

中南林业科技大学学报

Journal of Central South University of Forestry & Technology

ISSN 1673-923X，CN 43-1470/S

《中南林业科技大学学报》2026年第6期

Doi:10.14067/j.cnki.1673-923X.2026.06.018

论文标题： 新疆野苹果LEA基因家族鉴定与表达分析

作　　者： 房 震，李 静，李美芸，等

收稿日期： 2025-06-03

引文格式： 房震,李静,李美芸,等.新疆野苹果*LEA*基因家族鉴定与表达分析[J].中南林业科技大学学报,2026,46(6):184-199.

FANG Z, LI J, LI M Y, et al. Identification and expression analysis of *LEA* gene family in *Malus sieversii*[J]. Journal of Central South University of Forestry & Technology,2026,46(6):184-199.

相关论文推荐

Recommended articles

巨桉脱水素基因*EgrDHN*家族鉴定及表达模式分析

Identification and expression pattern analysis of *EgrDHN* gene family in *Eucalyptus grandis*

巨桉*FLA*基因家族成员的鉴定与分析

Identification and analysis of *FLA* gene family members in *Eucalyptus grandis*

宁夏枸杞*LbDXS*基因家族全基因组鉴定及表达分析

Genome-wide identification and expression analysis of *LbDXS* gene family in *Lycium barbarum*

山茶*WRKY*基因家族鉴定及其在花色形成中的表达分析

Identification and expression analysis of *WRKY* gene family in flower color formation in *Camellia japonica*

油茶*bHLH*基因家族鉴定与响应干旱胁迫的基因表达分析

Identification of *bHLH* gene family in *Camellia oleifera* and analysis of its expression under drought stress

油茶*WRKY*基因家族鉴定及逆境胁迫表达分析

Identification and expression analysis of the *WRKY* gene family in *Camellia oleifera* under stress conditions

第 46 卷 第 6 期
2026 年 6 月
中南林业科技大学学报
Journal of Central South University of Forestry & Technology
Vol. 46 No. 6
Jun. 2026
Doi:10.14067/j.cnki.1673-923x.2026.06.018
http: //qks.csuft.edu.cn

新疆野苹果 *LEA* 基因家族鉴定与表达分析

房　震，李　静，李美芸，旷莹香，叶春秀
（新疆农业大学 林学与风景园林学院，新疆 乌鲁木齐 830052）

摘　要：【目的】全面解析晚期胚胎发生丰富蛋白在高等植物胚胎发育和非生物胁迫反应中的调控机制。【方法】利用隐马尔可夫模型和生物信息学方法，从新疆野苹果全基因组中筛选鉴定 *LEA* 基因家族，对其序列特征、保守基序、基因结构、顺式作用元件和表达模式等进行分析。【结果】新疆野苹果全基因组中共鉴定出 119 个 *LEA* 基因。根据它们的保守结构域和系统发育关系，共分为 8 个不同的组（LEA1-LEA6，DHN，SMP），其中 MsiLEA2 组的基因最丰富。顺式作用元件的分析结果表明，大多数 *MsiLEA* 基因与光响应元件、非生物胁迫响应相关。此外，一些基因与生物合成途径有关，如脱落酸响应、生长激素响应、茉莉酸甲酯响应和分生组织表达等。新疆野苹果 *LEA* 基因在种子低温层积的不同时期和不同器官中的表达模式分析显示，*MsiDHN3* 在种子低温层积过程中表达量下降，在所有器官中表达量较高，表明 *MsiDHN3* 在种子低温层积过程中发挥作用。【结论】本研究为新疆野苹果 *MsiLEA* 基因家族的形成提供了新见解，同时也加深了对这些基因在新疆野苹果种子胚胎发育阶段和非生物胁迫响应过程中的潜在作用的理解。这些结果对今后 *MsiLEA* 基因家族的功能研究具有一定的指导意义，并为探索 *MsiLEA* 基因在新疆野苹果种子胚胎发育阶段的作用提供理论依据。

关键词：新疆野苹果；*LEA* 基因家族；全基因组鉴定；表达分析
中图分类号：S727.3；S661.1　　**文献标志码：**A　　**文章编号：**1673-923X(2026)06-0184-16

Identification and expression analysis of *LEA* gene family in *Malus sieversii*

FANG Zhen, LI Jing, LI Meiyun, KUANG Yingxiang, YE Chunxiu

(College of Forestry and Landscape Architecture, Xinjiang Agricultural University, Urumqi 830052, Xinjiang, China)

Abstract:【Objective】To comprehensively elucidate the regulatory mechanisms of late embryogenesis abundant (LEA) proteins in higher plant embryogenesis and abiotic stress responses.【Method】Hidden Markov model and bioinformatics were used to screen and identify *LEA* genes family from the whole genome of *Malus sieversii*, and their sequence features, conserved motifs, gene structures, cis-acting elements, and expression patterns were analyzed.【Result】The 119 *LEA* genes were identified in the whole genome of *Malus sieversii*. According to their conserved structural domains and phylogenetic relationships, they were categorized into eight different groups (LEA1-LEA6, DHN, SMP), among which the MsiLEA2 group was the most gene-rich. Analysis of the cis-acting elements showed that most of the *MsiLEA* genes were associated with light response elements, abiotic stress response. In addition, some genes were related to biosynthetic pathways, such as abscisic acid response, growth hormone response, methyl jasmonate response, and meristematic tissue expression. The analysis of the expression pattern of *Malus sieversii LEA* genes in different periods of seed low-temperature stratification and in different organs showed that the expression of *MsiDHN3* decreased during seed low-temperature stratification and was higher in all organs, indicating that *MsiDHN3* plays a role in the low-temperature stratification of *Malus sieversii* seeds.【Conclusion】This study provides new insights into the formation of the *MsiLEA* genes family in *Malus sieversii*, it also deepens the understanding of the potential roles of these genes during embryonic developmental stages and abiotic stress response in *Malus sieversii* seeds. These results are instructive for future functional studies of the *MsiLEA* genes family and provide a theoretical basis for exploring the role of *MsiLEA* genes in the embryonic developmental stages of *Malus sieversii* seeds.

Keywords: *Malus sieversii*; *LEA* gene family; genome-wide identification; expression analysis

收稿日期（Received）：2025-06-03　　**修订日期**（Revised）：2025-08-21
基金项目：新疆维吾尔自治区自然科学基金面上项目（2023D01A28）。
第一作者：房震（fz199971@163.com），硕士研究生。**通信作者：**叶春秀（yecx2008@163.com），副教授，博士，硕士研究生导师。
引文格式：房震 , 李静 , 李美芸 , 等 . 新疆野苹果 *LEA* 基因家族鉴定与表达分析 [J]. 中南林业科技大学学报 ,2026,46(6):184-199.
FANG Z, LI J, LI M Y, et al. Identification and expression analysis of *LEA* gene family in *Malus sieversii*[J]. Journal of Central South University of Forestry & Technology,2026,46(6):184-199.
* 本文为第二十八届中国科协年会学术论文成果。
Academic Papers of the 28th Annual Meeting of the China Association for Science and Technology.

晚期胚胎发育丰富蛋白（Late embryogenesis abundant，LEA）是30多年前在棉花*Gossypium hirsutum*种子胚胎发育后期首次发现的[1]。随后，这类蛋白质也在植物的根、茎、叶、花和其他组织中被发现。根据氨基酸序列和保守基序的相似性，并在Pfam数据库中检索，将LEA蛋白分为至少8个亚家族，分别是LEA1（PF03760）、LEA2（PF03168）、LEA3（PF03242）、LEA4（PF02987）、LEA5（PF00477）、LEA6（PF10714）、脱水蛋白（Dehydrin，DHN）（PF00257）和种子成熟蛋白（Seed maturation protein，SMP）（PF04927）[2]。LEA蛋白广泛分布于高等植物中[3]，迄今为止，仅在少数测序的植物物种基因组中进行LEA家族基因全基因组鉴定和分析。例如拟南芥*Arabidopsis thaliana*[4]、水稻*Oryza sativa*[5]、玉米*Zea mays*[6]、马铃薯*Solanum tuberosum*[7]、黄瓜*Cucumis sativus*[8]、番茄*Solanum lycopersicum*[9]、甘蓝型油菜*Brassica napus*[10]、木薯*Manihot esculenta* Crantz[11]、毛果杨*Populus trichocarpa*[12]、梅*Prunus mume*[13]，油松*Pinus tabuliformis*[14]和海刀豆*Canavalia rosea*[15]，亚麻*Linum usitatissimum*[16]、花生*Arachis hypogaea*[17]、铁皮石斛*Dendrobium officinale*[18]、油菜*Brassica campestris*[19]、毛竹*Phyllostachys edulis*[20]、西瓜*Citrullus lanatus*、甜瓜*Cucumis melo*[21]、杂交枫香*Liquidambar styraciflua* *Liquidambar formosana*[22]，枳*Poncirus trifoliata*[23]。除植物外，藻类、真菌、细菌[24]以及无脊椎动物中也有相关研究[25]。

由于大量带电氨基酸残基的存在，大部分LEA蛋白在自然形式下是无序的，其主要特征是其显著的亲水性。这些残基，连同甘氨酸或其他小氨基酸（如丙氨酸、丝氨酸和苏氨酸），明显缺乏半胱氨酸和色氨酸，在某些情况下，含量极低[26-27]。*LEA*基因被认为是植物生长发育过程中的关键基因，在缓解各种逆境胁迫的不利影响方面发挥着重要作用。一项研究揭示*LEA*基因的动态本质，揭示它们在种子成熟的最后阶段表现出一种有趣的积累模式，这与脱水的开始相吻合。具体地说，这些基因在这一阶段以高浓度存在，表明它们在保护发育中的种子免受环境干扰方面具有潜在的作用[20]。研究结果发现，*LEA*基因具有超强的适应能力，无芒隐子草*Cleistogenes songorica*在非生物胁迫条件下（炎热和干旱胁迫）被显著诱导[28]。一些科学研究提出*LEA*基因可能潜在地涉及对过多有害的非生物应激的防御机制。这些机制包括但不限于强化和维持膜的结构完整性[29-30]、保护细胞内蛋白和酶免受水分胁迫[31]。

新疆野苹果*Malus sieversii*是古老的第三纪孑遗物种，是现代栽培苹果*Malus domestica* cv. ‘Gala’的祖先种，集中分布在新疆伊犁河谷两侧的山地[32]，具有抗寒、抗旱、抗病虫和耐瘠薄等优良特性[33]。苹果*MdLEA*基因在干旱、高盐和低温胁迫处理（1、6、10、24和48 h）后，*MdLEA60*对3种胁迫均表现出一致的正响应，说明干旱、高盐和低温胁迫对*MdLEA*基因表达具有显著的诱导作用[34]。栽培苹果与2个主要野生祖先 新疆野苹果和欧洲野苹果*Malus sylvestris*全基因组测序组装已公开[35]，为研究新疆野苹果基因家族的系统发育进化和功能分析奠定了基础。

本研究利用隐马尔可夫模型（Hidden Markov model，HMM）对新疆野苹果*LEA*基因家族进行全基因组鉴定，并对其序列特征、系统发育关系、保守基序、基因结构等进行分析。根据转录组数据研究新疆野苹果*LEA*基因在种子不同低温沙藏层积时期的表达模式，同时分析其在种子、根、茎、叶、花和果实器官的表达水平，为新疆野苹果*LEA*基因家族研究提供新思路，为进一步研究种子胚胎发育相关基因的功能和机制奠定基础。

1 材料与方法

1.1 试材及取样

于2023年10月采集新疆塔城地区额敏县野果林（80°47′～83°58′E，43°20′～46°21′N）的新疆野苹果种子。对照组为室温保存0 d的种子，标记为CK0；处理组为室温清水浸泡24 h的种子，标记为M0；4 ℃沙藏层积4个时期（30、60、90、120 d）的种子样品，分别标记为M30、M60、M90和M120。各时期的种子均做3次生物学重复。样品迅速于液氮中冷冻，之后在-80 ℃超低温冰箱中存放。

1.2 新疆野苹果*LEA*基因的鉴定

从Pfam数据库（http://pfam.xfam.org/）下载LEA亚家族PF03760（LEA 1）、PF03168（LEA 2）、PF03242（LEA 3）、PF02987（LEA 4）、PF00477（LEA 5）、PF10714（LEA 6）、PF00257（DHN）和PF04927（SMP）的隐马尔可夫模型。新疆野苹果基因组数据来自苹果数据库（http://bioinfo.bti.cornell.edu/apple_genome/），使用TBtools[36]进行HMM搜索新疆野苹果*LEA*基因，通过Pfam数据库和NCBI保守结构域数据库[37]（www. ncbi.

nlm.nih.gov/Structure/cdd/wrpsb.cgi）以验证和确定 *LEA* 基因内保守结构域的存在，并去除缺乏保守结构域和完整 CDS 序列的 *LEA* 基因。

1.3 新疆野苹果 LEA 蛋白的理化性质和系统进化分析

使用 ProtParam[38]（http://web.expasy.org/）预测分析分子量（Molecular weight，MW）、等电点（Isoelectric point，pI）、不稳定指数、脂肪族指数和总平均疏水性（Grand average of hydropathicity，GRAVY）。此外，利用 PSORT 预测工具[39]（http://www.genscript.com/wolf-psort.html）阐明 LEA 蛋白的亚细胞定位。拟南芥全基因组数据从 Ensemble Plants 数据库（http://plants.ensemble.org/index.html）获取，*LEA* 基因家族成员获取同上述方法，使用 MEGA 11.0.13[40] 软件通过邻接法构建系统进化树，进行 LEA 蛋白序列的 ClustalW 比对（设置 boot-strap replicates=1 000）。系统进化树在 ChiPlot[41] 网站（https://www.chiplot.online/#）美化。

1.4 LEA 家族蛋白保守基序与基因结构分析

由 MEME（https://meme-suite.org/meme/）的经典模式进行保守基序分析，其中基序数量设置为 40 个，其他设置与默认参数一致。利用在线软件 GSDS 2.0[42]（http://gsds.gao-lab.org/）分析 *LEA* 基因结构。

1.5 *LEA* 基因共线性和顺式作用元件分析

将来自 *LEA* 家族基因的 DNA 序列定位到新疆野苹果基因组上，计算 *LEA* 基因重复事件，并使用 TBtools 软件进行可视化作图。使用 TBtools 软件的 Fasta Extract 工具提取新疆野苹果 *LEA* 家族基因转录起始位点上游 2 000 bp 的基因组序列，PlantCARE 数据库（https://bioinformatics.psb.ugent.be/webtools/plantcare/html/）分析启动子区域顺式元件[43]，使用 R 语言计数与植物激素、植物生长发育和非生物与生物胁迫响应相关的顺式元件。

1.6 *LEA* 基因低温层积过程和器官特异性表达模式分析

结合新疆野苹果种子的转录组测序，获取 *LEA* 基因家族在不同低温层积时期的表达量。从 NCBI GEO 数据库（https://www.ncbi.nlm.nih.gov/geo/）下载器官特异性表达数据，登录号为 GSE42873[29]。数据库中选定的基因表达数据涵盖苹果的所有器官（花、果实、种子、根、茎和叶）。

2 结果与分析

2.1 新疆野苹果 *LEA* 基因家族成员鉴定及基本理化性质

通过对 LEA1（PF03760）、LEA2（PF03168）、LEA3（PF03242）、LEA4（PF02987）、LEA5（PF00477）、LEA6（PF10714）、SMP（PF04927）和 DHN（PF00257）的 HMM 搜索，在新疆野苹果全基因组中共鉴定出 119 个 LEA 基因（表 1），编码氨基酸长度在 68（MsiLEA4-8）～ 958（MsiLEA2-3）aa，分子量在 7 005.68 ～ 108 725.32 D 之间，等电点为 4.62（MsiSMP-5）～ 10.55（MsiLEA2-23），不稳定系数在 4.36（MsiLEA2-8）～ 64.90（MsiLEA2-64），脂肪族指数在 25.38（MsiLEA4-6）～ 116.95（MsiLEA2-37），亲水性在 -1.69（MsiLEA5-1）～ 0.34（MsiLEA2-65）。亚细胞定位结果表明，LEA 蛋白主要定位在细胞核、细胞质和叶绿体中。

2.2 新疆野苹果 LEA 家族成员进化分析

在新疆野苹果中共鉴定出 119 个 *MsiLEA* 家族基因成员，并选择拟南芥 *AtLEA* 基因家族 51 条蛋白序列构建系统进化树（图 1）。以拟南芥 *AtLEA* 基因家族分类为标准，将 *MsiLEA* 家族成员分 8 个组，分别是 LEA1、LEA 2、LEA3、LEA 4、LEA5、LEA6、DHN 和 SMP，包含基因数目分别是 5、78、3、11、5、2、8 和 7，其中 *MsiLEA* 家族成员不含一个拟南芥特有的 ATM 组，这与杂交枫香分组结果一致[22]。

2.3 新疆野苹果 LEA 家族保守基序分析

分析新疆野苹果 LEA 蛋白保守基序，119 个 LEA 的 40 个 Motif 保守基序序列和氨基酸序列（表 2），这些保守基序在同一组成员基本相同，不同组间存在一定差异（图 2）。Motif1 ～ Motif10 特异分布于 LEA2，Motif23 特异分布于 LEA4，Motif20 特异分布于 DHN，SMP 组主要包含 Motif11、Motif12、Motif16 和 Motif35。同一组的成员被认为具有相对相似的生物学功能。结果表明，不同 LEA 亚家族的保守结构域基序组成不同，但同一亚家族的 *LEA* 基因保守结构域基序组成极为相似，而且这些结构域基序也是保守的。

表 1　新疆野苹果 *MsiLEA* 基因家族成员理化性质和亚细胞定位预测 †

Table 1　Predicted physical and chemical properties and subcellular localization of *MsiLEA* gene family members in *Malus sieversii*

基因名称 Gene name	基因 ID Gene ID	家族 Family	Pfam ID	长度 Length / aa	分子量 Molecular weight /D	等电点 pI	不稳定指数 Instability index	脂肪族指数 Aliphatic index	亲水性 GRAVY	亚细胞定位 Subcellular localization
MsiLEA1-1	Msi_06g003110	LEA1	PF03760	130	14 302.07	9.37	36.16	55.00	-1.06	Nucl、Mito
MsiLEA1-2	Msi_12g005900	LEA1	PF03760	165	16 825.54	9.16	20.34	42.79	-0.73	Cyto、Chlo
MsiLEA1-3	Msi_02g019760	LEA1	PF03760	110	11 783.37	7.11	29.51	64.27	-0.65	Nucl、Cyto、Mito
MsiLEA1-4	Msi_07g007370	LEA1	PF03760	110	11 696.20	8.93	30.23	65.18	-0.76	Nucl、Cyto、Mito
MsiLEA1-5	Msi_14g006940	LEA1	PF03760	98	10 256.38	8.99	50.97	51.84	-0.77	Cyto、Mito
MsiLEA2-1	Msi_15g013730	LEA2	PF03168	319	35 005.91	4.78	21.34	96.21	-0.31	Cyto、Nucl
MsiLEA2-2	Msi_02g000110	LEA2	PF03168	361	39 934.80	4.78	22.61	102.77	-0.11	Cyto、PM
MsiLEA2-3	Msi_14g003370	LEA2	PF03168	958	108 725.32	9.56	34.52	85.94	-0.18	PM
MsiLEA2-4	Msi_07g011120	LEA2	PF03168	172	18 576.44	5.63	17.48	104.24	-0.11	Cyto
MsiLEA2-5	Msi_01g007180	LEA2	PF03168	177	19 193.25	6.15	16.88	105.20	-0.13	PM、Cyto、ER
MsiLEA2-6	Msi_14g000970	LEA2	PF03168	555	59 723.68	9.30	27.82	106.54	0.25	PM
MsiLEA2-7	Msi_07g011100	LEA2	PF03168	150	16 516.16	4.90	8.84	108.47	0.11	Nucl、Cyto
MsiLEA2-8	Msi_01g007200	LEA2	PF03168	150	16 511.15	5.31	4.36	105.20	0.06	Cyto
MsiLEA2-9	Msi_05g032200	LEA2	PF03168	398	45 530.41	10.02	41.26	82.81	-0.38	Cyto、Nucl
MsiLEA2-10	Msi_10g028740	LEA2	PF03168	397	45 831.93	9.91	44.80	90.65	-0.33	Nucl、Cyto
MsiLEA2-11	Msi_07g022510	LEA2	PF03168	447	49 018.14	9.87	37.18	96.55	-0.07	PM
MsiLEA2-12	Msi_06g009160	LEA2	PF03168	487	54 462.15	10.28	28.87	90.23	-0.09	Chlo
MsiLEA2-13	Msi_06g009030	LEA2	PF03168	422	45 733.64	9.96	24.77	96.59	-0.02	PM
MsiLEA2-14	Msi_06g009200	LEA2	PF03168	559	61 206.08	9.95	24.47	98.32	0.00	PM
MsiLEA2-15	Msi_12g016780	LEA2	PF03168	338	36 293.49	9.82	20.52	99.70	0.15	Mito、Cyto
MsiLEA2-16	Msi_14g012830	LEA2	PF03168	423	46 529.38	10.09	38.02	89.20	-0.14	PM
MsiLEA2-17	Msi_11g013610	LEA2	PF03168	483	53 262.41	9.78	39.38	93.56	0.07	PM、ER
MsiLEA2-18	Msi_06g009190	LEA2	PF03168	440	47 878.23	9.91	25.92	95.05	-0.01	PM
MsiLEA2-19	Msi_06g009300	LEA2	PF03168	403	43 744.34	9.87	34.80	90.45	0.04	PM
MsiLEA2-20	Msi_04g016060	LEA2	PF03168	339	36 085.10	8.97	22.86	104.28	0.32	Cyto、Vacu
MsiLEA2-21	Msi_01g018790	LEA2	PF03168	313	34 239.55	9.70	28.61	97.38	-0.09	Cyto、ER
MsiLEA2-22	Msi_03g012810	LEA2	PF03168	181	19 642.78	9.30	33.29	102.93	0.10	Chlo、Mito
MsiLEA2-23	Msi_09g017050	LEA2	PF03168	276	30 759.12	10.55	42.20	88.01	-0.21	Cyto
MsiLEA2-24	Msi_12g003750	LEA2	PF03168	228	25 913.09	9.51	32.65	88.46	-0.16	Chlo
MsiLEA2-25	Msi_12g002330	LEA2	PF03168	446	50 056.57	9.05	17.86	84.10	-0.09	Vacu
MsiLEA2-26	Msi_14g012770	LEA2	PF03168	534	59 137.93	9.84	30.03	88.35	-0.15	Chlo、Cyto
MsiLEA2-27	Msi_02g004590	LEA2	PF03168	234	26 155.07	7.64	29.18	104.06	0.13	Chlo、Mito
MsiLEA2-28	Msi_15g037450	LEA2	PF03168	267	28 681.71	9.53	38.21	85.43	-0.16	Chlo
MsiLEA2-29	Msi_08g020580	LEA2	PF03168	258	27 785.69	9.41	35.77	87.64	-0.09	Chlo
MsiLEA2-30	Msi_14g004720	LEA2	PF03168	412	46 200.50	7.00	32.85	92.40	0.06	Cyto
MsiLEA2-31	Msi_15g017550	LEA2	PF03168	306	34 556.51	8.82	35.51	91.37	-0.16	Chlo、Vacu
MsiLEA2-32	Msi_12g003760	LEA2	PF03168	233	26 334.67	9.72	41.74	100.26	0.05	Chlo、Mito
MsiLEA2-33	Msi_17g017530	LEA2	PF03168	219	24 570.37	6.42	48.06	95.57	0.08	Chlo
MsiLEA2-34	Msi_04g016050	LEA2	PF03168	202	22 390.44	9.93	41.66	109.55	0.10	Cyto、Mito
MsiLEA2-35	Msi_09g005610	LEA2	PF03168	386	42 399.97	4.92	42.85	81.40	-0.01	Chlo、Cyto
MsiLEA2-36	Msi_14g004700	LEA2	PF03168	563	62 101.39	9.32	44.58	63.57	-0.36	Nucl
MsiLEA2-37	Msi_12g016770	LEA2	PF03168	200	22 007.07	10.33	36.34	116.95	0.21	Cyto、Nucl
MsiLEA2-38	Msi_09g008290	LEA2	PF03168	190	20 978.55	9.80	12.41	94.74	-0.07	Chlo
MsiLEA2-39	Msi_06g020900	LEA2	PF03168	578	62 302.39	8.95	34.77	96.44	0.03	Cyto、Nucl、ER
MsiLEA2-40	Msi_12g003820	LEA2	PF03168	544	60 478.28	8.88	38.81	91.14	0.04	PM、Vacu
MsiLEA2-41	Msi_08g005650	LEA2	PF03168	271	29 598.09	10.13	39.60	83.91	-0.16	Chlo、Mito
MsiLEA2-42	Msi_15g005820	LEA2	PF03168	264	29 150.75	10.35	28.67	80.91	-0.23	Chlo、Mito

续表 1
Continuation of table 1

基因名称 Gene name	基因 ID Gene ID	家族 Family	Pfam ID	长度 Length / aa	分子量 Molecular weight /D	等电点 pI	不稳定指数 Instability index	脂肪族指数 Aliphatic index	亲水性 GRAVY	亚细胞定位 Subcellular localization
MsiLEA2-43	Msi_14g012760	LEA2	PF03168	213	23 387.44	9.82	31.26	95.63	0.04	Nucl
MsiLEA2-44	Msi_01g007480	LEA2	PF03168	251	28 139.31	9.91	45.14	90.80	-0.12	Chlo、Cyto
MsiLEA2-45	Msi_06g010370	LEA2	PF03168	213	23 176.62	9.51	42.62	72.25	0.04	Chlo、Mito
MsiLEA2-46	Msi_07g012320	LEA2	PF03168	251	28 234.34	10.03	44.67	86.85	-0.18	Chlo、Cyto
MsiLEA2-47	Msi_12g003770	LEA2	PF03168	405	44 351.85	9.33	37.97	83.09	-0.13	Cyto、ER
MsiLEA2-48	Msi_01g003080	LEA2	PF03168	266	29 215.93	9.35	35.67	81.65	-0.49	Cyto、Nucl
MsiLEA2-49	Msi_06g009240	LEA2	PF03168	286	31 494.88	9.76	39.41	108.71	0.23	Chlo
MsiLEA2-50	Msi_02g002000	LEA2	PF03168	205	22 697.38	9.70	30.44	87.90	-0.24	Nucl
MsiLEA2-51	Msi_06g009220	LEA2	PF03168	176	19 169.28	9.52	28.79	90.80	-0.08	Nucl、Chlo
MsiLEA2-52	Msi_06g009210	LEA2	PF03168	210	22 800.84	9.85	22.97	94.71	-0.01	Chlo、PM、ER
MsiLEA2-53	Msi_17g008780	LEA2	PF03168	192	21 086.83	9.61	26.04	104.06	0.28	Nucl、Cyto
MsiLEA2-54	Msi_06g009180	LEA2	PF03168	225	24 558.77	9.83	21.68	97.47	0.02	Chlo
MsiLEA2-55	Msi_02g004300	LEA2	PF03168	305	33 670.53	9.93	64.06	61.61	-0.42	Chlo、Cyto
MsiLEA2-56	Msi_06g009280	LEA2	PF03168	210	22 921.85	9.80	28.91	92.38	-0.04	Chlo、Nucl
MsiLEA2-57	Msi_06g009230	LEA2	PF03168	210	22 948.96	9.89	28.37	93.29	-0.03	Chlo、Nucl
MsiLEA2-58	Msi_11g004390	LEA2	PF03168	884	97 131.13	9.15	44.24	92.48	-0.16	PM
MsiLEA2-59	Msi_05g014150	LEA2	PF03168	286	32 569.85	9.16	58.09	78.29	-0.40	Nucl、Vacu
MsiLEA2-60	Msi_06g009260	LEA2	PF03168	173	18 573.39	9.36	21.66	95.72	0.12	Cyto、Nucl
MsiLEA2-61	Msi_10g013460	LEA2	PF03168	306	34 655.13	9.25	56.05	79.22	-0.39	Chlo
MsiLEA2-62	Msi_14g004710	LEA2	PF03168	424	46 989.57	8.59	40.37	89.86	-0.03	PM
MsiLEA2-63	Msi_15g017330	LEA2	PF03168	306	33 465.24	9.97	60.16	65.52	-0.36	Chlo、Cyto
MsiLEA2-64	Msi_02g014780	LEA2	PF03168	317	34 749.50	10.22	64.90	71.01	-0.34	Cyto、ER
MsiLEA2-65	Msi_14g012820	LEA2	PF03168	747	82 449.39	9.76	36.82	116.09	0.35	PM、Nucl
MsiLEA2-66	Msi_09g005590	LEA2	PF03168	217	23 871.42	5.84	43.92	94.88	0.20	Chlo、Nucl
MsiLEA2-67	Msi_12g003800	LEA2	PF03168	397	43 965.85	8.58	35.87	94.46	0.00	PM、Chlo
MsiLEA2-68	Msi_14g013730	LEA2	PF03168	213	23 011.51	9.55	35.88	83.71	0.12	Chlo、Mito
MsiLEA2-69	Msi_09g005550	LEA2	PF03168	217	23 892.63	5.84	45.56	94.88	0.25	Chlo
MsiLEA2-70	Msi_09g005680	LEA2	PF03168	223	24 563.94	4.79	44.57	84.44	0.07	Chlo、Vacu
MsiLEA2-71	Msi_14g003450	LEA2	PF03168	209	23 739.65	10.02	37.84	86.75	-0.01	Chlo、Cyto、PM
MsiLEA2-72	Msi_03g018040	LEA2	PF03168	264	28 846.59	9.92	29.16	92.31	0.17	PM、Cyto、Chlo
MsiLEA2-73	Msi_17g013830	LEA2	PF03168	248	28 226.03	8.62	37.17	110.77	0.31	PM、Vacu、ER
MsiLEA2-74	Msi_09g005640	LEA2	PF03168	171	19 132.62	5.01	41.68	81.52	-0.13	Chlo
MsiLEA2-75	Msi_09g005630	LEA2	PF03168	224	24 886.16	4.71	37.95	81.92	-0.06	Chlo
MsiLEA2-76	Msi_15g007930	LEA2	PF03168	318	35 135.20	10.15	56.85	65.25	-0.42	Nucl、Chlo
MsiLEA2-77	Msi_06g009290	LEA2	PF03168	241	26 603.34	9.79	38.66	107.97	0.29	Chlo、Nucl
MsiLEA2-78	Msi_14g012790	LEA2	PF03168	190	21 345.32	10.09	28.67	102.58	0.01	Nucl
MsiLEA3-1	Msi_11g025570	LEA3	PF03242	102	10 480.81	10.07	29.59	71.76	-0.29	Cyto、Chlo
MsiLEA3-2	Msi_03g024800	LEA3	PF03242	102	10 605.95	10.23	38.75	65.98	-0.36	Cyto、Chlo
MsiLEA3-3	Msi_13g014410	LEA3	PF03242	97	10 330.78	10.08	40.64	69.69	-0.50	Mito、Chlo、Nucl
MsiLEA4-1	Msi_09g023710	LEA4	PF02987	531	58 137.13	5.57	24.20	46.20	-1.20	Mito、Nucl、Cyto
MsiLEA4-2	Msi_09g026200	LEA4	PF02987	501	55 024.67	5.39	26.00	46.99	-1.19	Mito、Cyto
MsiLEA4-3	Msi_14g011330	LEA4	PF02987	242	26 807.56	8.54	19.55	45.79	-1.22	Nucl、Mito、Cyto
MsiLEA4-4	Msi_14g011320	LEA4	PF02987	108	11 987.54	9.49	17.44	40.93	-1.24	Nucl、Mito
MsiLEA4-5	Msi_13g012260	LEA4	PF02987	436	47 206.89	5.73	25.98	45.85	-1.15	Chlo
MsiLEA4-6	Msi_17g018290	LEA4	PF02987	171	18 882.57	8.84	32.09	25.38	-1.52	Mito、Nucl
MsiLEA4-7	Msi_10g026720	LEA4	PF02987	453	51 036.24	8.55	39.76	55.08	-1.10	Chlo
MsiLEA4-8	Msi_09g007410	LEA4	PF02987	68	7 005.68	8.10	60.39	29.12	-1.04	Nucl
MsiLEA4-9	Msi_17g006500	LEA4	PF02987	85	9 021.17	9.40	40.45	47.41	-0.79	Mito、Nucl
MsiLEA4-10	Msi_17g006520	LEA4	PF02987	88	9 337.44	9.40	50.43	41.36	-0.87	Nucl、Mito
MsiLEA4-11	Msi_12g022090	LEA4	PF02987	136	14 531.22	9.60	39.17	57.65	-0.71	Chlo

续表 1
Continuation of table 1

基因名称 Gene name	基因 ID Gene ID	家族 Family	Pfam ID	长度 Length / aa	分子量 Molecular weight /D	等电点 pI	不稳定指数 Instability index	脂肪族指数 Aliphatic index	亲水性 GRAVY	亚细胞定位 Subcellular localization
MsiLEA5-1	Msi_12g015080	LEA5	PF00477	159	17 699.13	6.27	43.05	40.57	-1.69	Nucl、Mito
MsiLEA5-2	Msi_04g011530	LEA5	PF00477	117	13 044.20	5.05	53.74	51.71	-1.29	Nucl
MsiLEA5-3	Msi_04g014420	LEA5	PF00477	99	10 789.80	6.85	43.40	40.51	-1.51	Nucl、Mito、Cyto
MsiLEA5-4	Msi_04g014400	LEA5	PF00477	133	14 659.53	9.25	51.17	58.80	-0.91	Nucl、Mito
MsiLEA5-5	Msi_04g014410	LEA5	PF00477	70	7 721.38	8.26	43.62	36.29	-1.61	Nucl、Mito
MsiLEA6-1	Msi_02g015490	LEA6	PF10714	505	56 422.71	9.56	41.70	85.72	-0.37	Nucl
MsiLEA6-2	Msi_15g026630	LEA6	PF10714	156	17 476.75	8.91	45.99	58.14	-0.86	Chlo、Cyto、Mito
MsiDHN-1	Msi_02g012340	DHN	PF00257	565	59 916.04	8.89	12.26	32.35	-1.30	Nucl、Cyto
MsiDHN-2	Msi_02g012330	DHN	PF00257	380	39 821.92	8.57	20.49	30.89	-1.30	Nucl
MsiDHN-3	Msi_10g007490	DHN	PF00257	200	20 823.65	7.28	33.24	44.80	-1.12	Nucl
MsiDHN-4	Msi_10g013470	DHN	PF00257	221	24 511.02	7.03	38.93	40.23	-1.56	Nucl
MsiDHN-5	Msi_08g000440	DHN	PF00257	277	31 802.53	5.20	60.06	41.91	-1.68	Nucl
MsiDHN-6	Msi_15g000510	DHN	PF00257	284	32 816.69	5.28	54.97	42.25	-1.65	Nucl
MsiDHN-7	Msi_05g014160	DHN	PF00257	198	22 115.24	6.31	45.94	37.07	-1.63	Nucl
MsiDHN-8	Msi_15g023160	DHN	PF00257	331	35 382.76	5.27	32.66	44.26	-0.90	Nucl
MsiSMP-1	Msi_13g016660	SMP	PF04927	260	26 802.76	4.88	34.41	74.46	-0.33	Chlo、Cyto、Nucl
MsiSMP-2	Msi_12g020340	SMP	PF04927	291	29 976.45	5.26	34.57	76.94	-0.39	Chlo、Cyto
MsiSMP-3	Msi_04g019350	SMP	PF04927	267	27 370.56	5.27	33.44	76.93	-0.33	Chlo、Cyto
MsiSMP-4	Msi_13g002120	SMP	PF04927	267	27 370.56	5.27	33.44	76.93	-0.33	Chlo、Cyto
MsiSMP-5	Msi_03g022920	SMP	PF04927	259	26 833.68	4.62	39.59	75.83	-0.44	Chlo、Cyto、Nucl
MsiSMP-6	Msi_11g023520	SMP	PF04927	249	27 007.16	9.77	53.55	87.15	-0.33	Chlo、Cyto、Mito
MsiSMP-7	Msi_16g017360	SMP	PF04927	94	9 965.27	8.77	29.35	82.23	-0.28	Chlo、Mito、Cyto

† Cyto：细胞质；Nucl：细胞核；Chlo：叶绿体；Mito：线粒体；Vacu：液泡；ER：内质网；PM：细胞质膜。
Cyto: cytoplasm; Nucl: nucleus; Chlo: chloroplast; Mito: mitochondrion; Vacu: vacuole; ER: endoplasmic reticulum; PM: plasma membrane.

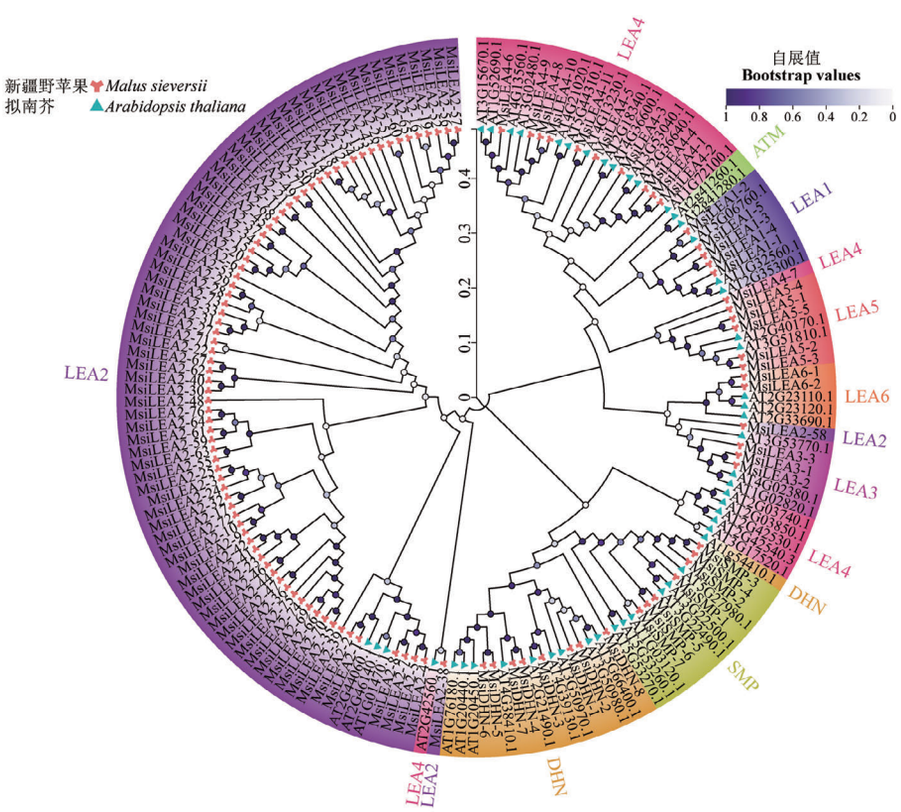


图 1　拟南芥和新疆野苹果 LEA 蛋白的系统发生树
Fig. 1　Phylogenetic tree of *Arabidopsis thaliana* and *Malus sieversii* LEA proteins

表 2　新疆野苹果 LEA 保守基序
Table 2　Motifs of LEA proteins in *Malus sieversii*

基序 Motif	序列 Sequence
1	TVMKVKTPKVRLGNINVQELNSVPATPSFDTKFTTQIRVKNTNFGPYKFD
2	TFLYQGATVGQVSIPKSKAGMLSTKKIBVEVSLSSSALSS
3	SNLGSELSSGVLTLNSAAKLTGKVELMLIMKKK
4	RSDAESLZSADELKRKKRIKLAIYIGIFIVFQIIVITVISL
5	SSTMDCTIAFDLSSKTLKSLQ
6	LILWLVLKPKKPKFRLDSVSV
7	LNATLDVTJTVRNPN
8	LLYRGTAVGQVSIPKFKQGAKSTRKVNVT
9	DDKESDSNLKLGTAPLPPIYVSKMNQTTLDFSVVMARSYVGTDVANDLII
10	CYGSGRLWLKLFGYFRYNSTLFALAGVQLSRLSCDLFPPNGTAGQSRACQ
11	KLPADKPATRRDAEGVTGAEMRNDPFLTTHPTGVAASVAAAARLNENN
12	AITIGEAAEETAQTAGDKPVEYSDAAAEQAAEVKATARQS
13	NLGSELSSGVJPLNSAAKVTGKVELMFIM
14	MAEKTQQAYPLAPAN
15	MASEQERRDPQKRQELDZKARRGETVVPGGTGGKSLEAQEHLAEGRHRGG
16	GGELADKPVAPKDAAIMQAAENAVLGQTLKGGPAAIMQSAATENE
17	NPYPHPLPICEIKYTLKSVNRVIASGELPDPGSJKASGVTVLEVPIKVPH
18	ZTRKEQIGZEGYQEMGKKGGLSTKDKSGGERAAEEGIEIDES
19	CICCCAFJIFLLIVI
20	GTTGESEEKKGILEKIKEKLPGHHKTEE
21	EEGKEEKGFMEKIKDKLPGTGSKDEEDA
22	PRSPRRPIYYVQSPSRDSHDG
23	AGEYTBYAAEKAKEAKDTTVNKAGEYTNYAAEKAKEAKDETVKTAKEAKD
24	PLNLTFVVRSRAYVLGKLVKSKFYRRISCSVT
25	KLDLYYENLQADIFY
26	PQHSYSHPSYGHHSRESSSSRFSGSLK
27	RGAVSFRLEMFAWVRFKTGWW
28	SIPALLFGIETCPTVAPWNKNVTMPASYLYGPKFVFLTLIWVVNLVSNEG
29	ASGLJPPFYQGHKNTTIFRSVLQGQQVLV
30	VVSTSSCDFRIDVGLTNYKPKDPITLVDALVLRNLEVCHHAPKLITMVMS
31	MGMAKSSSAGPLCKIVFFCFAALLAIFLF
32	VDKAKNYVAEKIANVPKPDAELTDVDFKKVSFDSVDYJAKV
33	KGKKKKKGLKDKIKEKISGNGEKEDEYH
34	KSSTMBCTIAFDLSS
35	MNQEQPRRPQPQQA
36	IGNITIPLNKKGEIKLPTLP
37	GVSVKEADLPGRRIISESMGGQVVGQYRQRAPLAAPNTIQP
38	EQEQKKHETLAZKLHRSGSSSSSSSSDEE
39	RHRMHANYEEMCEAKNFNCWALQFLAFLGLLALCLWLALRPKDPQYTIVD
40	RLDTGVKIRVGNFDSPKVGIRVTCEGII

2.4　新疆野苹果 *LEA* 基因家族共线性分析

对新疆野苹果基因组进行共线性分析，在 119 个 *MsiLEA* 基因中共检测到 107 个片段复制基因对（图 3），其中，chr2、4、7、11、12、14、15 染色体上的基因复制事件较多（＞ 10）。综合来说，新疆野苹果 *LEA* 基因家族扩张与基因复制有必然联系，而全基因组复制或片段重复事件在所有导致基因加倍的事件中占比较大。

2.5　新疆野苹果 LEA 基因家族结构分析

新疆野苹果 119 个 *LEA* 基因序列长度整体较短，介于 0 ～ 2 kb 之间（图 4），大多数 *LEA* 基因内含子较少。Msi_01g003080、Msi_02g002000、

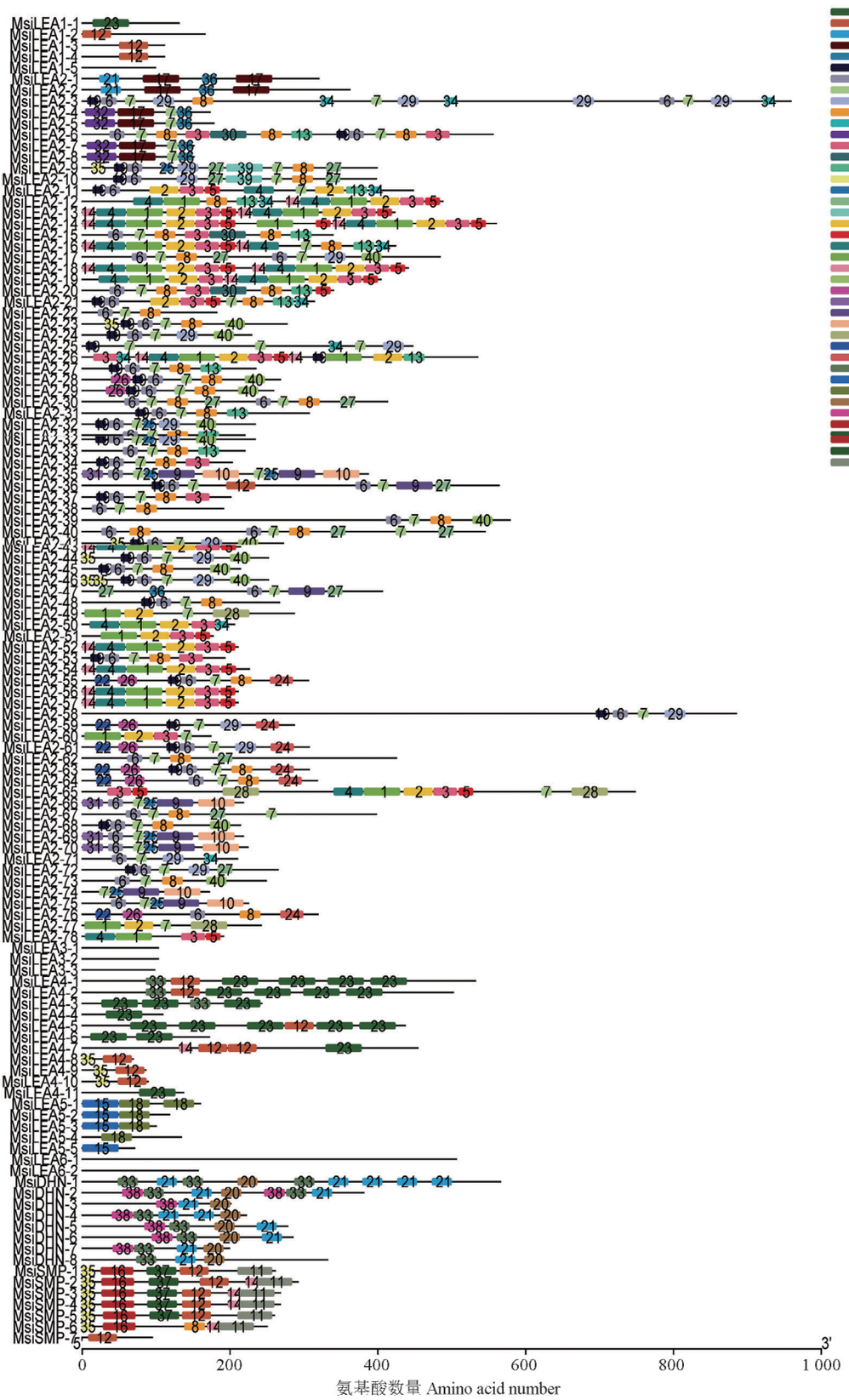


图 2　新疆野苹果 LEA 蛋白基序的分布

Fig. 2　Motif distributions of LEA proteins in *Malus sieversii*

Msi_04g016050、Msi_06g009210、Msi_06g009230、Msi_06g009280、Msi_06g010370、Msi_08g005650、Msi_09g005550、Msi_09g005590、Msi_09g005630、Msi_09g005680、Msi_12g003750、Msi_14g003450、Msi_15g005820 和 Msi_17g008780 无内含子，Msi_01g007180、Msi_01g007200、Msi_01g007480、Msi_02g019760、Msi_03g012810、Msi_03g018040、Msi_03g024800、Msi_04g011530、Msi_

04g014400、Msi_04g014410、Msi_04g014420、Msi_04g016060、Msi_05g014160、Msi_05g032200、Msi_06g003110、Msi_06g009030、Msi_06g009160、Msi_06g009180、Msi_06g009220、Msi_06g009260、Msi_06g009300、Msi_07g007370、Msi_07g011100、Msi_07g011120、Msi_07g012320、Msi_08g000440、Msi_08g020580、Msi_09g005610、Msi_09g005640、Msi_09g008290、Msi_09g017050、Msi_09g023710、Msi_09g026200、Msi_10g007490、Msi_10g013470、Msi_10g028740、Msi_11g025570、Msi_12g003760、Msi_12g003800、Msi_12g005900、Msi_12g015080、Msi_12g016770、Msi_12g016780、Msi_13g012260、Msi_13g014410、Msi_14g004720、Msi_14g006940、Msi_14g011320、Msi_14g011330、Msi_14g012760、Msi_14g012790、Msi_14g012830、Msi_14g013730、Msi_15g000510、Msi_15g013730、Msi_15g026630、Msi_15g037450、Msi_16g017360、Msi_17g013830、Msi_17g017530 和 Msi_17g018290 只含 1 个内含子，说明 *LEA* 基因结构保守性较高。值得注意的是，LEA2 亚家族中的 2 个基因 Msi_06g009200 和 Msi_14g003370，其长度最长，可达 23 kb，且其最长的内含子长度达到 10 kb。

2.6　新疆野苹果 *LEA* 基因家族启动子顺势作用元件分析

分析 *MsiLEA* 上游 2 000 bp 启动子序列中顺式作用元件分布情况（图 5），发现 119 个 *MsiLEA* 基因的启动子顺式作用元件均与植物生长发育、植物激素响应和非生物与生物胁迫有关。

启动子中预测到光响应相关元件（G-box、GT1-motif、TCT-motif 和 GATA-motif）。此外，少数基因含有与植物次生代谢和生长发育相关的顺式作用元件，涉及玉米蛋白代谢调控（O2-site）、负责昼夜节律控制的位点和种子特异性调控元件（RY-element）。存在激素响应元件，包括水杨酸响应元件（Salicylic acid response elements，TCA-element）、赤霉素响应元件（P-box 和 GARE-motif）、脱落酸响应元件（Abscitic acid responsiveness，ABRE）、生长素响应元件（Auxin，AuxRR-core, as-1, TGA-element）和茉莉酸甲酯响应元件（Methyl jasmonate，CGTCA-motif 和 TGACG-motif）。进一步研究发现，*MsiLEA* 基因的启动子中含有 ABRE、CGTCA-motif 和 TGACG-motif，分别能被脱落酸和茉莉酸甲酯激

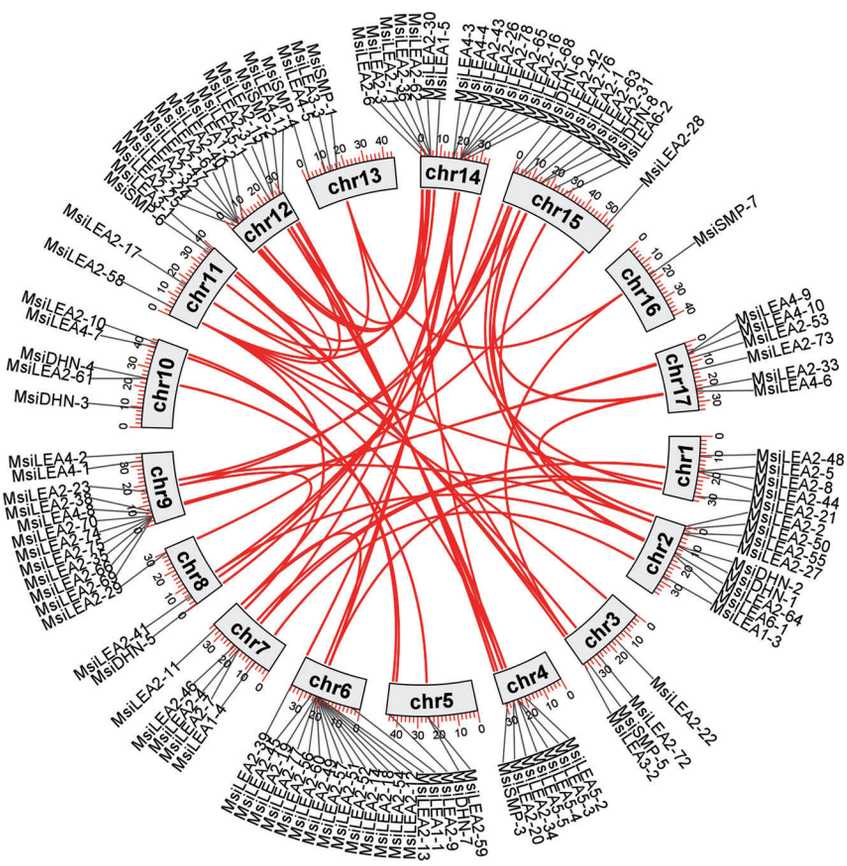
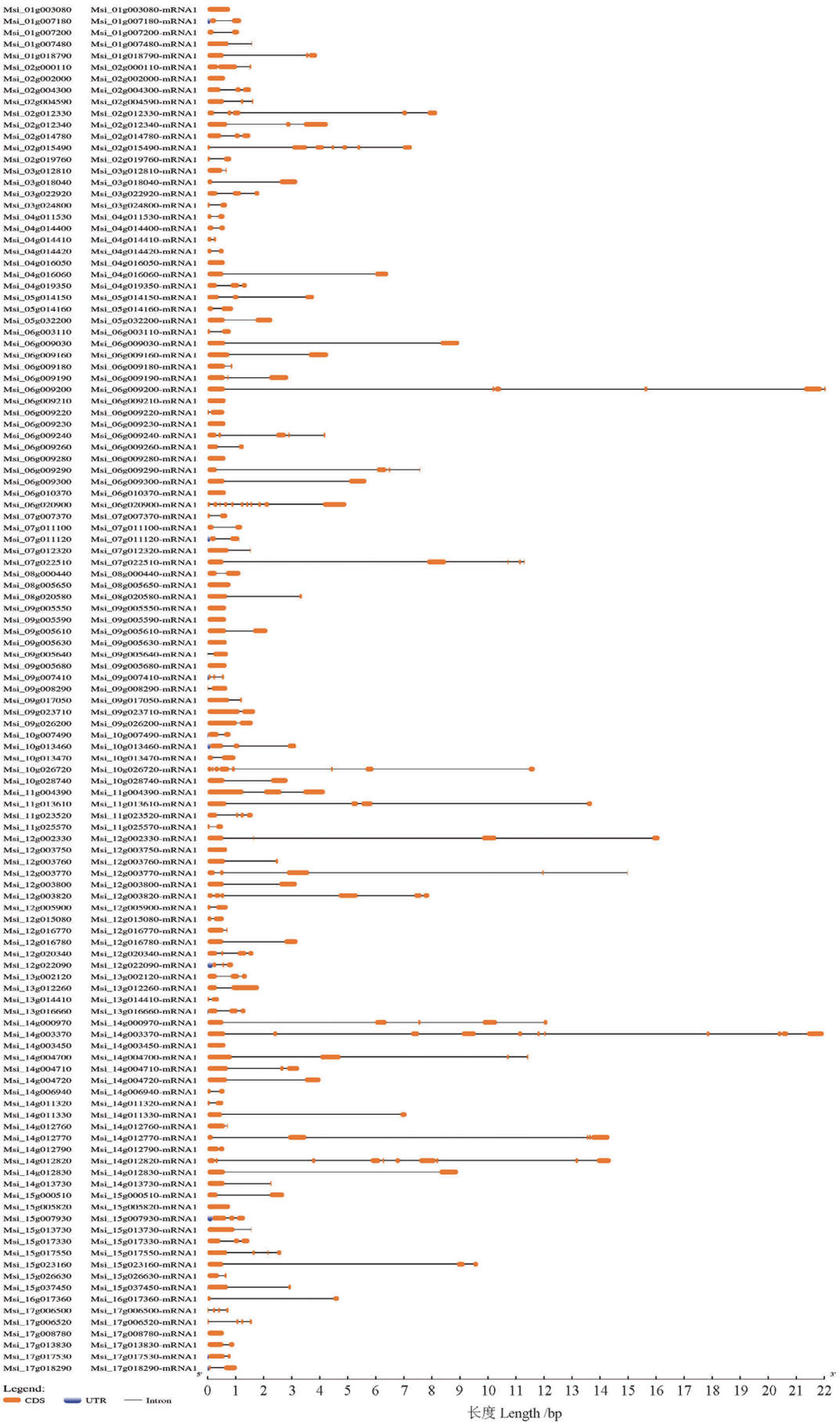


图 3　新疆野苹果 *LEA* 基因家族的种内共线性分析

Fig. 3　Intraspecific colinearity analysis of *LEA* gene family in *Malus sieversii*



棒状结构代表外显子部分，线状代表内含子部分。Exons are represented by bars, introns are represented by connecting lines.

图 4　新疆野苹果 *LEA* 基因结构

Fig. 4　Intron-exon structures of *LEA* genes in *Malus sieversii*

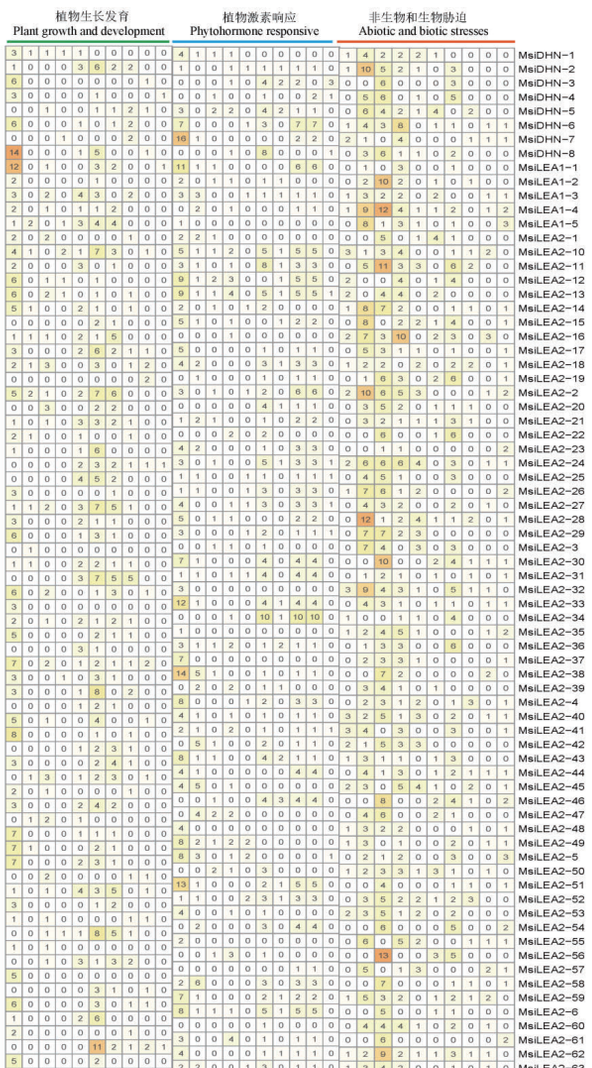


图 5 新疆野苹果 LEA 家族成员的顺式作用元件

Fig. 5 Cis-acting elements in promoters of *LEA* family members in *Malus sieversii*

活。大量与植物对胁迫的响应相关的顺式元件，主要元件分别是 MBS（MYB 结合位点，参与耐旱性）、LTR（低温响应）、DRE（干旱响应）、STRE（应激响应）、ARE（厌氧诱导）和 TC-rich repeats（防御和应激响应）、W-Box（病原体攻击响应）、WUN（损伤响应）。这些 *MsiLEA* 启动子区域含有各种类型的顺式元件，表明它们可能参与多样化的生物过程和调控途径。

2.7 新疆野苹果 *LEA* 基因家族的表达模式分析

结合新疆野苹果种子低温层积不同时期的转录组数据，分析 *LEA* 基因在种子低温层积过程中的表达情况（图 6A）。在低温层积 0 d、30 d、60 d、90 d 和 120 d 的 5 个不同时期的新疆野苹果种子中，*MsiLEA* 基因表达量差异较大，119 个成员在种子不同时期均有表达。

MsiLEA1-1、*MsiLEA1-2*、*MsiLEA1-4*、*MsiLEA2-1*、*MsiLEA2-3*、*MsiLEA2-10*、*MsiLEA2-11*、*MsiLEA2-30*、*MsiLEA2-31*、*MsiLEA2-32*、*MsiLEA2-33*、*MsiLEA2-37*、*MsiLEA2-44*、*MsiLEA2-46*、*MsiLEA2-56*、*MsiLEA2-57*、*MsiLEA2-60*、*MsiLEA2-65*、*MsiLEA2-69*、*MsiLEA2-72*、*MsiLEA2-78*、*MsiLEA3-1*、*MsiLEA4-7*、*MsiLEA4-9*、*MsiLEA5-1*、*MsiLEA5-2*、*MsiLEA5-5*、*MsiDHN-1*、*MsiDHN-5*、*MsiDHN-7*、*MsiDHN-8* 和 *MsiSMP-2* 基因的表达量升高，*MsiLEA2-24*、*MsiLEA2-34*、*MsiLEA2-48*、*MsiLEA2-49*、*MsiLEA2-50*、*MsiLEA2-51*、*MsiLEA2-52*、*MsiLEA2-55*、*MsiLEA2-66*、*MsiLEA2-68*、*MsiLEA2-74*、*MsiLEA2-75*、*MsiLEA4-5*、*MsiLEA4-6*、*MsiLEA5-3*、*MsiLEA5-4*、*MsiLEA6-2*、*MsiDHN-4*、*MsiSMP-1* 和 *MsiSMP-7* 在种子低温层积的前 60 天表达量很低或检测不到转录本。*MsiDHN3* 在种子低温层积过程中表达量下降。

为进一步研究新疆野苹果 *LEA* 家族基因的表达情况，结合苹果所有器官（花、果实、种子、根、茎和叶）的转录组数据，分析 *MsiLEA* 基因在各器官的表达情况（图 6B）。*MsiLEA1-1* 主要在花中有较高的表达；*MsiLEA2-4* 主要在花、叶和茎中有较高的表达；*MsiLEA2-12*、*MsiLEA6-1* 和 *MsiSMP-7* 主要在花、叶、果实、根和茎中有较高的表达；*MsiLEA2-18* 主要在花、叶、根和茎中有较高的表达；*MsiLEA2-42*、*MsiLEA2-48*、

MsiDHN-3 和 *MsiSMP-5* 在所有器官中有较高的表达；*MsiLEA2-68* 主要在果实和根中有较高的表达；*MsiLEA4-2*、*MsiLEA6-2* 主要在花、叶和果实中有较高的表达；*MsiLEA4-7* 主要在花、叶和根中

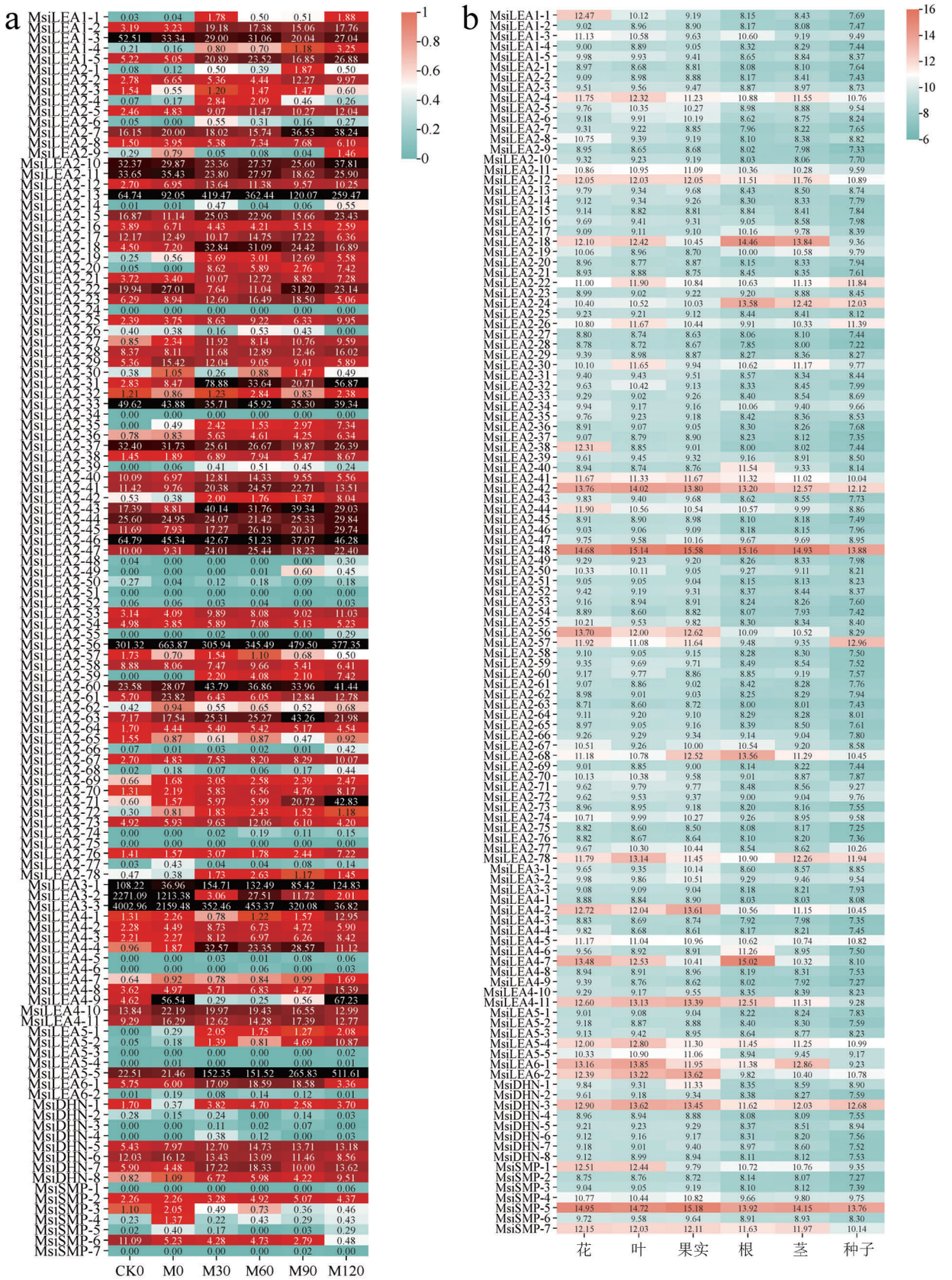


图 6　新疆野苹果 *MsiLEA* 基因在种子低温层积不同时期（A）和不同器官（B）的表达分析

Fig. 6　Expression analysis of *MsiLEA* genes during seed low-temperature stratification in different periods（A）and different organs（B）in *Malus sieversii*

有较高的表达；*MsiLEA4-11* 主要在花、叶、果实和根中有较高的表达；*MsiLEA* 在各器官中都有表达，但不同器官中表达水平差异较大。

3 讨 论

近年来，随着基因组分析技术的兴起，越来越多的 LEA 家族成员被鉴定和分析。LEA 蛋白是一种著名的应激相关蛋白因子，植物无法改变自然环境，必须通过自己的机制来适应。新疆野苹果作为栽培苹果的主要祖先之一，新疆野苹果中 *LEA* 家族基因的鉴定和分析尚未见报道。因此，本研究对新疆野苹果 *LEA* 家族基因进行全基因组筛选和鉴定，并对其表达水平进行分析。共鉴定新疆野苹果 119 个 *MsiLEA* 家族基因，并将其分为 8 个亚家族。与其他已报道物种的 *LEA* 基因数量相比，如毛果杨 88 个 [44]，花生 126 个 [17]，三七 *Panax notoginseng* 61 个 [45]，辣椒 *Capsicum annuum* 82 个 [46]，苹果 87 个 [34]，新疆野苹果 *LEA* 基因数量较多。在 MsiLEA 的 8 个亚科中，LEA2 所占比例最大，高达 65%，占绝对优势，这与目前已报道的烟草 *Nicotiana tabacum*[47]、其他种类所占比例一致 [7,48]。新疆野苹果中大量的 *LEA* 基因更可能是基因复制和 *LEA* 基因的保守性造成的。新疆野苹果包含 LEA4 亚组，但在苹果 LEA4 亚组中没有发现完整的 LEA 蛋白，这在 LEA 蛋白家族的报道中是罕见的 [34]。

保守结构域分析结果表明，LEA 亚家族的每个成员都包含一组独特的保守结构域，其中已经在拟南芥 [4]、水稻 [5]、小麦 *Triticum aestivum*[49] 和栽培花生 [50] 等植物中观察到 LEA 蛋白保守结构域，这些结果表明 MsiLEA 蛋白不同亚家族的功能不同。此外，在每个 LEA 亚家族中的保守基序表明，它们的成员可能起源于该亚家族内的基因扩增 [51]。研究表明，基因的内含子序列虽然不直接参与蛋白质编码，但在基因功能中起着至关重要的作用。基因结构中内含子数量的减少归因于进化过程，主要目的是加速从转录到翻译的转变。这种流线型的过程使基因能够快速合成功能蛋白质以响应外部刺激基因结构 [52-53]。*MsiLEA* 基因含有非常少的内含子，这与其他植物物种如甘蓝型油菜 [10]、茶树 *Camellia sinensis*[54]、玉米 [55] 和杂交枫香 [22] 一致，数量少的内含子有利于 *MsiLEA* 基因在非生物胁迫下的快速表达和功能蛋白的产生。MsiLEA 蛋白的亚细胞定位结果显示其广泛存在于细胞核、细胞质、细胞膜、线粒体和叶绿体中的亚细胞区室。

基因复制是基因组结构的一个主要特征，在植物基因组和生物体进化过程中发挥着重要作用，产生新的原始遗传物质用于遗传漂变、突变和选择，最终导致新的基因功能的产生和基因网络的进化 [48]。基因复制机制不仅导致基因组大小的扩展，而且有助于基因功能的多样化，以确保植物具有足够的适应性和进化能力。新疆野苹果 *LEA* 基因家族扩张与基因复制有必然联系，而全基因组复制或片段重复事件在所有导致基因加倍的事件中占比较大。

LEA 基因在促进植物界的生长、发育和植物激素响应中发挥广泛作用 [56-57]。先前的研究已经证明，*LEA* 基因参与植物对非生物逆境的响应，最显著的包括干旱、低温限制和高盐胁迫 [58]。结果显示，*MsiLEA* 的启动子序列含有丰富的胁迫相关的顺式作用元件，其表达受多种胁迫诱导。基于在低温层积 0 d、30 d、60 d、90 d 和 120 d 的 5 个时期新疆野苹果种子的表达数据，部分 LEA1、LEA2、LEA4、LEA5、SMP 亚家族成员在种子发育过程中逐渐升高，与小麦 [59]、胡麻 *Linum usitatissimum*[60] 和栽培花生 [50] 的结果一致，推测这些 *LEA* 基因可能对种子成熟脱水阶段发生的生理反应具有保护作用。部分成员的表达量变化不明显，另一部分成员的表达量很低或检测不到转录本，不同基因的表达方式不同，其在响应干旱胁迫、低温胁迫及盐胁迫等中发挥的作用也有差别。栽培花生中有 45 个 *AhLEA* 基因受低温胁迫后发生变化，其中 28 个基因表达量升高，17 个基因表达量降低 [50]。

4 结 论

本研究从新疆野苹果基因组共揭示了 119 个 *MsiLEA* 基因的存在，分布在 8 个亚家族。对 *MsiLEA* 基因的系统发育关系、保守结构域、外显子 - 内含子结构和顺式作用元件进行分析发现，不同亚家族基序之间差异较大。转录组结果表明，*MsiDHN3* 在种子低温层积过程中表达量下降，在所有器官中表达量较高。本研究可为深入理解 *LEA* 基因家族在新疆野苹果种子胚胎发育阶段中的作用提供基础数据。

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: 4162-4168.

[2] FINN R D, COGGILL P, EBERHARDT R Y, et al. The Pfam protein families database: towards a more sustainable future[J]. Nucleic Acids Research,2016,44:D279-D285.

[3] SHAO H B, LIANG Z S, SHAO M A. LEA proteins in higher plants: structure, function, gene expression and regulation[J]. Colloids and Surfaces B: Biointerfaces,2005,45:131-135.

[4] HUNDERTMARK M, HINCHA D K. LEA (Late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*[J]. BMC Genomics,2008,9:118.

[5] 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, 2007,172:414-420.

[6] LI X, CAO J. Late embryogenesis abundant (*LEA*) gene family in maize: identification, evolution, and expression profiles[J]. Plant Molecular Biology Reporter,2016,34:15-28.

[7] CHEN Y K, LI C H, ZHANG B, et al. The role of the late embryogenesis-abundant (LEA) protein family in development and the abiotic stress response: a comprehensive expression analysis of potato (*Solanum tuberosum*)[J]. Genes,2019,10:148.

[8] ALTUNOGLU Y C, BALOGLU P, YER E N, et al. Identification and expression analysis of *LEA* gene family members in cucumber genome[J]. Plant Growth Regulation,2016,80:225-241.

[9] CAO J, LI X. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*)[J]. Planta,2015,241:757-772.

[10] 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.

[11] WU C L, HU W, YAN Y, et al. The late embryogenesis abundant protein family in cassava (*Manihot esculenta* Crantz): genome-wide characterization and expression during abiotic stress[J]. Molecules,2018,23:1196.

[12] LAN T, GAO J, ZENG Q Y. Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in *Populus trichocarpa*[J]. Tree Genetics & Genomes,2013,9:253-264.

[13] DU D L, ZHANG Q X, CHENG T R, et al. Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*[J]. Molecular Biology Reports, 2013,40:1937-1946.

[14] GAO J, LAN T. Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabuliformis* (Pinaceae) in *Escherichia coli*[J]. Scientific Reports,2016,6:19467.

[15] LIN R Y, ZOU T, MEI Q M, 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[J]. International Journal of Molecular Sciences,2021,22:4554.

[16] LI Z, CHI H, LIU C Y, et al. Genome-wide identification and functional characterization of *LEA* genes during seed development process in linseed flax (*Linum usitatissimum* L.)[J]. BMC Plant Biology,2021,21:193.

[17] HUANG R L, XIAO D, WANG X, et al. Genome-wide identification, evolutionary and expression analyses of LEA gene family in peanut (*Arachis hypogaea* L.)[J]. BMC Plant Biology,2022,22:155.

[18] LING H, ZENG X, GUO S X. Functional insights into the late embryogenesis abundant (LEA) protein family from *Dendrobium officinale* (Orchidaceae) using an *Escherichia coli* system[J]. Scientific Reports,2016,6:39693.

[19] WANG G P, XU X F, GAO Z Y, et al. Genome-wide identification of LEA gene family and cold response mechanism of *BcLEA4-7* and *BcLEA4-18* in non-heading Chinese cabbage [*Brassica campestris* (syn. *Brassica rapa*) ssp. *chinensis*][J]. Plant Science,2022,321:111291.

[20] HUANG Z, ZHONG X J, HE J, et al. Genome-wide identification, characterization, and stress-responsive expression profiling of genes encoding LEA (Late embryogenesis abundant) proteins in moso bamboo (*Phyllostachys edulis*)[J]. PLoS One,2016,11(11):e0165953.

[21] ALTUNOGLU Y C, BALOGLU M C, BALOGLU P, et al. Genome-wide identification and comparative expression analysis of *LEA* genes in watermelon and melon genomes[J]. Physiology and Molecular Biology of Plants,2017,23(1):5-21.

[22] LI Y, QI S, CHEN S Z, CHEN S Y, et al. 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*)[J]. Forestry Research, 2023,3:12.

[23] 陈凯 , 佟晓楠 , 张晓媛 , 等 . 枳 *LEA* 基因家族鉴定及其对非生物胁迫的响应 [J]. 西北植物学报 ,2023,43(6):918-928.
CHEN K, TONG X N, ZHANG X Y, et al. Genome-wide identification and abiotic stress responses of *LEA* gene family in *Poncirus trifoliata*[J]. Acta Botanica Boreali-Occidentalia Sinica,2023,43(6):918-928.

[24] GARAY-ARROYO A, COLMENERO-FLORES J M, GARCIARRUBIO 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.

[25] GAL T Z, GLAZER I, KOLTAI H. An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress[J]. FEBS Letters,2004,577:21-26.

[26] BATTAGLIA M, COVARRUBIAS A A. Late embryogenesis abundant (LEA) proteins in legumes[J]. Frontiers in Plant Science,2013,4:190.

[27] OLVERA-CARRILLO Y, REYES J L, COVARRUBIAS A A. Late embryogenesis abundant proteins: versatile players in the plant adaptation to water limiting environments[J]. Plant Signaling & Behavior,2011,6(4):586-589.

[28] MUVUNYI B P, YAN Q, WU F, et al. Mining late embryogenesis abundant (LEA) family genes in *Cleistogenes songorica*, a xerophyte perennial desert plant[J]. International Journal of Molecular Sciences,2018,19:3430.

[29] CELTON J M, GAILLARD S, BRUNEAU M, et al. Widespread anti-sense transcription in apple is correlated with siRNA production and indicates a large potential for transcriptional and/or post-transcriptional control[J]. New Phytologist,2014,203: 287-299.

[30] TOLLETER D, HINCHA D K, MACHEREL D. A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state[J]. Biochimica et Biophysica Acta, 2010,1798:1926-1933.

[31] CHAKRABORTEE S, TRIPATHI R, WATSON M, et al. Intrinsically disordered proteins as molecular shields[J]. Molecular Biosystems,2012,8(1):210-219.

[32] 郑凤 , 李建贵 , 秦伟 , 等 . 新疆野苹果自然落储种实的萌发特性 [J]. 经济林研究 ,2024,42(3):197-206.
ZHENG F, LI J G, QIN W, et al. Germination characteristics of naturally dropped seed of *Malus sieversii*[J]. Non-wood Forest Research,2024,42(3):197-206.

[33] 王菁菁 , 刘立强 , 秦伟 , 等 . 新疆野苹果矮化性状形态学与生理指标评价 [J]. 经济林研究 ,2022,40(3):151-161.
WANG J J, LIU L Q, QIN W, et al. Dwarf evaluation on morphology and physiological indexes of Xinjiang wild apple[J]. Non-wood Forest Research,2022,40(3):151-161.

[34] WANG X Y, LI Y W, ZHANG L S, et al. Comprehensive identification of LEA protein family genes and functional analysis of MdLEA60 involved in abiotic stress responses in apple (*Malus domestica*)[J]. International Journal of Biological Macromolecules,2024,283:137641.

[35] SUN X P, JIAO C, SCHWANINGER H, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication[J]. Nature Genetics,2020,52: 1423-1432.

[36] CHEN C J, WU Y, LI J W, et al. TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining[J]. Molecular Plant,2023,16:1733-1742.

[37] MARCHLER-BAUER A, DERBYSHIRE M K, GONZALES N R, et al. CDD: NCBI’s conserved domain database[J]. Nucleic Acids Research,2015,43:D222-D226.

[38] GASTEIGER E, GATTIKER A, HOOGLAND C, et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis[J]. Nucleic Acids Research,2003,31(13):3784-3788.

[39] HORTON P, PARK K J, OBAYASHI T, et al. WoLF PSORT: protein localization predictor[J]. Nucleic Acids Research,2007,35: W585-W587.

[40] TAMURA K, STECHER G, KUMAR S. MEGA11: Molecular evolutionary genetics analysis version 11[J]. Molecular Biology and Evolution,2021,38(7):3022-3027.

[41] XIE J M, CHEN Y R, CAI G J, et al. Tree visualization by one table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees[J]. Nucleic Acids Research,2023,51:W587-W592.

[42] HU B, JIN J P, GUO A Y, et al. GSDS 2.0: an upgraded gene feature visualization server[J]. Bioinformatics,2015,31(8): 1296-1297.

[43] LESCOT M, DÉHAIS P, THIJS G, et al. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences[J]. Nucleic Acids Research,2002,30(1):325-327.

[44] CHENG Z H, ZHANG X M, YAO W J, et al. Genome-wide search and structural and functional analyses for late embryogenesis- abundant (LEA) gene family in poplar[J]. BMC Plant Biology,2021,21:110.

[45] JIA J S, GE N, WANG Q Y, et al. Genome-wide identification and characterization of members of the LEA gene family in *Panax notoginseng* and their transcriptional responses to dehydration of recalcitrant seeds[J]. BMC Genomics,2023,24:126.

[46] ZHAO Y Y, HAO Y P, DONG Z Y, et al. Identification and expression analysis of LEA gene family members in pepper (*Capsicum annuum* L.)[J]. FEBS Open Bio,2023,13:2246-2262.

[47] GENG W B, WANG Y N, ZHANG J, et al. 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[J]. Gene, 2022,836:146665.

[48] MAGWANGA R O, LU P, KIRUNGU J N, et al. Characterization of the late embryogenesis abundant (LEA) proteins family and

their role in drought stress tolerance in upland cotton[J]. BMC Genetics,2018,19:6.

[49] LIU H, XING M Y, YANG W B, et al. Genome-wide identification of and functional insights into the late embryogenesis abundant (*LEA*) gene family in bread wheat (*Triticum aestivum*)[J]. Scientific Reports,2019,9:13375.

[50] 付娆 , 李佳佳 , 单燕 , 等 . 栽培种花生 LEA 蛋白基因家族的鉴定及其低温胁迫下表达分析 [J]. 花生学报 ,2022,51(3):1-12.
FU R, LI J J, SHAN Y, et al. Genome-wide identification and expression analysis of the late embryogenesis abundant (LEA) protein gene family in cultivated peanut (*Arachis hypogaea* L.) on its response to cold stress[J]. Journal of Peanut Science, 2022,51(3):1-12.

[51] ZENG X, LING H, YANG J W, et al. LEA proteins from *Gastrodia elata* enhance tolerance to low temperature stress in *Escherichia coli*[J]. Gene,2018,646:136-142.

[52] JEFFARES D C, PENKETT C J, BÄHLER J. Rapidly regulated genes are intron poor[J]. Trends in Genetics,2008,24(8): 375-378.

[53] LANE N, MARTIN W. The energetics of genome complexity[J]. Nature,2010,467:929-934.

[54] WANG W D, GAO T, CHEN J F, et al. The late embryogenesis abundant gene family in tea plant (*Camellia sinensis*): genome-wide characterization and expression analysis in response to cold and dehydration stress[J]. Plant Physiology and Biochemistry, 2019,135: 277-286.

[55] QIAN Y X, CHEN C L, JIANG L Y, et al. Genome-wide identification, classification and expression analysis of the JmjC domain-containing histone demethylase gene family in maize[J]. BMC Genomics,2019,20:256.

[56] 李雪宝 , 周璇 , 鄢波 . 纤枝短月藓 *LEA5* 基因表达及耐盐性分析 [J]. 西北植物学报 ,2021,41(6):919-925.
LI X B, ZHOU X, YAN B. Expression and salt tolerance of *LEA5* gene from *Brachymenium exile*[J]. Acta Botanica Boreali-Occidentalia Sinica,2021,41(6):919-925.

[57] 陈益 , 杨妮 , 罗微 , 等 . 逆境与赤霉素调控下茶树 *BRX* 同源基因的系统性鉴定与响应模式 [J]. 西北植物学报 ,2024,44(11): 1725-1734.
CHEN Y, YANG N, LUO W, et al. Systematic identification and response patterns of *BRX* orthologs from *Camellia sinensis* under stress and gibberellin regulation[J]. Acta Botanica Boreali-Occidentalia Sinica,2024,44(11):1725-1734.

[58] RODRIGUEZ-SALAZAR J, MORENO S, ESPÍN G. LEA proteins are involved in cyst desiccation resistance and other abiotic stresses in *Azotobacter vinelandii*[J]. Cell Stress and Chaperones,2017,22:397-408.

[59] JIN X F, CAO D, WANG Z J, et al. Genome-wide identification and expression analyses of the LEA protein gene family in tea plant reveal their involvement in seed development and abiotic stress responses[J]. Scientific Reports,2019,9:14123.

[60] 李震 . 胡麻 LEA 基因家族的鉴定及其在种子发育过程中的遗传效应 [D]. 北京 : 中国农业科学院 ,2021.
LI Z. Identification and genetic effect of *LEA* genes during seed development in *Linseed flax*[D]. Beijing: Chinese Academy of Agricultural Sciences,2021.

[本文编校：吴　彬]