

蓖麻 *LEA* 基因家族的全基因组鉴定、分类及其进化分析

邹 智¹, 黄启星², 安 锋¹

(1. 中国热带农业科学院橡胶研究所/农业部橡胶树生物学与遗传资源利用重点实验室, 海南 儋州, 571737;
2. 中国热带农业科学院热带生物技术研究所/农业部热带作物生物学与遗传资源利用重点实验室, 海南 海口, 571101)

摘要: 基于已公布的基因组和 EST 数据对蓖麻 *LEA* 基因家族进行全面鉴定和系统命名, 并在此基础上分析了其基因结构、生化特征、进化关系和表达特性。结果显示, 蓖麻基因组中存在 27 个 *LEA* 基因, 分属于 *LEA*_1、*LEA*_2、*LEA*_3、*LEA*_4、*LEA*_5、*LEA*_6、*dehydrin* 和 *SMP* 8 个亚家族, 它们散布于 24 条 scaffold 上, 含有 0~2 个内含子不等, 且其中有 3 个基因存在可变剪接; 根据编码蛋白中的 Pfam 结构域类型及其进化关系, 将这些基因依次命名为 *RcLEA*1-1 和 -2、*RcLEA*2-1 和 -2、*RcLEA*3-1 至 -3、*RcLEA*4-1 至 -7、*RcLEA*5-1 和 -2、*RcLEA*6-1 和 -2、*RcLEA*7-1 至 -5 和 *RcLEA*8-1 至 -4; 利用 BLAT 分析基因表达谱显示, 在叶片、花、II/III 期胚乳、V/VI 期胚乳和种子等组织中, 所有 *RcLEA* 基因都有表达; 启动子分析表明, *RcLEA* 基因的启动子区富含 LTRE、ABRE、MYC、MYB 和 W-box 等逆境应答相关顺式作用元件。

关键词: 蓖麻; 全基因组; *LEA* 基因家族; 分类; 进化分析

中图分类号: S565.603 **文献标识码:** A **文章编号:** 1007-9084(2013)06-0637-07

Genome-wide identification, classification and phylogenetic analysis of *LEA* gene family in castor bean (*Ricinus communis* L.)

ZOU Zhi¹, HUANG Qi-xing², AN Feng¹

(1. Key Laboratory of Biology and Genetic Resources of Rubber Tree, Ministry of Agriculture/Rubber Research Institute (RRI), Chinese Academy of Tropical Agricultural Sciences (CATAS), Danzhou 571737, China;

2. Institute of Tropical Biosciences and Biotechnology, CATAS/Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture, Haikou 571101, China)

Abstract: Late embryogenesis abundant (LEA) proteins, a class of extremely hydrophilic proteins which were first identified in plant embryogenesis, were also found in vegetative tissues. They play important roles in plant desiccation tolerance. In this study, a genome-wide search was carried out to identify LEA genes in castor bean (*Ricinus communis* L.), a Euphorbiaceae plant that is well known for its plasticity to unfavorable environments. As a result, a total of 27 RcLEAs representing 8 groups were identified, 3 of which were found to have alternative splicing isoforms. Based on protein characteristics and phylogenesis, the 27 *RcLEAs* contained 0 to 2 introns and distributed across 24 released scaffolds, and were named as *RcLEA*1-1 and -2, *RcLEA*2-1 and -2, *RcLEA*3-1 to -3, *RcLEA*4-1 to -7, *RcLEA*5-1 and -2, *RcLEA*6-1 and -2, *RcLEA*7-1 to -5 and *RcLEA*8-1 to 4. Using the BLAT method, all *RcLEA* genes were expressed in at least one of the examined tissues such as leaf, flower, endosperm_II_III, endosperm_V_VI and seed. Their promoters enriched stress-responsive cis-acting elements including LTRE, ABRE, MYC, MYB and W-box.

Key words: *Ricinus communis* L.; Genome-wide; *LEA* gene family; Classification; Phylogenetic analysis

胚胎发育晚期丰富蛋白 (Late embryogenesis abundant proteins, LEAP) 是一类广泛存在于生物体

收稿日期: 2013-02-25

基金项目: 国家自然科学基金 (31100460); 海南省重点科技项目 (90107); 海南省自然科学基金 (312026); 中国热带农业科学院橡胶研究所基
本科研业务费专项 (1630022011014)

作者简介: 邹 智, 助理研究员, 主要研究方向为植物抗逆生物学, E-mail: zouzhi2008@126.com

内参与渗透调节的小分子量蛋白质,最早因发现其在胚胎发育的后期大量积累而被命名^[1,2]。研究表明,除种子外,LEA 蛋白在根、茎、叶等组织中也广泛存在,它们大多受干旱、低温、高渗透、盐胁迫、ABA、紫外辐射和 NaHCO₃ 等环境因子所诱导^[3]。至今,LEA 类基因已在各种植物、细菌、真菌、绿藻和低等动物中得到了克隆和鉴定,过量表达 LEA 类基因可以提高拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)、小麦(*Triticum aestivum*)、烟草(*Nicotiana tabacum*)、柑桔(*Citrus reticulata*)、草莓(*Fragaria ananassa*)、酵母和大肠杆菌等的抗寒、抗旱或耐盐能力^[3~7]。基因组范围内的分析表明,根据 Pfam 结构域类型,植物中的 LEAP 可分为 LEA_1、LEA_2、LEA_3、LEA_4、LEA_5、LEA_6、dehydrin 和 SMP (seed maturation protein) 等 8 大类^[8~13]。

蓖麻(*Ricinus communis* L., 2n = 20)是大戟科(Euphorbiaceae)一年生或多年生灌木。与大多数大戟科植物类似,蓖麻的适应能力特别强。2010 年蓖麻基因组草图的绘制使其成为大戟科第一个完成基因组测序的物种^[14],为通过比较基因组学研究大戟科植物耐胁迫的分子机制创造了条件。本研究基于已公布的基因组和 EST 数据对蓖麻 LEA 基因家族进行全面鉴定和系统命名,并在此基础上分析其基因结构、生化特征、进化关系及其表达特性,以期为进一步的功能研究与利用提供参考。

1 材料与方法

1.1 材料

蓖麻的全基因组序列下载于 phytozome v9.0 (<http://www.phytozome.net/ricinus.php>);转录组数据、cDNA 及 EST 序列下载于 NCBI (<http://www.ncbi.nlm.nih.gov/>)。拟南芥和水稻的 LEA 基因序列和蛋白序列分别下载自 TAIR (<http://www.arabidopsis.org>) 和 TIGR (<http://rice.plantbiology.msu.edu/>)。

1.2 蓖麻 LEA 基因的鉴别与分类

以 Pfam 数据库 (<http://www.sanger.ac.uk/Software/Pfam/>) 中 LEA 蛋白家族相关的结构域 LEA_1 (PF03760)、LEA_2 (PF03168)、LEA_3 (PF03242)、LEA_4 (PF02987)、LEA_5 (PF00477)、LEA_6 (PF10714)、dehydrin (PF00257) 和 SMP (PF04927) 及拟南芥和水稻的 LEA 蛋白作为 tBLASTn^[15] 分析的查询序列,搜索(为避免遗漏,将 E 值设为 10⁻³)^[16] 并获取蓖麻基因组和 EST 库中的同源序列,候选基因进行 Pfam 分析(<http://pfam.sanger.ac.uk/search>) 以确认其编码蛋白中存在 LEA 相关结构域(注:如未特别声明,在线软件的分析均采用默认参数,下同)。

1.3 蓖麻 LEA 蛋白的生化特性、系统发生学分析及系统命名

采用 ProtParam (<http://web.expasy.org/prot-param/>) 预测蛋白的理论分子量 (Mw)、等电点 (pI)、GRAVY;采用 TargetP 1.1 (<http://www.cbs.dtu.dk/services/TargetP/>) 预测蛋白的亚细胞定位;采用 ClustalW2^[17] 进行多序列比对后,应用 MEGA 4.0^[18] 采用 Neighbor - Joining 法构建系统发生树,bootstrap 值设为 1 000;根据 Pfam 结构域类型和进化关系对各家族成员进行分类和系统命名。

1.4 蓖麻 LEA 基因的表达特性及其启动子区域的顺式作用元件分析

基因表达谱分析采用 BLAT^[19],即将叶片 (NCBI SRA 登录号 ERX021378)、花 (ERX021379)、II/III 期胚乳 (ERX021375)、V/VI 期胚乳 (ERX021376) 和种子 (ERX021377) 的转录组数据比对到蓖麻的基因序列上,按照最小相似度 (min Identity) > 95,匹配长度 > 50bp 的条件过滤后,以 RPKM (Reads per kilobases per million reads) 值比较基因在不同组织中的表达丰度。RPKM 计算公式如下:

$$RPKM = (10^9 \times C) \div (N \times L)$$

C 为比对到某个基因的 reads (读段) 数;N 为样本比对到的所有基因总读段数;L 为基因长度。

采用 PLACE (<http://www.dna.affrc.go.jp/PLACE/signalscan.html>) 分析基因起始密码子上游 1 000bp 序列中的 LTRE (low - temperature responsive element, CCGAC)、ABRE (ABA - responsive element, ACGTG)、MYC (CACATG)、MYB (TAACCA) 和 W - box (TGAC) 等顺式作用元件。

2 结果与分析

2.1 蓖麻 LEA 基因的鉴别与分类

综合 Pfam 结构域和先前的报道^[8~10],分别从拟南芥和水稻中鉴定出 45 个 (无 Pfam 结构域的 At1g22600、At5g15960、At5g15970、At3g02480、At5g38760、At5g53820、At2g42530、At2g42540 和 At4g13230 未计入) 和 36 个 LEA 基因 (补充了早期未被鉴定的 LOC_Os09g30438 和 LOC_Os09g30418)。

通过结构域和同源搜索蓖麻的基因组和 EST 序列,总共鉴定出 27 个 LEA 基因,它们分属于 LEA

_1(2 个)、LEA_2(2 个)、LEA_3(3 个)、LEA_4(7 个)、LEA_5(2 个)、LEA_6(2 个)、dehydrin(5 个)和 SMP(4 个)等 8 个亚家族。在最新释放的 phytozome v9.0 中,这 27 个 *LEA* 基因被注释成“Late embryo-genesis abundant protein”的有 9 个,被注释成“late embryogenesis abundant”的有 4 个,被注释成“De-hydrin Xero”的有 2 个,被注释成“Late seed maturation protein P8B6”、“Embryonic abundant protein”、

“Phosphoprotein ECPP44”、“Indole – 3 – acetic acid – induced protein ARG2”的各有 1 个,另外 8 个则被注释成“conserved hypothetical protein”(表 1)。
从基因结构来看,这些基因包含 0~2 个内含子不等(无内含子的有 4 个,约占 14.8%;含有 1 个内含子的有 18 个,约占 66.7%;含有 2 个内含子的有 5 个,约占 18.5%),cDNA 编码区长度在 243~1 677 bp 之间,基因编码区长度在 243~2 494 bp 之间(表 1)。

表 1 蓖麻基因组中鉴定的 *LEA* 基因
Table 1 List of *LEA* genes identified from castor bean

Pfam	基因组注释号 Phytozome ID	注释 Phytozome description	cDNA ORF/bp	ORF /bp	内含子 Intron	EST	Scaffold 编号 Scaffold ID	可变剪接 Alternative splicing
LEA_1	29709. m001231	Conserved hypothetical protein	315	392	1	1	scf1106159298126	
	29827. m002548	Conserved hypothetical protein	375	471	1	1	scf1106159300308	
LEA_2	30174. m008670	Late embryogenesis abundant protein Lea14 – A, putative	456	978	1	6	scf1106159290416	Yes
	28266. m000192	Late embryogenesis abundant protein Lea14 – A, putative	942	1 589	1	4	scf1106159309564	
LEA_3	29308. m000187	Conserved hypothetical protein	276	358	1	0	scf1106159302826	Yes
	29726. m004064	Late embryogenesis abundant protein Lea5, putative	294	436	1	15	scf1106159301046	
	29986. m001632	Indole – 3 – acetic acid – induced protein ARG2, putative	288	288	0	38	scf1106159292744	
LEA_4	29634. m002127	Late embryogenesis abundant, putative	1 350	1 950	2	3	scf1106159300426	
	29815. m000495	Late embryogenesis abundant, putative	1 221	1 507	2	44	scf1106159295146	
	29889. m003401	Late embryogenesis abundant, putative	717	841	1	5	scf1106159297646	
	30190. m011335	Late embryogenesis abundant, putative	798	990	1	9	scf1106159293958	
	29841. m002806	Late embryogenesis abundant protein D – 7, putative	393	608	1	2	scf1106159304484	
	29646. m001088	Conserved hypothetical protein	1 677	2 494	1	16	scf1106159291412	
	30074. m001413	Conserved hypothetical protein	1 485	1 818	1	0	scf1106159305650	
LEA_5	29601. m000448	Late seed maturation protein P8B6, putative	339	429	1	5	scf1106159292012	
	29814. m000729	Embryonic abundant protein, putative	285	453	1	0	scf1106159306036	
LEA_6	29172. m000221	Conserved hypothetical protein	291	291	0	0	scf1106159307590	
	29804. m001543	Conserved hypothetical protein	243	243	0	0	scf1106159302598	
dehydrin	30072. m000963	Conserved hypothetical protein	291	291	0	2	scf1106159303528	Yes
	30170. m014323	Phosphoprotein ECPP44, putative	693	820	1	18	scf1106159304512	
	30131. m007082	Late embryogenesis abundant protein, putative	570	810	1	11	scf1106159298566	
	29634. m002070	Dehydrin Xero, putative	450	621	1	3	scf1106159300426	
SMP	29634. m002071	Dehydrin Xero, putative	441	533	1	0	scf1106159300426	
	30128. m008646	Late embryogenesis abundant protein D – 34, putative	729	1 003	2	0	scf1106159305188	
	29844. m003281	Late embryogenesis abundant protein D – 34, putative	795	1 055	2	2	scf1106159296318	
	29647. m002057	Late embryogenesis abundant protein D – 34, putative	774	882	1	7	scf1106159296320	
	29647. m002058	Late embryogenesis abundant protein D – 34, putative	771	945	2	3	scf1106159296320	

从基因分布来看,这些基因散布于 24 条 scaffold 上(表 1),其中 scf1106159300426 分布最多,有 3 个。

从基因表达来看,除注释号为 29308. m000187、30074. m001413、29814. m000729、29172. m000221、29804. m001543、29634. m002071 和 30128. m008646

的 8 个基因在 NCBI EST 库中暂时还没有对应的 EST 外,其它基因都含有 1~44 条不等。此外,目前的 EST 数据显示,有 3 个 *RcLEA* 基因存在可变剪接(表 1)。

2.2 蓖麻 LEA 蛋白的生化特性、系统发生学分析及系统命名

蓖麻 LEA 蛋白最长的包含 558 个氨基酸残基,最短的仅有 80 个残基,其理论分子量在 8.67~60.64kD 之间;等电点在 4.54 到 10.16 之间;所有 *RcLEA* 蛋白的 GRAVY 值都小于 0,最小的为 -2.000;亚细胞定位显示,除 3 个蛋白被预测为线粒体蛋白,4 个被预测为分泌型蛋白外,其余的均无信号肽(表 2)。

表 2 蓖麻 LEA 蛋白的生化特性
Table 2 Biochemical features of castor bean LEAPs

基因组注释号 Phytozome ID	名称 Name	长度 Len	Mw /kDa	pI	GRAVY	Loc	拟南芥中同源基因 TAIR ID of AtLEA ortholog	水稻中同源基因 TIGR ID of OsLEA ortholog
29709. m001231	RcLEA1 -1	124	13.48	10.05	-0.891	other	AT1G32560	LOC_Os06g02040
29827. m002548	RcLEA1 -2	104	11.22	7.93	-0.628	other	AT1G32560	LOC_Os06g02040
30174. m008670	RcLEA2 -1	151	16.62	4.87	-0.084	other	AT1G01470	LOC_Os01g12580
28266. m000192	RcLEA2 -2	313	34.71	4.64	-0.341	other	AT2G44060	LOC_Os03g62620
29308. m000187	RcLEA3 -1	91	10.55	9.00	-0.436	mitochondrion	AT3G53770	LOC_Os01g21250
29726. m004064	RcLEA3 -2	97	10.57	9.35	-0.338	secreted	AT4G15910	LOC_Os01g21250
29986. m001632	RcLEA3 -3	95	9.95	10.16	-0.206	mitochondrion	AT4G02380	LOC_Os01g21250
29634. m002127	RcLEA4 -1	449	48.52	4.83	-0.828	secreted	AT4G36600	LOC_Os03g07180
29815. m000495	RcLEA4 -2	406	44.48	5.34	-1.322	other	AT2G36640	LOC_Os03g20680
29889. m003401	RcLEA4 -3	238	26.90	5.35	-1.260	secreted	AT2G18340	LOC_Os03g07180
30190. m011335	RcLEA4 -4	265	29.09	6.95	-1.191	mitochondrion	AT4G21020	LOC_Os04g52110
29841. m002806	RcLEA4 -5	130	14.43	7.91	-1.558	other	AT1G52690	LOC_Os05g46480
29646. m001088	RcLEA4 -6	558	60.64	5.89	-1.141	other	AT2G42560	LOC_Os02g15250
30074. m001413	RcLEA4 -7	494	54.11	7.68	-0.409	secreted	AT1G72100	LOC_Os11g44940
29601. m000448	RcLEA5 -1	112	12.10	5.80	-1.316	other	AT2G40170	LOC_Os05g28210
29814. m000729	RcLEA5 -2	94	10.23	5.25	-1.433	other	AT2G40170	LOC_Os05g28210
29172. m000221	RcLEA6 -1	96	10.60	5.36	-1.215	other	AT2G33690	LOC_Os09g30418
29804. m001543	RcLEA6 -2	80	8.67	6.72	-1.107	othe	AT2G33690	LOC_Os09g30418
30072. m000963	RcLEA7 -1	96	10.91	6.64	-2.000	other	AT1G54410	LOC_Os03g45280
30170. m014323	RcLEA7 -2	230	26.02	5.25	-1.565	other	AT1G20450	LOC_Os02g44870
30131. m007082	RcLEA7 -3	189	20.33	5.88	-1.161	other	AT2G21490	LOC_Os01g50700
29634. m002070	RcLEA7 -4	149	16.72	7.25	-1.608	other	AT3G50980	LOC_Os01g50700
29634. m002071	RcLEA7 -5	146	16.03	8.74	-1.384	other	AT3G50980	LOC_Os01g50700
30128. m008646	RcLEA8 -1	242	25.05	5.48	-0.453	other	AT3G22490	LOC_Os01g50700
29844. m003281	RcLEA8 -2	264	26.92	5.18	-0.272	other	AT3G22490	LOC_Os01g50700
29647. m002057	RcLEA8 -3	257	26.73	4.96	-0.437	other	AT3G22490	LOC_Os01g50700
29647. m002058	RcLEA8 -4	256	26.74	4.54	-0.364	other	AT3G22500	LOC_Os01g50700

系统发生学分析显示,与拟南芥和水稻的建树结果类似(结果未显示),除注释号为 30074. m001413 的 LEA 蛋白外,其它的蛋白都依 Pfam 结构域的分类聚在一起,形成 8 个明显的分支,其中, LEA_4 又可分成 2 个亚支(图 1)。为此,综合 Pfam 结构域类型和进化关系,依次将蓖麻 *LEA* 基因命名如表 2 和图 1 所示(虽然 30074. m001413 与 dehydrin 亚家族成员聚在一起,但因其含有的 Pfam 结

同源分析表明,27 个 *RcLEA* 蛋白在拟南芥和水稻中都存在同源蛋白(E 值设为 10^{-5}),其中,拟南芥中亲缘关系较近的有 24 个,除 AT1G32560(2)、AT2G40170(2)、AT2G33690(2)、AT3G50980(2)、AT3G22490(3)对应多个蓖麻 *RcLEA* 蛋白外,其余的均为——对应关系;相对而言,水稻中对应的同源蛋白数目相对较少,有 16 个,除 LOC_Os06g02040(2)、LOC_Os01g21250(3)、LOC_Os03g07180(2)、LOC_Os05g28210(2)、LOC_Os09g30418(2)、LOC_Os02g44870(2)、LOC_Os11g26570(2)和 LOC_Os03g06360(4)外,其余的均为——对应关系(表 2)。

构域为 LEA_4,将其命名为 *RcLEA4-7*)。

2.3 蓖麻 LEA 基因的表达特性及其启动子区域的顺式作用元件分析

基因表达谱的 BLAT 分析显示,在叶片、花、II/III 期胚乳、V/VI 期胚乳和种子等组织中,所有 *RcLEA* 基因都有表达,除 *RcLEA6-2* 外,其它 26 个 *RcLEA* 基因至少在 2 个以上的样本中表达:在叶片中,除 *RcLEA4-6*、*RcLEA6-2*、*RcLEA7-4*、*RcLEA7*

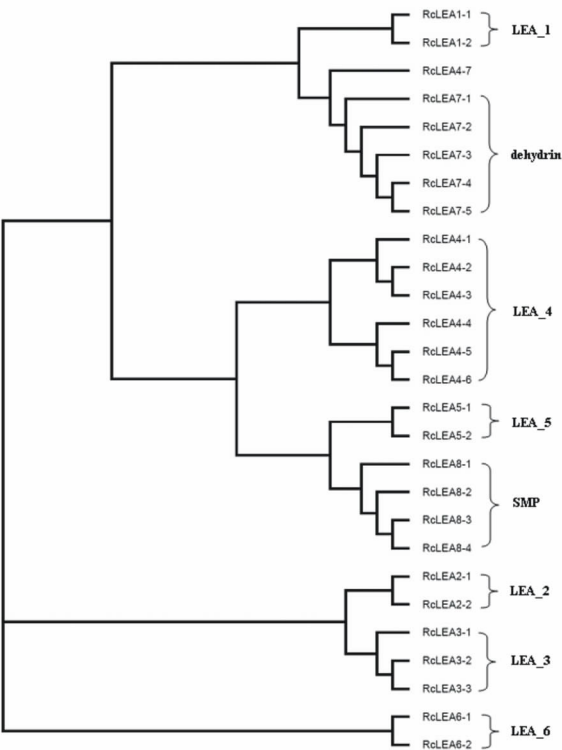
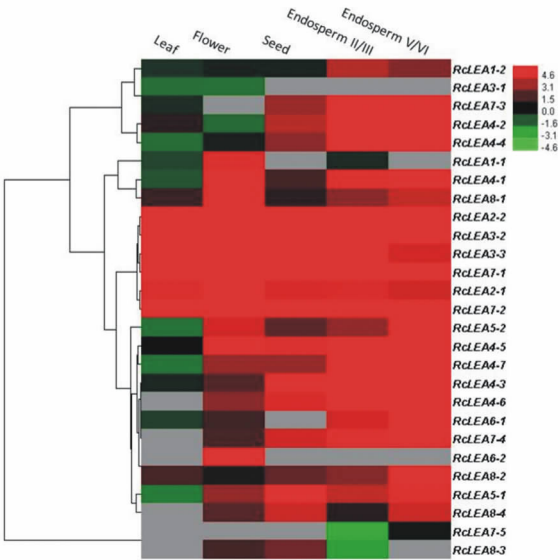


图1 蓖麻 *LEA* 蛋白的系统发生树
Fig.1 Phylogenetic analysis of RcLEAs using MEGA 4.0 with pairwise distance and Neighbor – Joining algorithm

-5、RcLEA8-3 和 RcLEA8-4 外,其它 21 个都有表达(总 *RPKM* 值为 2 486.67),其中,丰度最高的为 RcLEA7-1(*RPKM* 值为 1 913.62),丰度最低的为 RcLEA5-1(*RPKM* 值为 0.21),两者的表达差异接近 10 000 倍;在花中,除 RcLEA7-3 和 RcLEA7-5 外,其它 25 个都有表达(总 *RPKM* 值 3 170.78),其中,丰度最高的为 RcLEA7-1(*RPKM* 值为 1 071.22),丰度最低的为 RcLEA3-1(*RPKM* 值为 0.24),两者的表达差异接近 5 000 倍;在 II/III 期胚乳中,除 RcLEA3-1 和 RcLEA6-2 外,其它 26 个都有表达(总 *RPKM* 值为 2 119.74),其中,丰度最高的为 RcLEA7-2(*RPKM* 值为 598.10),丰度最低的为 RcLEA7-5(*RPKM* 值为 0.11),两者的表达差异超过 5 000 倍;在 V/VI 期胚乳中,除 RcLEA1-1、RcLEA3-1、RcLEA6-2 和 RcLEA8-3 外,其它 23 个都有表达(总 *RPKM* 值为 3 187.96),其中,丰度最高的为 RcLEA4-4(*RPKM* 值为 519.77),丰度最低的为 RcLEA7-5(*RPKM* 值为 0.86),两者的表达差异超过 600 倍;在种子中,除 RcLEA1-1、RcLEA3-1、RcLEA6-1、RcLEA6-2 和 RcLEA7-5 外,其它 22 个都有表达(总 *RPKM* 值为 2 275.77),其中,丰度最高的为 RcLEA7-1(*RPKM* 值为 1 075.50),丰度最低的为 RcLEA1-2(*RPKM* 值为 0.70),两者的表达差异超过 1 500 倍(图 2)。



注:色标表示 RPKM 的对数值;绿色代表低表达,红色代表高表达
Note: Color scale represents RPKM normalized log₁₀ transformed counts; light green indicates low expression and red indicates high expression

图2 蓖麻 *LEA* 基因在叶片、花、II/III 期胚乳、V/VI 期胚乳和种子的表达模式
Fig.2 Expression profiles of castor bean *LEA* genes in leaf, flower, endosperm_II_III, endosperm_V_VI and seed

起始密码子的 PLACE 分析显示,所有 *RcLEA* 基因的启动子区域都含有多个 MYC(最多的含有 20 个)和 WRKY(最多的含有 8 个)元件;约有 33.3% 的成员含有 LTRE 元件(最多的含有 4 个);约有 85.2% 的成员含有 ABRE 元件(最多的含有 6 个);约有 70.4% 的成员含有 MYB 元件(最多的含有 3 个)(表 3)。

3 讨论

虽然 LEAP 早在 1981 年就被发现^[1],并被证实与生物抗逆密切相关^[4],然而,至今其参与抗逆的分子机制还鲜为人知。随着越来越多物种基因组数据的公布,从全基因组水平分析和鉴定这一基因家族成为可能。至今,*LEA* 基因家族已在水稻^[8]、拟南芥^[9,10]、梅花(*Prunus mume*, 30 个)^[11]和大豆(36 个)^[12]得到了系统的鉴定。为探讨 LEAP 在蓖麻耐胁迫中可能作用,本研究基于已公布的基因组和 EST 序列对蓖麻的 *LEA* 基因进行了全面鉴定。

蓖麻基因组中存在 27 个 *LEA* 基因,这是迄今报道的家族成员最少的高等植物。根据其编码蛋白中 Pfam 结构域的类型,这些 *LEA* 基因分归为 LEA_1、LEA_2、LEA_3、LEA_4、LEA_5、LEA_6、dehydrin 和 SMP 8 个亚家族。与其它物种类似,蓖麻的 *LEA* 基因不含或少含内含子,内含子数目仅为 1 个的高达 66.7%(拟南芥为 66.7%^[10]、水稻为 63.8%^[8]、

大豆为 72.2%^[12]、梅花为 72.0%^[11]),而含 2 个内含子的基因(最多的也只有 2 个)所占比例不到 18.5%(表 1)。蓖麻有 3 个 *LEA* 基因存在可变剪

接,所占比例约为 11.1%,与水稻的 11.8%^[8]接近,但低于大豆的 16.7%^[12]和拟南芥的 22.2%^[10]。

表 3 蓖麻 *LEA* 基因启动子顺式元件的数量和分布
Table 3 Distribution of cis – acting elements in castor bean *LEA* gene promoters

基因 Gene	LTRE	ABRE	MYC	MYB	W – box	基因 Gene	LTRE	ABRE	MYC	MYB	W – box
<i>RcLEA1</i> – 1	0	0	4	0	6	<i>RcLEA5</i> – 1	0	1	20	1	5
<i>RcLEA1</i> – 2	0	2	16	2	6	<i>RcLEA5</i> – 2	0	4	10	2	6
<i>RcLEA2</i> – 1	0	3	10	1	6	<i>RcLEA6</i> – 1	0	2	4	1	6
<i>RcLEA2</i> – 2	0	0	10	2	3	<i>RcLEA6</i> – 2	0	0	12	0	3
<i>RcLEA3</i> – 1	0	0	4	1	8	<i>RcLEA7</i> – 1	1	3	12	1	4
<i>RcLEA3</i> – 2	1	5	10	0	4	<i>RcLEA7</i> – 2	4	6	10	0	7
<i>RcLEA3</i> – 3	0	1	4	3	3	<i>RcLEA7</i> – 3	0	4	14	1	6
<i>RcLEA4</i> – 1	0	4	2	1	2	<i>RcLEA7</i> – 4	0	4	4	1	1
<i>RcLEA4</i> – 2	1	1	8	2	7	<i>RcLEA7</i> – 5	0	4	4	1	1
<i>RcLEA4</i> – 3	1	6	18	1	3	<i>RcLEA8</i> – 1	0	3	6	0	4
<i>RcLEA4</i> – 4	1	2	8	0	4	<i>RcLEA8</i> – 2	1	3	8	1	7
<i>RcLEA4</i> – 5	3	3	16	2	7	<i>RcLEA8</i> – 3	0	2	20	1	3
<i>RcLEA4</i> – 6	1	5	14	0	5	<i>RcLEA8</i> – 4	0	2	12	0	2
<i>RcLEA4</i> – 7	0	4	2	3	2						

从各亚家族的成员数量来看,不同物种中亚族的成员组成并无明显的规律,例如蓖麻 *LEA*_4 亚族成员数量最多(约占 25.9%),dehydrin 亚族次之(约占 18.5%),这与拟南芥类似^[9,10],然而,水稻中成员数量最多的为 dehydrin 亚族^[8],大豆中成员数量最多的为 *LEA*_3 亚族^[12],梅花中成员数量最多的为 dehydrin 和 *LEA*_2 亚族^[11],这表明 *LEA* 类基因在这些物种的分化后又出现了独立的进化。就 *LEA* 基因的起源而言,纵观各亚家族在不同物种中分布,*LEA*_4 亚家族基因分布很广,基本覆盖了低等动物、细菌、真菌和各种高低等植物,显示出较早的起源;*LEA*_5 亚家族主要存在于植物和细菌中;*LEA*_1、*LEA*_2、*LEA*_3、*LEA*_6、dehydrin 和 SMP 集中于植物中^[3]。蓖麻中,除 *LEA*_4、dehydrin 和 SMP 亚族外,其它亚家族基因的扩增速度都比较慢。

这 27 个 *RcLEA* 基因至少在叶片、花、II/III 期胚乳、V/VI 期胚乳和种子等一种组织或发育时期中表达,虽然如此,这些基因中约有 37% 的基因在公共数据库中暂时还没有对应的 EST 序列,主要原因可能包括基因表达存在组织和发育时期特异性(如 *RcLEA6* – 2 可能只在花组织中表达,*RcLEA7* – 5 在 V/VI 期胚乳中的表达水平约为 II/III 期胚乳的 8 倍)、基因表达丰度太低而测序深度又不够(截至到 2013 年 1 月,NCBI 中释放了 62 592 条 EST,涉及根、叶、韧皮部、花、胚乳和种子^[20~22],基本囊括了上述表达谱组织)等。

蓖麻 *LEA* 基因的启动子区都富含 LTRE (33.3%)、ABRE (85.2%)、MYC (100%)、MYB

(70.4%) 和 W – box (100%) 等逆境应答相关顺式作用元件。ABRE、MYC 和 MYB 是逆境激素 ABA 的主要调控元件^[23,24],在拟南芥中,82% 的 *LEA* 基因启动子含有 ABRE 元件,其中约有 76.2% 可被 ABA 显著诱导^[10];在梅花中,分别有 66.7%、50% 和 83.3% 的 *LEA* 基因启动子含有 ABRE、MYC、MYB 元件,约有 63.3% 的 *LEA* 基因受 ABA 诱导^[11]。LTRE 及 CRT (C – repeat)/DRE (drought – responsive) (RCCGAC) 是低温和脱水(如干旱、高盐等)响应的重要元件,在拟南芥中,有 69% 的 *LEA* 基因含有 LTRE 元件,其中约有 34.3% 可被低温和干旱显著诱导^[10]。此外,W – box 是 WRKY 转录因子的结合元件,而 WRKY 是植物参与各种生物与非生物胁迫应答的主要转录因子^[25]。这表明,LEAP 在蓖麻的耐胁迫中起着非常重要的作用,有深入研究的价值。

参考文献:

[1] Dure III L, Greenway S C, Galau G A. Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by *in vitro* and *in vivo* protein synthesis[J]. Biochemistry,1981,20(14):4 162 – 4 168.

[2] Galau G A, Hughes D W, Dure III L. Absciscic acid induction of cloned cotton late embryogenesis – abundant (*Lea*) mRNAs [J]. Plant Mol Biol,1986,7(3):155 – 170.

[3] Tunnacliffe A, Wise M J. The continuing conundrum of the LEAP[J]. Naturwissenschaften,2007,94(10):791 –

- 812.
- [4] 李 剑,赵常玉,张富生,等. LEA 蛋白与植物抗逆性[J]. 植物生理学通讯,2010,46(11):1 101 – 1 108.
- [5] 刘 洋,邢 鑫,李德全. LEA 蛋白的分类与功能研究进展[J]. 生物技术通报,2011,(8):36 – 43.
- [6] 周 丽,姬向楠,何 非,等. 植物胚胎发育晚期丰富蛋白的结构与功能[J]. 热带生物学报,2012,3(2):191 – 196.
- [7] Hand S C, Menze M A, Toner M, et al. LEA proteins during water stress: not just for plants anymore[J]. Annu Rev Physiol,2011,73:115 – 134.
- [8] 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 Sci,2007,172:414 – 420.
- [9] Bies – Ethève N, Gaubier – Comella P, Debures A, et al. Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*[J]. Plant Mol Biol,2008,67(1 – 2):107 – 124.
- [10] Hundertmark M, Hinch D K. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*[J]. BMC Genomics,2008,9:118.
- [11] Du D, Zhang Q, Cheng T, et al. Genome – wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*[J]. Mol Biol Rep,2013,40(2):1 937 – 1 946.
- [12] 李 乐,许红亮,杨兴露,等. 大豆 *LEA* 基因家族全基因组鉴定、分类和表达[J]. 中国农业科学,2011,44(19):3 945 – 3 954.
- [13] Hunault G, Jaspard E. LEAPdb: a database for the late embryogenesis abundant proteins [J]. BMC Genomics, 2010,11:221.
- [14] Chan A P, Crabtree J, Zhao Q, et al. Draft genome sequence of the oilseed species *Ricinus communis*[J]. Nat Biotechnol,2010,28(9):951 – 956.
- [15] Altschul S F, Madden T L, Schäffer A A, et al. Gapped BLAST and PSI – BLAST: a new generation of protein database search programs [J]. Nucleic Acids Res,1997,25(17):3 389 – 3 402.
- [16] Yang B, Jiang Y, Rahman M H, et al. Identification and expression analysis of WRKY transcription factor genes in canola (*Brassica napus* L.) in response to fungal pathogens and hormone treatments [J]. BMC Plant Biol,2009,9:68.
- [17] Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position – specific gap penalties and weight matrix choice [J]. Nucleic Acids Res,1994,22(22):4 673 – 4 680.
- [18] Tamura K, Dudley J, Nei M, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4. 0 [J]. Mol Biol Evol,2007,24(8):1 596 – 1 599.
- [19] Kent W J, Kent W J. BLAT – the BLAST – like alignment tool [J]. Genome Res,2002,12(4):656 – 664.
- [20] Van de Loo F J, Broun P, Turner S, et al. An oleate 12 – hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog [J]. Proc Natl Acad Sci USA,1995,92:6 743 – 6 747.
- [21] Doering – Saad C, Newbury H J, Couldridge C E, et al. A phloem – enriched cDNA library from *Ricinus*: insights into phloem function [J]. J Exp Bot,2006,57:3 183 – 3 193.
- [22] Lu C F, Wallis J G, Browse J. An analysis of expressed sequence tags of developing castor endosperm using a full – length cDNA library [J]. BMC Plant Biol,2007,7:42.
- [23] Bartels D, Sunkar R. Drought and salt tolerance in plants [J]. Crit Rev Plant Sci,2005,24:23 – 58.
- [24] Yamaguchi – Shinozaki K, Shinozaki K. Organization of cis – acting regulatory elements in osmotic – and cold – stress – responsive promoters [J]. Trends Plant Sci, 2005,10:88 – 94.
- [25] Chen L, Song Y, Li S, et al. The role of WRKY transcription factors in plant abiotic stresses [J]. Biochim Biophys Acta,2012,1 819(2):120 – 128.

(责任编辑:王丽芳)