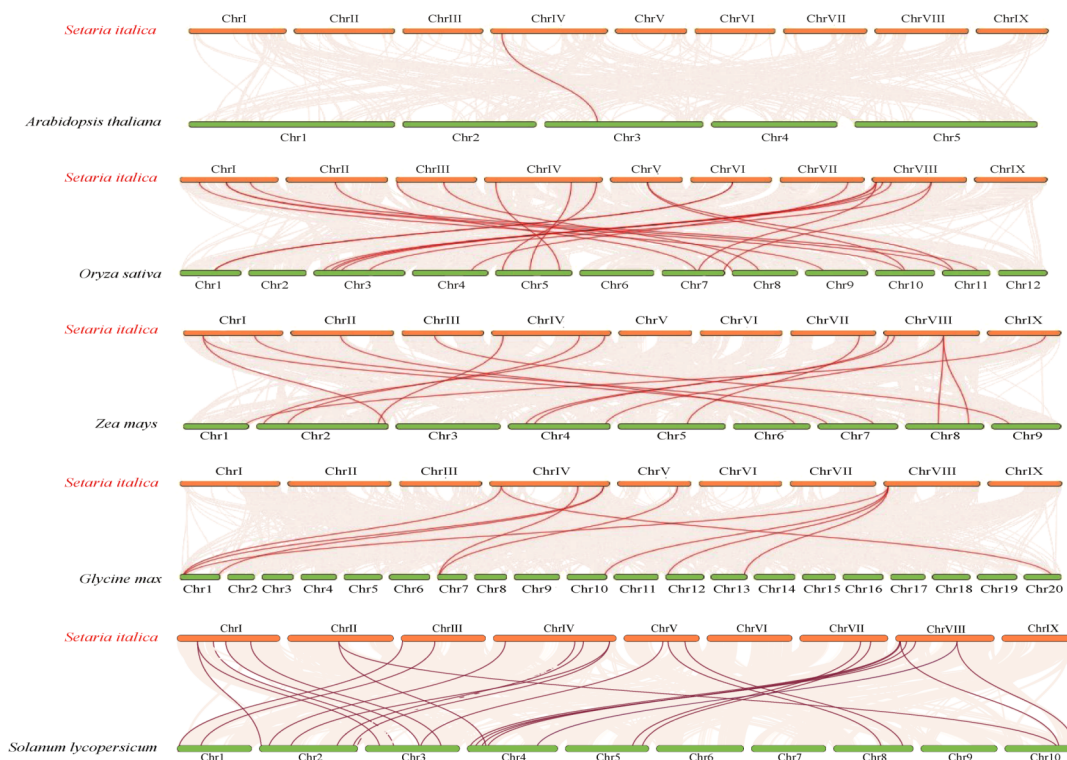


图6 谷子与拟南芥、水稻、玉米、大豆和高粱 *LEAs* 基因家族成员的共线性分析

Fig.6 Collinearity analysis of *LEAs* gene family members in foxtail millet with *Arabidopsis*, rice, maize, soybean, and sorghum

### 2.6 谷子 *LEAs* 基因的编码蛋白互作分析

谷子 *LEAs* 基因成员编码蛋白互作如图 7 所示。

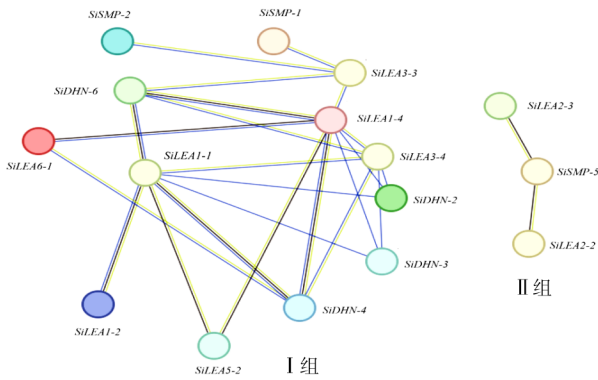


图 7 谷子 *LEAs* 基因成员编码蛋白互作

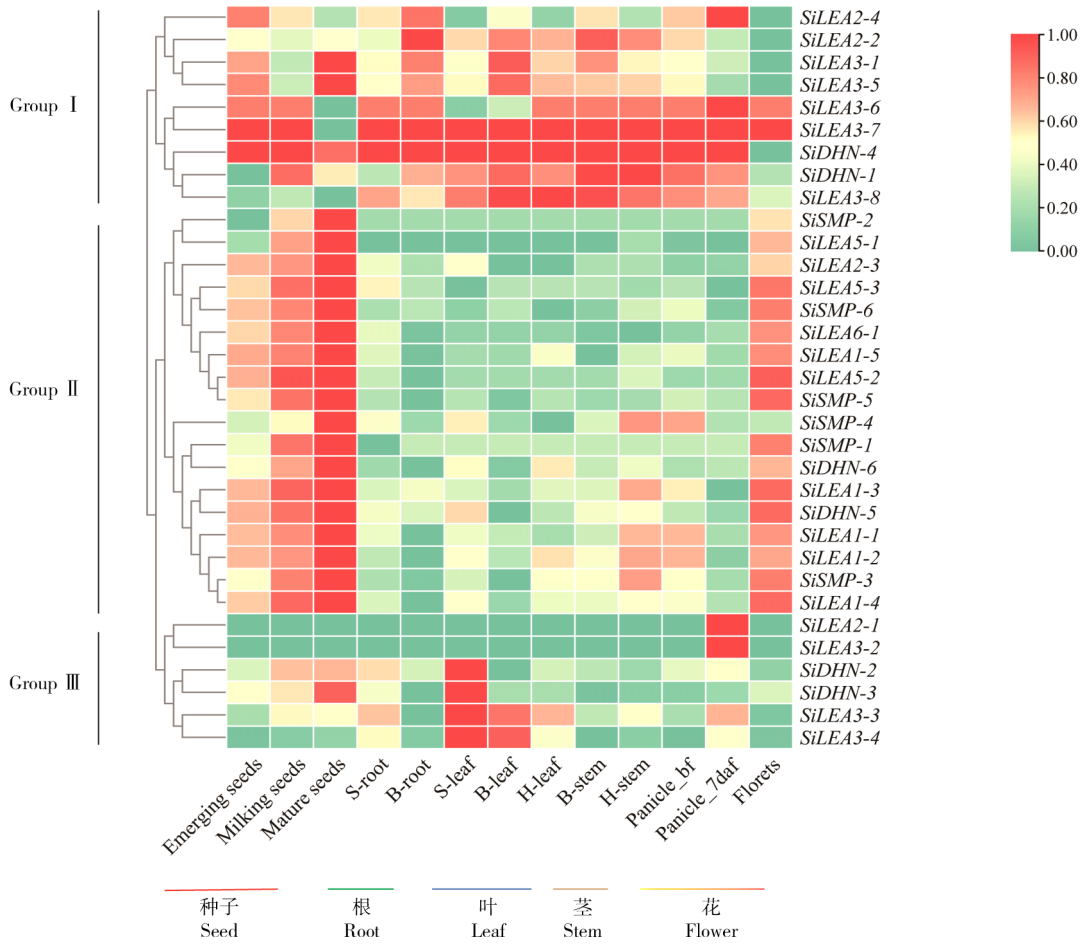
Fig.7 Interactions of proteins encoded by *LEAs* gene members in foxtail millet

为更好了解 *SiLEAs* 蛋白之间的关联,本研究

对 *SiLEAs* 的编码蛋白进行蛋白互作分析(图 7),结果显示,*SiLEAs* 中有 16 个蛋白(48%)存在互作关系,共分为 2 组。I 组包括 13 个 *SiLEAs* 蛋白(*SiSMP*-1、*SiSMP*-2、*SiDHN*-2、*SiDHN*-3、*SiDHN*-4、*SiDHN*-6、*SiLEA3*-3、*SiLEA3*-4、*SiLEA1*-1、*SiLEA1*-2、*SiLEA1*-4、*SiLEA5*-2), II 组包括 3 个 *SiLEAs* 蛋白(*SiLEA2*-3、*SiLEA2*-2、*SiSMP*-5)。从分组中可以看出, I 组中包含 *SMP*、*DHN*、*LEA\_6*、*LEA\_3*、*LEA\_1* 和 *LEA\_5* 这 6 个亚家族, II 组中包含 *SMP* 和 *LEA\_2* 这 2 个亚家族。表明同一分组中的亚家族可能在谷子生长发育过程中某阶段发挥同一作用。

### 2.7 谷子 *LEAs* 基因的器官表达模式分析

为了确定 *SiLEAs* 在谷子各组织中的表达模式,本研究分析了 *SiLEAs* 在豫谷 1 号不同组织中的表达情况(图 8)。



S. 幼苗期; B. 孕穗期; H. 抽穗期

S. Seedling stage; B. Booting stage; H. Heading stage

图 8 谷子 *LEAs* 基因成员的不同阶段不同器官表达模式

Fig.8 Expression patterns in different organs at different stages of the *LEAs* gene members in foxtail millet

从图8可以看出, *SiLEAs* 在不同时期不同组织中的表达模式主要分为3组。I组在谷子整个发育阶段均有较高表达, *SiLEA3-6*、*SiLEA3-7*、*SiDHN-4*、*SiDHN-1*和*SiLEA3-8*在根、茎、叶、开花前和开花后7d中高表达,说明这5个基因在谷子根、茎、叶和开花的生长发育过程中发挥重要作用。II组在种子和花中有较高表达,尤其是在成熟种子中,表明在种子成熟时可能发挥一定的作用。III组中, *SiLEA2-1*和*SiLEA3-2*在开花后7d有较高表达,说明这2个基因可能在开花过程中有一定作用。*SiLEA3-3*在发育中的叶中特异性高表达,可能与叶的发育有关。不同的表达模式表明, *SiLEAs* 在谷子生长发育过程中不同阶段可能发挥不同的作用。

### 3 结论与讨论

本研究对谷子 *LEA* 基因家族成员进行了鉴定和生物信息学分析,共鉴定了33个 *SiLEAs* 基因,分为7个亚家族。拟南芥 *LEA* 基因含有51个基因,分为9个亚家族<sup>[3]</sup>,水稻 *LEA* 基因含有34个基因,分为7个亚家族<sup>[6]</sup>,谷子中基因数目介于拟南芥和水稻之间,亚家族数目与水稻相同。这种现象的发生可能是由基因组和环境变化所导致。根据 *SiLEAs* 蛋白的亲水性平均值(GRAVY),本研究表明,大多数 *SiLEAs* 蛋白都是亲水性的。这与许多植物中鉴定出的 *LEA* 蛋白特性相似。

基因结构分析表明,在鉴定的33个 *SiLEAs* 基因中,大部分成员(约90.9%)含有1个内含子或不内含子,这与其他植物中已报道的 *LEA* 基因的基因结构一致。例如,小麦中约60%的 *LEAs* 基因缺乏内含子,而拟南芥中超过1/2的 *LEAs* 基因每个成员仅包含1个内含子<sup>[25]</sup>。从植物体胁迫反应的角度来看,较少的内含子结构减少了防御胁迫从转录到翻译的时间,降低了能量消耗的成本,有利于植物对不利环境条件引起的刺激做出快速反应<sup>[26]</sup>。这也说明了大多数的 *SiLEAs* 含有0~1个内含子,在非生物胁迫下能被迅速、强烈地诱导或抑制。此外,在同一亚家族中,不仅具有相似的基因结构和保守基序,在亚细胞定位方面也具有相似的结果。这些发现表明, *SiLEAs* 属于同一亚家族的成员可能在谷子生长发育过程中发挥相似的作用。在 *SiLEAs* 基因与拟南芥、玉米、小麦、高粱的 *LEA* 基因亚家族的系统发育树中可以看出,谷子、拟南芥、玉米、小麦和高粱的 *LEAs* 基因成员

在各亚家族的类群中分布较均匀,表明其起源可能出现在单子叶植物和双子叶植物分化之前<sup>[3,6]</sup>。

对顺式元件结果分析,发现在 *SiLEAs* 基因中含有各种激素响应类元件、逆境响应元件和生长发育响应元件。其中,全部的 *SiLEAs* 基因成员都含有激素响应元件,32个 *SiLEAs* 基因成员含有逆境响应元件和生长发育响应元件。表明 *SiLEAs* 基因成员可能在非生物胁迫过程中发挥重要作用并具有复杂的调节机制。已有研究表明,这些顺式作用元件对于基因响应相应的外界胁迫具有重要的作用。顺式作用元件的存在以及数量的多少对分析基因能否响应相关胁迫具有一定的参考价值,为进一步研究谷子 *LEA* 基因调控谷子生长发育与逆境响应的分子机制提供一定的理论基础。

对 *SiLEAs* 基因在谷子的根、茎、叶和花中的表达模式进行研究,70%的 *SiLEAs* 基因在成熟种子中表达相对较高。这与 *LEA* 基因在种子发育晚期积累的特性一致<sup>[27]</sup>。6个 *LEA\_3* (*SiLEA3-1*、*SiLEA3-3*、*SiLEA3-5*、*SiLEA3-6*、*SiLEA3-7*和*SiLEA3-8*)和4个 *DHN* (*SiDHN-1*、*SiDHN-2*、*SiDHN-3*和*SiDHN-4*)亚家族基因成员在根中表达相对较高, *LEA\_1* 亚家族的全部成员和大部分 *SMP* (5个基因: *SiSMP-1*、*SiSMP-2*、*SiSMP-3*、*SiSMP-5*和*SiSMP-6*)亚家族成员在花中表达较高。推测 *LEA\_3* 和 *DHN* 亚家族在谷子根部生长期起到一定作用, *LEA\_1* 和 *SMP* 亚家族在开花期发挥一定作用。此外, *SiLEA3-6*、*SiLEA3-7*、*SiDHN-4*、*SiDHN-1*和*SiLEA3-8*在谷子生长发育中5个期间表达量较高,并且其含有分生组织表达(CAT-box)元件或玉米蛋白代谢调控( $O_2$ -site)元件或种子特异性调控(RY-element)元件,推测这5个基因可能在谷子生长发育中发挥重要作用,在接下来的研究中,可以进一步研究和挖掘其生物学功能。这些结果初步表明,谷子 *LEA* 基因家族成员可能参与了植物种子的生长发育及对胁迫的响应等多种生理功能,从而使该家族的基因表达呈现多种模式。然而, *SiLEAs* 基因在谷子生长发育与胁迫过程中发挥的具体功能及机制还需要通过试验进行全面系统验证。

#### 参考文献:

- [1] INGRAM J, BARTELS D. The molecular basis of dehydration tolerance in plants[J]. Annual Review of Plant Physiology and Plant Molecular Biology, 1996, 47: 377-403.
- [2] DURE 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[J]. *Biochemistry*, 1981, 20(14):4162-4168.
- [3] HUNDERTMARK M, HINCHA D K. LEA (Late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*[J]. *BMC Genomics*, 2008, 9:118.
- [4] ZHANG Y P, ZHANG X J, ZHU L J, et al. Identification of the maize *LEA* gene family and its relationship with kernel dehydration[J]. *Plants*, 2023, 12(21):3674.
- [5] 刘浩. 小麦 *LEA* 基因家族的鉴定及脱水蛋白 WZY2 响应干旱胁迫的调节机制[D]. 杨凌:西北农林科技大学, 2021.
- LIU H. Identification of wheat *LEA* gene family and regulatory mechanism of dehydrin WZY2 in response to drought stress [D]. Yangling: North West Agriculture and Forestry University, 2021.
- [6] WANG X S, ZHU H B, JIN G L, et al. Genome-scale identification and analysis of *LEA* genes in rice (*Oryza sativa* L.) [J]. *Plant Science*, 2007, 172(2):414-420.
- [7] JIA C P, GUO B, WANG B K, et al. The *LEA* gene family in tomato and its wild relatives: genome-wide identification, structural characterization, expression profiling, and role of *SlLEA6* in drought stress[J]. *BMC Plant Biology*, 2022, 22(1):596.
- [8] LIANG Y, XIONG Z Y, ZHENG J X, et al. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*[J]. *Scientific Reports*, 2016, 6:24265.
- [9] İBRAHİME M, KIBAR U, KAZAN K, et al. Genome-wide identification of the *LEA* protein gene family in grapevine (*Vitis vinifera* L.) [J]. *Tree Genetics & Genomes*, 2019, 15(4):55.
- [10] GARAY-ARROYO A, COLMENERO-FLORES J M, GARCİARRUBIO A, et al. Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit[J]. *Journal of Biological Chemistry*, 2000, 275(8):5668-5674.
- [11] OLVERA-CARRILLO Y, CAMPOS F, REYES J L, et al. Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in *Arabidopsis*[J]. *Plant Physiology*, 2010, 154(1):373-390.
- [12] HAND S C, MENZE M A, TONER M, et al. LEA proteins during water stress: not just for plants anymore[J]. *Annual Review of Physiology*, 2011, 73:115-134.
- [13] XIAO B Z, HUANG Y M, TANG N, et al. Over-expression of a *LEA* gene in rice improves drought resistance under the field conditions[J]. *Theoretical and Applied Genetics*, 2007, 115(1):35-46.
- [14] LIU Y, LIANG J N, SUN L P, et al. Group 3 *LEA* protein, *ZmLEA3*, is involved in protection from low temperature stress[J]. *Frontiers in Plant Science*, 2016, 7:1011.
- [15] SHI H F, HE X Y, ZHAO Y J, et al. Constitutive expression of a group 3 *LEA* protein from *Medicago falcata* (*MfLEA3*) increases cold and drought tolerance in transgenic tobacco[J]. *Plant Cell Reports*, 2020, 39(7):851-860.
- [16] 闫留延, 李剑峰, 张世文, 等. 谷子 *SiPRR73* 基因的光温调控模式及非生物胁迫响应特性[J]. *华北农学报*, 2022, 37(4):11-19.
- YAN L Y, LI J F, ZHANG S W, et al. Regulation pattern of photoperiod, temperature and abiotic stress on *SiPRR73* in foxtail millet[J]. *Acta Agriculturae Boreali-Sinica*, 2022, 37(4):11-19.
- [17] 赵绍兴, 赵玲, 张皓珊, 等. 谷子苗期对不同干旱胁迫的生理响应[J]. *山西农业科学*, 2024, 52(5):18-23.
- ZHAO S X, ZHAO L, ZHANG H S, et al. Physiological responses of foxtail millet to different drought stress at seedling stage[J]. *Journal of Shanxi Agricultural Sciences*, 2024, 52(5):18-23.
- [18] SUDHAKAR C, VEERANAGAMALLAIAH G, NARESH-KUMAR A, et al. Polyamine metabolism influences antioxidant defense mechanism in foxtail millet (*Setaria italica* L.) cultivars with different salinity tolerance[J]. *Plant Cell Reports*, 2015, 34(1):141-156.
- [19] 侯思宇, 刘超, 马赟骁, 等. 谷子基因组学研究进展[J]. *山西农业科学*, 2024, 52(1):1-9.
- HOU S Y, LIU C, MA Y X, et al. Research advances in foxtail millet genomics[J]. *Journal of Shanxi Agricultural Sciences*, 2024, 52(1):1-9.
- [20] HE Q, WANG C C, HE Q, et al. A complete reference genome assembly for foxtail millet and *Setaria-db*, a comprehensive database for *Setaria*[J]. *Molecular Plant*, 2024, 17(2):219-222.
- [21] CHEN C J, CHEN H, ZHANG Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data[J]. *Molecular Plant*, 2020, 13(8):1194-1202.
- [22] SUBRAMANIAN B, GAO S H, LERCHER M J, et al. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees[J]. *Nucleic Acids Research*, 2019, 47:W270-W275.
- [23] TANG S, ZHAO Z Y, LIU X T, et al. An E2-E3 pair contributes to seed size control in grain crops[J]. *Nature Communications*, 2023, 14(1):3091.
- [24] 王震, 杜小云, 刘文, 等. 高粱 *LEA* 基因家族的鉴定及表达分析[J]. *生物资源*, 2019, 41(4):324-334.
- WANG Z, DU X Y, LIU W, et al. Genome-wide identification and expression analysis of *LEA* gene family in *Sorghum bicolor* L [J]. *Biotic Resources*, 2019, 41(4):324-334.
- [25] ZAN T, LI L Q, LI J T, et al. Genome-wide identification and characterization of late embryogenesis abundant protein-encoding gene family in wheat: evolution and expression profiles during development and stress[J]. *Gene*, 2020, 736:144422.
- [26] DEFOORT J, VAN DE PEER Y, CARRETERO-PAULET L. The evolution of gene duplicates in angiosperms and the impact of protein-protein interactions and the mechanism of duplication[J]. *Genome Biology and Evolution*, 2019, 11(8):2292-2305.
- [27] 王梦飞, 滑璐玢. *LEA* 蛋白及其在作物抗逆过程中的作用[J]. *北方农业学报*, 2018, 46(4):70-76.
- WANG M F, HUA L F. *LEA* protein and its role in the process of crop resistance[J]. *Journal of Northern Agriculture*, 2018, 46(4):70-76.