

doi:10.11937/bfyy.20172195

绿豆类甜蛋白家族鉴定与生物信息学分析

刘 潮, 韩利红, 王海波, 宋培兵, 唐利洲

(曲靖师范学院 云南高原生物资源保护与利用研究中心, 生物资源与食品工程学院,
云南省高校云贵高原动植物多样性及生态适应性进化重点实验室, 云南 曲靖 655011)

摘 要: 为了解全基因组水平上绿豆类甜蛋白家族基本生物学特征, 通过生物信息学手段, 鉴定绿豆类甜蛋白家族成员, 并对基因和蛋白结构及染色体定位进行了分析, 为进一步克隆绿豆类甜蛋白基因以及探讨其抗真菌活性奠定基础。结果表明: 共鉴定出 34 个绿豆类甜蛋白家族成员, 基因主要有 4 种结构类型, 分布在除 4 号和 9 号染色体外的其余 9 条染色体中。该家族蛋白功能主要构成细胞结构组分和参与细胞生物学进程。系统发育分析显示, 绿豆类甜蛋白家族归属 10 个聚类组, 其中成员较多的聚类组 6 和聚类组 7 中的基因主要来自 11 号染色体, 其中 7 个基因簇中的基因存在紧密连锁现象, 属于旁系同源基因, 可能发生了染色体复制。密码子使用性分析显示, 绿豆 TLP 家族基因偏好 A 或 T 作为第三位密码子, 但密码子使用总体偏性不强, 多数属于低表达基因, 多数基因进化受碱基突变和正向选择压力的双重影响。

关键词: 绿豆; 类甜蛋白; 密码子偏性; 生物信息学

中图分类号: S 552.501 **文献标识码:** A **文章编号:** 1001-0009(2017)20-0031-09

类甜蛋白(Thaumatococcus-like protein, TLP)常具有抗真菌活性, 属于病程相关蛋白第 5 家族, 因其氨基酸序列与西非竹筴(*Thaumatococcus daniellii*)的甜蛋白(Thaumatococcus)高度同源而得名, 广泛分布于多种植物、动物及微生物基因组中^[1-2]。植物 TLPs 主要通过裂解真菌孢子, 抑制孢子萌发, 降低幼嫩菌丝活力等方式对致病和非致病真菌产生抑制作用^[1,3]。TLP 基因能被多种胁迫诱导表达^[4-5], 体内或体外试验表明 TLPs 具有抗真菌活性^[6], TLP 转基因植物对真菌性病害的耐受性和抗性明显增强^[7-9]。TLPs 除了参与

植物的抗逆反应之外, 也参与植物生长发育的多项进程, 有些植物在果实成熟时很多 TLPs 高表达^[10-11]。目前, 多种植物 TLPs 已被鉴定^[7,12]。系统发育分析显示, 水稻和拟南芥 TLPs 分布于多个支系, 并存在染色体内和染色体间的复制现象, 推测单子叶和双子叶植物进化上发生分离后, TLP 基因在 10 个进化枝上发生了不对称的增加^[2], 相对于动物和真菌, 陆生植物进化过程中 TLP 基因含量和多样性也得到显著增加^[13]。

不同物种或同一物种的不同基因倾向于使用一种或几种特定的同义密码子, 称为密码子的使用偏性, 密码子的使用偏性与基因的表达和蛋白功能密切相关。因此, 分析物种中基因密码子的使用性有助于了解基因在生物体中的功能及其进化特征, 对具有特定功能的基因开发利用具有重要的应用价值。

尽管研究者对 TLP 家族的研究有了重大进展, 但这些研究主要限于拟南芥和水稻等模式植

第一作者简介: 刘潮(1980-), 男, 河北景县人, 博士, 讲师, 现主要从事植物病理学等研究工作。E-mail: liuchao@mail.qjnu.edu.cn.

基金项目: 国家自然科学基金资助项目(31460179); 云南省高校云南特境内生菌资源的开发与利用科技创新团队资助项目。

收稿日期: 2017-07-10

物。绿豆(*V. radiata*)属豆科豇豆属,是一种重要的食用豆类作物,富含谷氨酸、天冬氨酸、酪氨酸等多种氨基酸和磷脂酰胆碱、磷脂酰乙醇胺、磷脂酰肌醇、磷脂酰丝氨酸等磷脂成分以及多种维生素和矿质元素^[14]。目前,绿豆生产中出现品种老化、混杂严重等问题,绿豆炭疽病和细菌性疫病等病害普遍,在一定程度上影响了绿豆产业的发展。由于生长习性、成熟同步性和生物与非生物胁迫等原因,导致绿豆的平均产量较低。针对以上问题,亟需对绿豆现有品种进行资源开发和品种改良,筛选和培育优质、高产、抗病新品种。为全面了解绿豆基因组中 TLP 家族基因结构和蛋白功能,该研究通过生物信息学手段,从基因组水平上分析绿豆 TLP 基因的数目、基因结构、染色体定位以及系统进化,为进一步克隆绿豆类甜蛋白基因和抗病品种改良奠定基础。

1 材料与方法

1.1 绿豆 TLP 家族数据的获取

绿豆(*V. radiata*)全基因组数据下载于 GenBank 基因组数据库,其基因组大小为 459.27 Mb,分布于 11 条染色体^[15]。以典型 TLP 蛋白小果野蕉(*Musa acuminata*) 1Z3Q 为探针序列,搜索绿豆蛋白数据库,获得注释为 Thaumatin-like protein 的蛋白序列。

1.2 TLP 蛋白理化特性分析

通过 Expasy (<http://www.expasy.org/tools/>)对蛋白理化特征进行预测。使用 SMART(<http://smart.embl-heidelberg.de/>)对蛋白功能域进行预测。

1.3 TLP 基因结构及染色体定位分析

从 NCBI 中获得绿豆 TLP 蛋白对应的基因序列和 CDS 序列。使用基因结构显示系统(<http://gsds.cbi.pku.edu.cn/index.php>)绘制基因结构示意图。通过 MEME SUITE 在线工具(<http://meme-suite.org/tools/meme>)预测绿豆 TLP 转录因子蛋白序列的保守 Motif, Motif 搜索数目为 5,其它参数为默认设置。通过 Mapinspect (<http://mapinspect.software.informer.com/>)进行染色体定位作图。

1.4 绿豆 TLP 家族蛋白 GO 功能分析

利用 UniProt 数据库(<http://www.uniprot.org/>)对筛选的潜在致敏原蛋白进行结构域和 GO(Gene Ontology)功能分类进行分析。利用 WEGO 在线软件(Web Gene Ontology Annotation Plotting, <http://wego.genomics.org.cn/>)对蛋白 GO 富集度进行计算。

1.5 TLP 蛋白序列比对与系统进化分析

所有蛋白使用 ClustalX 进行序列比对,应用 MEGA 5.0 软件,采用邻接法(Neighbor-Joining)构建系统发育树。

1.6 绿豆 TLP 家族基因密码子偏性分析

采用 CodonW 软件对绿豆 TLP 家族基因编码序列密码子使用特性进行分析。分析参数为密码子适应指数(CAI)、有效密码子数(ENC)、密码子第 3 位的 G+C 含量(GC3s)、平均亲水性值(Gravy)。以 GC3s 为横坐标,ENC 为纵坐标,绘制 ENC 与 GC3s 的关联分布图^[14]。使用 EMBOSS explorer 网站在线软件对同义密码子相对使用度(RSCU)进行分析。

2 结果与分析

2.1 绿豆 TLP 家族成员的鉴定

鉴定出 34 个具有 THN 典型结构域的绿豆 TLP 成员(表 1)。蛋白理化性质分析显示,最长的蛋白序列 XP_014493317.1 包含 565 个氨基酸残基,最短的 XP_014493286.1 包含 153 个氨基酸残基,蛋白分子量介于 16.47~63.01 kDa,蛋白等电点介于 4.27~9.04,酸性蛋白占 64.7%。鉴定的 34 条 TLP 序列中,与 1Z3Q 序列相似性均高于 35%,所有序列 E 值均小于 1.00E-20。

利用 Clustal X 进行序列比对发现,所有绿豆 TLP 均含有 16 个半胱氨酸残基,均具有索玛甜家族标签 G-X-[GF]-X-C-X-T-[GA]-D-C-X-(1,2)-G-X-(2,3)-C 和 5 个保守的 REDDD 氨基酸残基,后者参与蛋白维持适当的拓扑结构和酸裂周围的表面静电势,对 TLPs 抗真菌活性必不可少^[7]。

表 1 绿豆中 TLP 蛋白家族信息

Table 1 Information of TLP protein family in mungbean

登录号 Accession	氨基酸数 AA	分子量 MW/kDa	等电点 pI	E 值 ^a E value	相似性 ^b Identity/%
XP_014494677.1	233	25.52	8.82	2.00E-34	36
XP_014503355.1	248	26.65	9.04	3.00E-49	43
XP_014509493.1	254	27.56	8.68	3.00E-43	39
XP_014495081.1	243	25.75	6.57	4.00E-43	41
XP_014523728.1	245	25.86	7.39	4.00E-42	38
XP_014492711.1	248	26.29	6.78	3.00E-35	35
XP_014493286.1	153	16.47	5.38	3.00E-21	35
XP_014493317.1	565	63.01	8.88	4.00E-23	36
XP_014508015.1	249	26.62	8.05	2.00E-115	80
XP_014520219.1	242	26.22	6.39	7.00E-107	73
XP_014520444.1	225	24.01	4.46	5.00E-98	67
XP_014502426.1	338	35.62	4.85	3.00E-52	44
XP_014505256.1	312	32.79	4.77	9.00E-51	46
XP_014506182.1	247	25.62	5.41	1.00E-39	43
XP_014519640.1	281	30.32	6.70	1.00E-33	38
XP_014519641.1	519	56.17	8.01	7.00E-33	40
XP_014519642.1	422	46.73	8.92	1.00E-30	37
XP_014519643.1	271	28.98	8.28	3.00E-33	37
XP_014519677.1	310	32.87	8.61	4.00E-48	40
XP_014520780.1	302	31.11	4.37	4.00E-50	45
XP_014492719.1	242	25.88	4.58	3.00E-48	42
XP_014494622.1	351	35.83	4.27	6.00E-50	46
XP_014503744.1	310	32.88	4.78	2.00E-40	38
XP_014507833.1	303	31.66	5.33	7.00E-47	44
XP_014512863.1	242	25.74	4.59	9.00E-47	41
XP_014519334.1	322	33.37	4.88	1.00E-50	45
XP_014519335.1	245	25.52	5.00	8.00E-51	45
XP_014504004.1	311	33.41	5.21	3.00E-50	41
XP_014504005.1	280	30.07	5.23	4.00E-50	41
XP_014503421.1	245	24.95	4.76	8.00E-51	43
XP_014504828.1	244	24.95	4.68	6.00E-49	44
XP_014504136.1	286	31.10	8.71	1.00E-52	47
XP_014518404.1	328	35.68	7.73	3.00E-41	41
XP_014518405.1	298	32.06	6.28	1.00E-41	41

注: ^a, ^b 表示与小果野蕉 1Z3Q 比对的 E 值和相似度。

Note: ^a, ^b match with 1Z3Q from *M. acuminata*.

2.2 绿豆 TLP 家族基因和蛋白结构分析

通过对绿豆 TLP 家族基因结构、内含子位置的分析发现(图 1),主要有 4 种类型的基因结构,其中不含内含子的基因数为 6,含有 1、2 和 3 个内含子的基因数分别为 6、12、12,含有 3 个以上内含子的基因数为 4,含有同一数量内含子的基因类型中,内含子的相位类型多样,说明基因结构类型较丰富,基因演变的来源多样。

蛋白保守氨基酸 Motif 分析发现 5 类 Motif 的保守性较强,其正则表达式分别为, Motif 1: CKPSLYSKIFKSACPKAYSAYDDATSTF; Motif 2: GGLDFYDVSLVDGYNLPMSVTP; Motif 3:

WSGRIWGRGTGCSFDS; Motif 4: GKCSATGC-VADLNAACPSELQVKGGDGAV; Motif 5: TFIIVNKCDYTVWPGILSGAG。34 条序列含有 Motif 1、2、4, 32 条序列含有 Motif 3, 33 条序列含有 Motif 5。其中蛋白 XP_014519641.1 含有 2 套保守 Motif, 这些基因序列可能发生了串联复制或重新拼接。

2.3 绿豆 TLP 家族蛋白功能分类

通过 GO 注释将基因产物分为细胞组分、分子功能和生物过程,可以对蛋白家族的功能进行分类以了解其在植物生命进程中的功能。在 34 条绿豆 TLPs 中,发现 23 条不同的 GO 注释子条目, 4 类作为细胞结构组分起作用,其中参与细胞 (cell, GO: 0005623) 和细胞组分 (cell part, GO: 0044464) 结构组成的均占 52.9%, 参与胞外区作用的 (extracellular region, GO: 0005576) 占 91.2%, 参与共质体构成的 (symplast, GO: 0055044) 占 8.8%。分子功能分类中,具有结合功能 (binding, GO: 0005488)、催化活性 (catalytic activity, GO: 0003824) 和分子转导活性 (molecular transducer activity, GO: 0060089) 的蛋白均只占 2.9%。5 类参与了生物学过程,其中参与胞内进程 (cellular process, GO: 0009987) 和代谢进程 (metabolic process, GO: 0008152) 的只占 2.9%, 参与多器官进程 (multi-organism process, GO: 0051704) 的占 91.2%, 参与免疫系统进程 (immune system process, GO: 0002376) 和响应应激 (response to stimulus, GO: 0050896) 的分别占 11.8% 和 94.1% (图 2)。

2.4 绿豆 TLP 家族系统进化分析

为了解绿豆 TLP 与其它物种同源基因的进化关系,选取了 10 个拟南芥 (*Arabidopsis thaliana*) TLPs、10 个水稻 (*Oryza sativa*) TLPs 和 10 个毛果杨 (*Populus trichocarpa*) TLPs 与 34 个绿豆 TLPs 一起构建 NJ 系统进化树(图 3)。发现拟南芥、水稻和毛果杨的 10 个 TLPs 分别属于 10 个聚类组^[2,12], 绿豆 TLPs 在 10 个聚类组中均有成员分布。其中聚类组 6 成员最多,包含 9 个绿豆 TLP 成员,聚类组 7 中有 7 个绿豆 TLP 成员,聚类组 5 含有 3 个绿豆 TLP 成员,有 4 个聚类组中仅包含 2 个绿豆 TLP 成员。聚类组 5 的其它

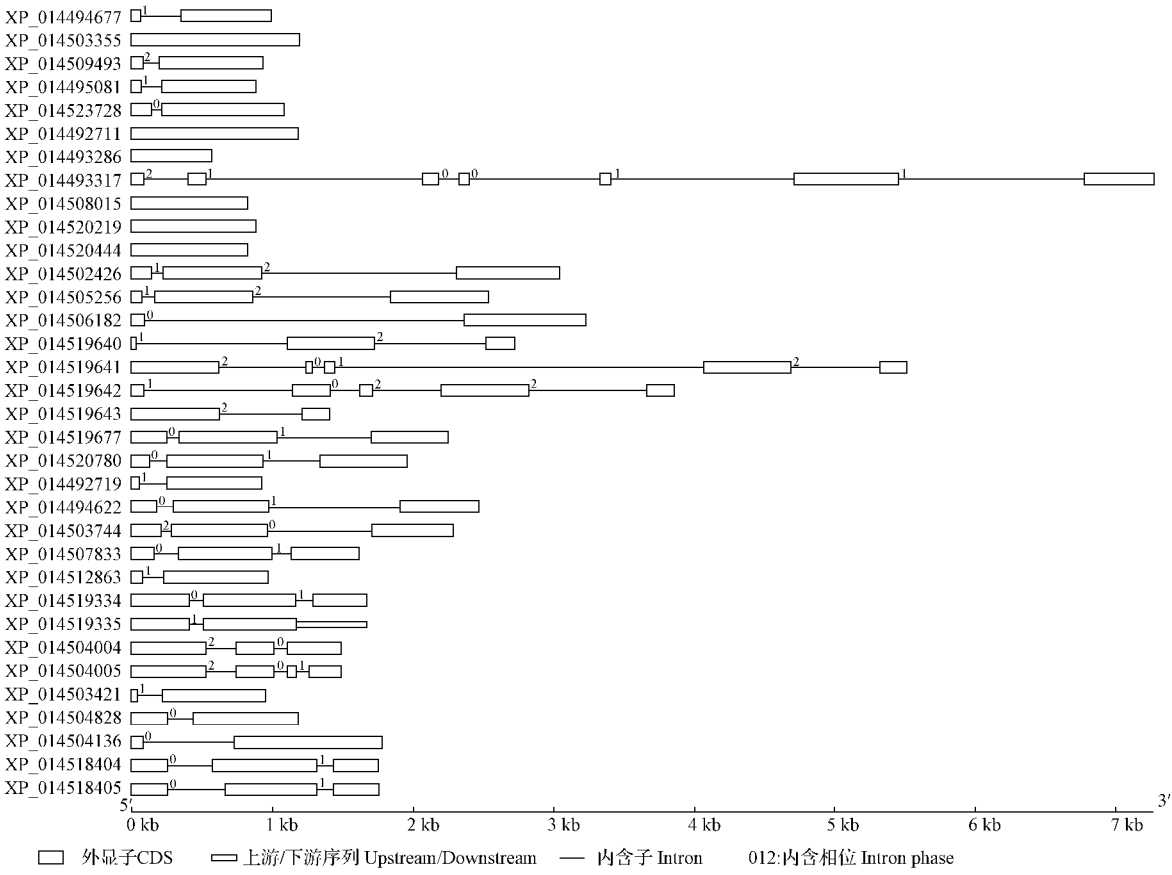


图 1 绿豆 TLP 家族基因结构
Fig. 1 Gene structure of TLPs gene family in mungbean

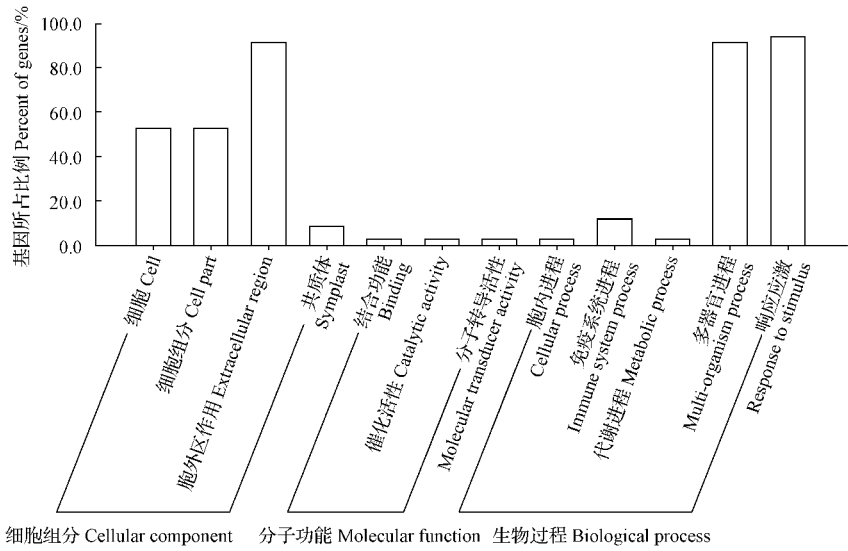
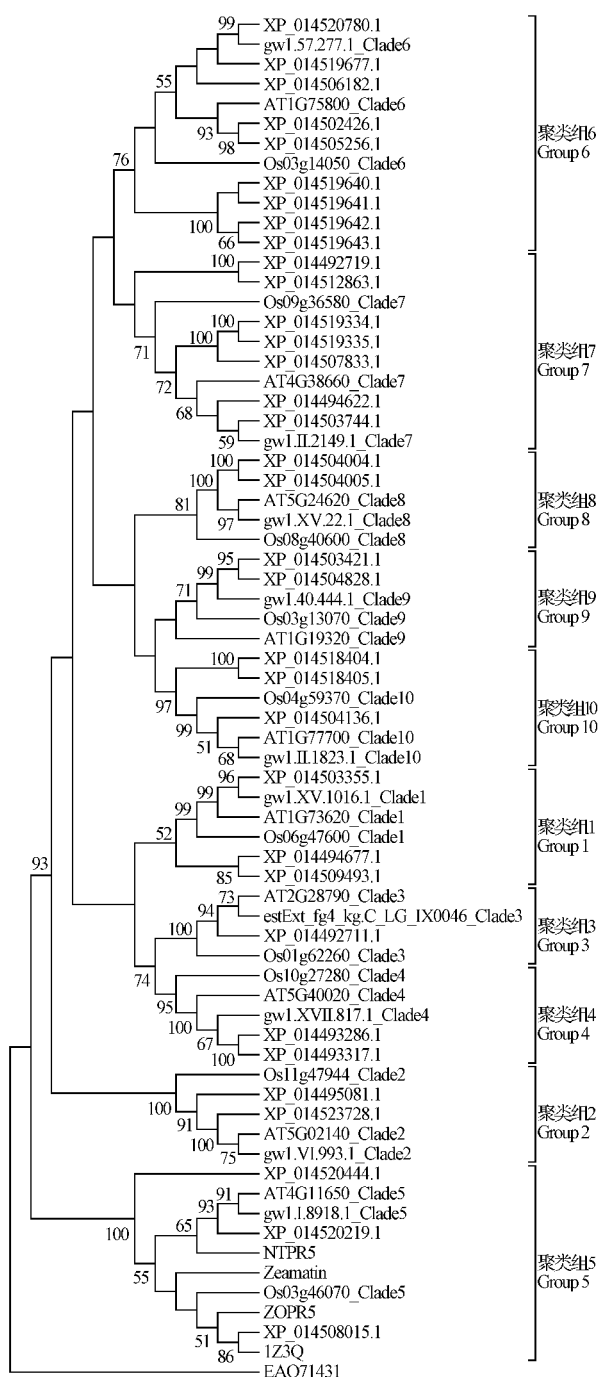


图 2 绿豆 TLP 蛋白 GO 功能分类
Fig. 2 GO classification of mungbean TLP

物种成员均能响应病原或环境胁迫,具有抗真菌活性^[16-17],该组绿豆 TLPs 可能也具有抑菌或葡萄糖苷酶活性。



注: TLPs 后缀的 Clade 序号参考文献[2, 12]。

Note: Clade number following the protein sequences refer to references[2, 12].

图 3 绿豆 TLP 家族蛋白系统发育树

Fig. 3 Phylogenetic of TLP family in mungbean

2.5 绿豆 TLP 家族基因染色体定位分析

根据基因位置信息,绘制了基因的染色体定位图(图 4),发现绿豆 TLP 基因分布在 11 条染色体中的 9 条上。其中最多的 11 号染色体上有 9 个 TLP 基因,其次为 6 号染色体,有 7 个基因,3 号、5 号和 8 号染色体最少,只有 1 个基因。分析发现,绿豆 TLP 基因在染色体上存在聚集现象,如在 1 号染色体的 3 Mb 和 8 Mb,2 号染色体 3 Mb,6 号染色体的 11、36、36 Mb,7 号染色体的 2 Mb、10 号染色体的 20 Mb、11 号染色体的 2、6 Mb 位置都存在 2 个或 2 个以上的基因集中分布的现象。研究发现,聚类组 5、6 和 7 中的成员主要来自 11 号染色体,聚类组 9 中的成员主要来自 6 号染色体,其中 7 个基因簇(XP_014493286.1/XP_014493317.1、XP_014494677.1/XP_014509493.1、XP_014504004.1/XP_014504005.1、XP_014503421.1/XP_014504828.1、XP_014518404.1/XP_014518405.1、XP_014520219.1/XP_014520444.1、XP_014519640.1/XP_014519641.1/XP_014519642.1/XP_014519643.1)在染色体上紧密连锁,属于旁系同源基因,说明这些基因可能是染色体内复制的结果。

2.6 绿豆 TLP 家族基因密码子使用特性

对基因密码子使用的多个参数进行分析发现,绿豆 TLP 家族基因密码子第三位的核苷酸使用总体偏性不强,基因偏好 A 或 T 作为第三位密码子,其中 5 个基因的 GC3s 数值低于 40%,仅基因 XP_014506182.1 的 GC3s 高于 60%,说明这些基因密码子偏好性较强。ENC 值反映基因编码对密码子选择性的强弱,高表达基因密码子 ENC 值较小^[14]。CAI 值为密码子适应指数,其值越高表示基因表达水平越强。分析发现,绿豆 TLP 家族基因 ENC 值介于 45.80~61.00,91% 的基因 ENC 值大于 50,CAI 值介于 0.146~0.296,只有 6 个基因的 CAI 值高于 0.250(表 2),说明绿豆 TLP 家族基因密码子使用偏好性不强,多数属于低表达基因,响应植物生长或外界刺激信号的能力较弱,仅有少数基因属于高表达基因,这部分基因可能在植物响应环境胁迫中发挥作用。59% 的绿豆 TLP 家族基因 Gravy 值为负值,表示多数蛋白属于亲水性蛋白。

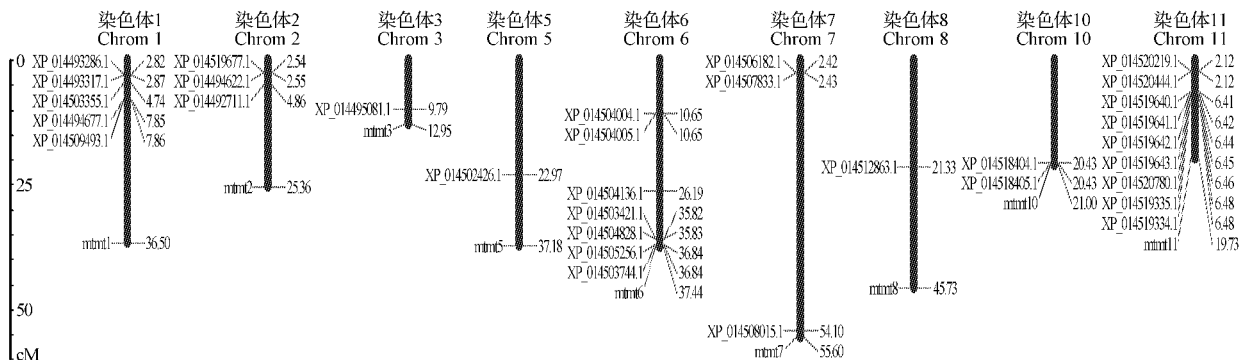


图4 绿豆 TLP 家族基因染色体定位

Fig. 4 Chromosome location of TLPs gene family in mungbean

表 2 绿豆 TLP 家族基因密码子使用特性

Table 2 Characterization of codon usage of TLP gene family in mungbean

登录号 Accession	密码子适 应指数 CAI	密码子第 3 位 的 G+C 含量 GC3s/%	有效密 码子数 ENC	平均亲 水性值 Gravy	登录号 Accession	密码子适 应指数 CAI	密码子第 3 位 的 G+C 含量 GC3s/%	有效密 码子数 ENC	平均亲 水性值 Gravy
XP_014494677.1	0.213	48.7	52.63	0.054	XP_014519643.1	0.217	50.4	57.09	0.056
XP_014503355.1	0.165	45.9	53.34	-0.132	XP_014519677.1	0.180	45.2	58.76	-0.254
XP_014509493.1	0.158	45.7	55.25	-0.419	XP_014520780.1	0.168	43.7	53.67	-0.211
XP_014495081.1	0.209	38.4	48.23	0.123	XP_014492719.1	0.296	52.8	54.05	-0.289
XP_014523728.1	0.173	42.7	52.05	0.587	XP_014494622.1	0.177	45.4	56.30	-0.316
XP_014492711.1	0.213	49.5	56.38	-0.085	XP_014503744.1	0.160	46.0	58.04	-0.056
XP_014493286.1	0.177	37.5	45.80	0.114	XP_014507833.1	0.152	42.7	55.51	-0.137
XP_014493317.1	0.178	45.6	56.78	0.031	XP_014512863.1	0.273	47.9	54.67	-0.297
XP_014508015.1	0.165	41.7	57.21	0.268	XP_014519334.1	0.148	40.6	53.33	-0.197
XP_014520219.1	0.156	49.0	54.28	-0.621	XP_014519335.1	0.240	59.2	54.73	0.045
XP_014520444.1	0.146	48.0	55.69	-0.648	XP_014504004.1	0.184	44.9	55.91	-0.187
XP_014502426.1	0.251	48.0	59.12	-0.066	XP_014504005.1	0.171	47.0	54.53	-0.291
XP_014505256.1	0.255	55.5	61.00	0.195	XP_014503421.1	0.261	57.9	57.06	0.169
XP_014506182.1	0.251	60.7	56.41	0.030	XP_014504828.1	0.183	47.4	54.45	-0.078
XP_014519640.1	0.218	49.6	58.08	0.089	XP_014504136.1	0.147	38.8	49.59	-0.120
XP_014519641.1	0.222	50.0	58.71	-0.007	XP_014518404.1	0.164	37.9	50.94	0.655
XP_014519642.1	0.214	51.5	61.00	-0.083	XP_014518405.1	0.159	38.1	50.70	0.688

ENC 与 GC3s 关联分布图中的 GC3s 分布反映了植物所受的选择压力,GC3s 分布越广泛,表明密码子使用偏好性受碱基突变压力越大,GC3s 分布范围越小,表明密码子使用偏好性受自然选择压力影响越大^[21]。在 ENC 与 GC3s 关联分析中,曲线为密码子偏性仅受碱基突变影响时的 ENC 预期值的位置。基因分布越靠近标准曲线表明密码子使用偏好性越不受正向选择压力的影响,基因分布在标准曲线下方或较远区域,表明该基因受到选择压力或其它因素的影响较大^[22]。绿豆 TLP 家族多数基因 ENC 值位于标准曲线下方,且多数 GC3s 分布相对较集中,说明基因进化受碱基突变和正向选择压力的双重影响,少数

ENC 值较小且远离标准曲线的基因主要受正向选择压力的影响(图 5)。

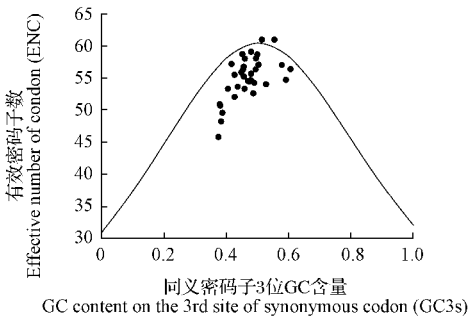


图 5 绿豆 TLP 家族基因 ENC 与 GC3s 的关系

Fig. 5 Correlative analysis of ENC and GC3s of TLP gene family in mungbean

同义密码子相对使用度(RSCU)是指同义密码子实际使用与理论使用数的比值。若 RSCU=1,表示密码子使用无偏性;若 RSCU>1,表示密码子使用频率高;若 RSCU<1,表示密码子使用频率低。使用 EMBOSS explorer 网站 Cusp 软件

分析并计算 TLP 家族基因 RSCU 值(表 3)。结果表明,绿豆不同基因编码蛋白的氨基酸密码子 RSCU 值不同,说明绿豆 TLP 家族蛋白氨基酸密码子存在偏性,但多数 RSCU 值偏离中心数值 1 的强度较弱,又说明密码子使用偏性并不强。

表 3
Table 3
绿豆 TLP 家族基因同义密码子使用情况
Usage of synonymous codon of TLP gene family in mungbean

氨基酸	密码子	RSCU	氨基酸	密码子	RSCU	氨基酸	密码子	RSCU	氨基酸	密码子	RSCU
Amino acid	Codon		Amino acid	Codon		Amino acid	Codon		Amino acid	Codon	
Ala	GCA	<u>1.11</u>	Gly	GGT	<u>1.02</u>	Asn	AAT	0.94	Ser	TCA	<u>1.50</u>
Ala	GCC	0.91	His	CAC	<u>1.01</u>	Pro	CCA	<u>1.24</u>	Ser	TCC	<u>1.01</u>
Ala	GCG	<u>1.04</u>	His	CAT	0.99	Pro	CCC	0.85	Ser	TCG	0.59
Ala	GCT	0.95	Ile	ATA	0.82	Pro	CCG	0.90	Ser	TCT	<u>1.32</u>
Cys	TGC	<u>1.03</u>	Ile	ATC	0.83	Pro	CCT	<u>1.01</u>	Thr	ACA	<u>1.08</u>
Cys	TGT	0.97	Ile	ATT	<u>1.35</u>	Gln	CAA	<u>1.18</u>	Thr	ACC	<u>1.05</u>
Asp	GAC	0.89	Lys	AAA	<u>1.03</u>	Gln	CAG	0.82	Thr	ACG	0.81
Asp	GAT	<u>1.11</u>	Lys	AAG	0.97	Arg	AGA	<u>1.20</u>	Thr	ACT	<u>1.07</u>
Glu	GAA	<u>1.11</u>	Leu	CTA	0.66	Arg	AGG	<u>1.11</u>	Val	GTA	0.72
Glu	GAG	0.89	Leu	CTC	<u>1.05</u>	Arg	CGA	0.80	Val	GTC	0.76
Phe	TTC	0.93	Leu	CTG	0.98	Arg	CGC	0.72	Val	GTG	<u>1.35</u>
Phe	TTT	<u>1.07</u>	Leu	CTT	<u>1.28</u>	Arg	CGG	<u>1.36</u>	Val	GTT	<u>1.17</u>
Gly	GGA	<u>1.07</u>	Leu	TTA	0.94	Arg	CGT	0.81	Tyr	TAC	0.94
Gly	GGC	<u>1.14</u>	Leu	TTG	<u>1.09</u>	Ser	AGC	0.82	Tyr	TAT	<u>1.06</u>
Gly	GGG	0.77	Asn	AAC	<u>1.06</u>	Ser	AGT	0.76			

注:下划线表示 RSCU>1。
Note: The underline mean RSCU>1.

3 讨论

绿豆主要在亚洲国家被广泛种植,因其营养丰富而成为一种重要的粮食作物。当前的研究主要涉及产量和品质改良、抗病育种^[18-19]。绿豆全基因组数据的公布,为绿豆抗病和遗传育种研究奠定了基础^[15]。目前,超过 20 个来自不同物种的 TLPs 具有抗真菌活性^[6-7]。转 TLP 基因的作物显著提高了对病原真菌的抗性和对多种非生物胁迫的耐受性^[8-9,16],因此,对植物 TLP 家族基因的研究受到越来越多的关注。该试验通过生物信息学方法共获得 34 条绿豆 TLPs 序列,所有绿豆 TLPs 成员均具有典型的 THN 功能域和类甜蛋白家族标签,说明该家族基因具有一定的保守性。绿豆 TLP 家族基因结构主要有 4 种基因类型,内含子相位类型多样,基因结构类型较丰富。蛋白 GO 分类显示 91.2% 的蛋白参与胞外作用和多器官发育进程,94.1% 的蛋白参与植物响应应激反应。说明该家族基因容易发生可变剪切,蛋白在

信号肽的引导下在胞外空间参与器官发育和应对胁迫等进程,在植物的生长发育和抵御环境胁迫过程中发挥重要作用。

比较基因组通过对不同种生物的基因组数据及其垂直进化、水平演化过程进行比较研究,可以了解基因的结构及其调控作用。基因组分析显示,大约 9 万年前绿豆发生了异源多倍体的事件^[15]。染色体定位分析发现,绿豆 TLP 家族多个基因存在部分集中分布现象,说明这些基因可能是通过基因复制进化而来的。有 7 个基因簇紧密连锁在 1 号、6 号、10 号和 11 号染色体上,说明这些基因可能是染色体内复制的结果。系统发育分析显示,绿豆 TLPs 家族基因数目和亚家族分类与拟南芥、水稻和杨树类似^[7,12],主要分为 10 个聚类组,说明该家族基因存在高度分化,这与该家族基因功能多样性是一致的。SHATTERS 等^[2]发现水稻和拟南芥 TLP 蛋白存在染色体内和染色体间的复制,单子叶和双子叶植物进化上发生分离后,TLP 基因在 10 个进化枝上发生了

不对称的增加。CAO等^[20]通过对多个物种 TLP 家族基因进化分析显示,串联和片段复制主导了该家族基因的扩张。在物种形成中,串联复制是 TLP 家族不对称扩张的重要机制之一^[7]。该试验结果与 SHATTERS 等^[2]研究一致,即绿豆的 TLP 蛋白也分布于多个支系,且可能存在染色体内和染色体间的复制现象,但与其它物种相比,绿豆 TLP 家族发生的复制事件较少。绿豆 TLP 各聚类组内序列间一致性较高,聚类组间一致性普遍较低,说明不同聚类组基因在植物体内的功能发生了明显的分化。聚类组 5 中的其它物种 TLPs 能响应病原或环境胁迫,如 1Z3Q 和 AT4G11650 具有抗真菌和致敏原活性^[16-17],预示该组中绿豆 TLPs 很可能也具有类似功能,具体原因有待进一步研究。多项研究显示 TLP 转基因植物表现出对真菌病原和环境胁迫的耐受性增强^[5,8],表明该家族蛋白在植物抵抗或耐受环境胁迫过程中发挥了作用,但也有部分研究显示 TLP 转基因作物并未增强植株抗病性^[21],因此在抗病基因育种工程中,既要考虑基因的病原菌抗性专化性,也要考虑与其它防御相关蛋白的协同作用,以及作物生长地区的生态系统复杂性^[7]。绿豆中该家族生物信息学研究将为筛选胁迫抗性相关基因和转基因育种提供理论基础。双子叶植物基因偏好 A/T 作为第三位密码子,单子叶植物偏好 G/C 作为第三位密码子^[22]。双子叶植物绿豆 TLP 家族多数基因 GC3s 小于 50%,偏爱 A 或 T 作为第三位密码子。GC3s 较小基因的 ENC 值也相对较低,表明这些基因密码子偏性较强,表达潜力较强,可能在植物的生命活动中发挥较强作用。ENC 与 GC3s 关联分析显示,ENC 值相对较低的基因主要受正向选择压力的影响,该结果与其表达潜力较强一致,进一步说明其发挥功能的重要性。研究显示,多个物种 TLP 家族基因主要受纯化选择影响^[20]。总体而言,绿豆 TLP 家族基因密码子使用偏性不强,主要受碱基突变和正向选择压力的双重影响。该研究将为 TLP 转基因研究中的密码子优化提供借鉴。

参考文献

- [1] ABAD L R, D'URZO M P, LIU D, et al. Antifungal activity of tobacco osmotin has specificity and involves plasma membrane permeabilization[J]. *Plant Sci* 1996, 118(1): 11-23.
- [2] SHATTERS R G J R, BOYKIN L M, LAPOINTE S L, et al. Phylogenetic and structural relationships of the PR5 gene family reveal an ancient multigene family conserved in plants and select animal taxa[J]. *J Mol Evol* 2006, 63(1): 12-29.
- [3] LIU C, CHENG F, SUN Y, et al. Structure-function relationship of a novel PR-5 protein with antimicrobial activity from soy hulls[J]. *Journal of Agricultural and Food Chemistry*, 2016, 64(4): 948-959.
- [4] BEATRICE C, LINTHORST J M H, CINZIA F, et al. Enhancement of *PR1* and *PR5* gene expressions by chitosan treatment in kiwifruit plants inoculated with *Pseudomonas syringae* pv. *actinidiae* [J]. *European Journal of Plant Pathology*, 2017, 148(1): 163-179.
- [5] ROUT E, NANDA S, JOSHI R K. Molecular characterization and heterologous expression of a pathogen induced *PR5* gene from garlic (*Allium sativum* L.) conferring enhanced resistance to necrotrophic fungi[J]. *European Journal of Plant Pathology*, 2016, 144(2): 345-360.
- [6] GUO J, ZHAO X, WANG H, et al. Expression of the *LePR5* gene from cherry tomato fruit induced by *Cryptococcus laurentii* and the analysis of *LePR5* protein antifungal activity [J]. *Postharvest Biology and Technology*, 2016, 111: 337-344.
- [7] LIU J J, STURROCK R, EKRAMODDOULLAH A K M. The superfamily of thaumatin-like proteins: Its origin, evolution, and expression towards biological function[J]. *Plant Cell Rep* 2010, 29(5): 419-436.
- [8] MISRA R C, SANDEEP M K, KUMAR S, et al. A thaumatin-like protein of *Ocimum basilicum* confers tolerance to fungal pathogen and abiotic stress in transgenic *Arabidopsis* [J]. *Scientific reports*, 2016, 6: 25340.
- [9] ODENY P O. Enhancement of resistance against *Colletotrichum Gloeosporioides* in cassava over-expressing rice thaumatin-like protein (Pr-5) Gene[D]. Kenya: Kenyatta University, 2015.
- [10] FILS-LYCAON B R, WIERSMA P A, EASTWELL K C, et al. A cherry protein and its gene, abundantly expressed in ripening fruit, have been identified as thaumatin-like [J]. *Plant Physiol* 1996, 111(1): 269-273.
- [11] HO V S, WONG J H, NG T B. A thaumatin-like antifungal protein from the emperor banana[J]. *Peptides*, 2007, 28(4): 760-766.
- [12] ZHAO J P, SU X H. Patterns of molecular evolution and predicted function in thaumatin-like proteins of *Populus trichocarpa* [J]. *Planta*, 2010, 232(4): 949-962.
- [13] PETRE B, MAJOR I, ROUHIER N, et al. Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar[J]. *Plant Biology*, 2011, 11(1): 334.
- [14] DAHIYA P K, LINNEMANN A R, van BOEKEL M, et al. Mung bean: Technological and nutritional potential[J]. *Critical Reviews in Food Science and Nutrition*, 2015, 55(5): 670-688.

[1] ABAD L R, D'URZO M P, LIU D, et al. Antifungal activity of tobacco osmotin has specificity and involves plasma membrane

- [15] KANG Y J, KIM S K, KIM M Y, et al. Genome sequence of mungbean and insights into evolution within *Vigna* species[J]. Nature Communications, 2014, 5(5543): 5443.
- [16] LEONE P, MENU-BOUAOUICHE L, PEUMANS W J, et al. Resolution of the structure of the allergenic and antifungal banana fruit thaumatin-like protein at 1.7-Å[J]. Biochimie, 2006, 88(1): 45-52.
- [17] NARASIMHAN M L, BRESSAN R A, D'URZO M P, et al. Unexpected turns and twists in structure/function of PR-proteins that connect energy metabolism and immunity[J]. Advances in Botanical Research, 2009, 51: 439-489.
- [18] ISEMURA T, KAGA A, TABATA S, et al. Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*) [J]. PLoS One, 2012, 7(8): e41304.
- [19] KITSANACHANDEE R, SOMTA P, CHATCHAWAN-KANPHANICH O, et al. Detection of quantitative trait loci for mungbean yellow mosaic India virus (MYMIV) resistance in mungbean (*Vigna radiata* (L.) Wilczek) in India and Pakistan [J]. Breeding Science, 2013, 63(4): 367-373.
- [20] CAO J, LYU Y, HOU Z, et al. Expansion and evolution of thaumatin-like protein (TLP) gene family in six plants[J]. Plant Growth Regulation, 2016, 79(3): 299-307.
- [21] SZWACKA M, KRZYMOWSKA M, OSUCH A, et al. Variable properties of transgenic cucumber plants containing the thaumatin II gene from *Thaumatococcus daniellii* [J]. Acta Physiologiae Plantarum, 2002, 24(2): 173-185.
- [22] TATARINOVA T V, ALEXANDROV N N, BOUCK J B, et al. GC3 biology in corn, rice, sorghum and other grasses [J]. BMC Genomics, 2010, 11(1): 308.

Identification and Bioinformatics Analysis of Thaumatin-like Protein Family in *Vigna radiata*

LIU Chao, HAN Lihong, WANG Haibo, SONG Peibing, TANG Lizhou

(Center for Yunnan Plateau Biological Resources Protection and Utilization, Qujing Normal University/College of Biological Resource and Food Engineering/Key Laboratory of Yunnan Province Universities of the Diversity and Ecological Adaptive Evolution for Animals and Plants on YunGui Plateau, Qujing, Yunnan 655011)

Abstract: In order to understand the basic biological characteristics of mungbean thaumatin-like protein (TLP) family on the genomic level, this study analyzed and identified gene structure and phylogenesis of 34 mungbean TLP family members through genome database and bioinformatics method. The results showed that the *TLP* gene structure classified into four types, *TLP* genes distributed on nine of eleven chromosomes, phylogenesis analysis showed that mungbean TLP family proteins classified into ten evolutionary groups. Most members of group sixth and seventh mainly derived from chromosome eleven. Analysis revealed mungbean TLP family genes preference A or T as the third codon, but generally codon usage bias was not strong. Most of genes belonged to the low expression genes, and affected by mutation and positive selection pressure. This study laid a foundation for cloning of *TLP* genes and further functional dissection in antifungal activity.

Keywords: *Vigna radiata*; thaumatin-like protein; codon usage bias; bioinformatics