

# 枣类甜蛋白基因家族的鉴定与生物信息学分析

刘 潮, 韩利红, 王海波, 高 永, 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:**【目的】从枣全基因组中鉴定类甜蛋白(thaumatin like protein, TLP)基因, 为枣 TLP 基因的功能研究与利用提供参考。【方法】通过生物信息学方法从枣基因组数据库中鉴定 TLP 基因家族成员, 并对基因结构、进化、启动子区顺式作用元件和密码子使用性等进行分析。【结果】共鉴定到 34 个枣 TLP 基因, 位于 9 条染色体上, 分为 4 种结构类型, 基因结构相对较简单; 系统发育进化分析归为 9 个聚类组, 其中聚类组 5 和聚类组 6 中的成员最多; 基因启动子区发现多个与增强基因表达、激素响应和胁迫响应相关的元件; 基因密码子使用偏性弱, 基因进化主要受碱基突变的影响。【结论】枣基因组中含有 34 个 TLP 基因家族成员, 数量相对偏少; 在植物生长、果实发育和抵御胁迫过程中发挥作用; 进化过程中主要受突变选择压力影响。该研究为 TLP 基因家族的功能分析和遗传育种应用提供了参考和借鉴。

**关键词:** 枣; 类甜蛋白; 生物信息学; 密码子使用性

中图分类号: S665.1

文献标志码: A

文章编号: 1009-9980(2018)04-0393-09

## Identification and bioinformatics analysis of thaumatin-like protein gene in Chinese jujube

LIU Chao, HAN Lihong, WANG Haibo, GAO Yong, TANG Lizhou\*

(Center for Yunnan Plateau Biological Resources Protection and Utilization · 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 Normal University, Qujing 655011, Yunnan, China)

**Abstract:** 【Objective】Thaumatococcal proteins (TLPs) include thaumatin, osmotin, zeamatin, and related proteins, which belong to the pathogenesis-related family (PR5) of plant defense proteins and are encoded by a multiple gene family. TLPs play a vital role in the enhancement of plant resistance to biotic and abiotic stresses due to involvement in defense system. Like other PRs, TLPs belong to a general plant stress response pathway rather than being specific to distinct stresses. TLPs possess antifungal, glucanase, allergens activities. Fruit diseases and storage would affect yield and fruit quality of Chinese jujube (*Ziziphus jujuba* Mill.). A genome-wide identification was performed to investigate the TLP gene number, evolution and gene expression ability in Chinese jujube, through comparison with those of *Arabidopsis thaliana*, *Oryza sativa* and *Populus trichocarpa*. 【Methods】TLP family member were identified based on the genome database of Chinese jujube. TLP sequences of *A. thaliana* were used as queries to obtain the member of the TLP family through searching the database of Chinese jujube protein on NCBI website. The sequences were then confirmed as encoding TLP for the presence of a thaumatin domain signature by SMART database searches. Chromosomal locations were determined using the annotation information of TLP genes genome data. The physiological and biochemical characteristics of protein were predicted using ProtParam at ExpASY server. The schematic diagrams of the gene structure were drawn using Gene Structure Display Server. Multiple sequence alignments of full-length protein

收稿日期: 2017-11-13 接受日期: 2018-01-27

基金项目: 国家自然科学基金(31460179、31760103); 云南省地方本科高校基础研究联合专项(2017FH001-034)

作者简介: 刘潮, 男, 博士, 讲师, 研究方向为分子植物病理学。Tel: 18213511867, E-mail: liuchao@mail.qjnu.edu.cn

\*通信作者 Author for correspondence. Tel: 0874-8987890, E-mail: tanglizhou@163.com

sequences were accomplished using ClustalW. And the neighbor-joining (NJ) method in MEGA 5 software was used to carry out phylogenetic tree analyses of the Chinese jujube TLPs with the default evolutionary model, and bootstraps with 1 000 replicates were used to test the convergence of phylogenetic trees. The *cis*-acting elements distribution of the gene promoter was analyzed by PlantCARE database. Codon usage parameters were analysed using CodonW software. Relative synonymous codon usage were calculated using Cusp at the EMBOSS explorer server. The optimal codon was defined with the abundance values relative synonymous codon usage (RSCU)  $>1$  in the high expression genes group [the first 10% genes of low effective number of codons (ENC) value] and the abundance values  $RSCU < 1$  in the low expression genes group (the last 10% genes of high ENC value). **【Results】**Thirty-four genes of the TLP family were identified in Chinese jujube. They were located on nine chromosomes. Four types of gene structures were found. The numbers of genes containing 1, 2, 3 and 4 exons were 10, 11, 11, and 2, respectively. Gene structure analysis indicated that TLP family gene structures were relatively simple in Chinese jujube, implying that the gene structures were relatively stable, and alternative splicing was rare during replication. Phylogenetic analysis indicated that Chinese jujube TLP gene family was classified into nine clusters, and the cluster five and six had the largest number of TLPs, suggesting that partial TLP members were involved in the tandem and segmental duplication events. The disequilibrium of the distribution of TLP family members in each cluster might be related to gene functions. Multiple *cis*-acting elements related to strengthen expression, hormone and stress response were found on the promoter region of *TLP* genes, indicating that these genes would play a role in plant defense against stresses. Codon analysis showed that the mean value of CAI was 0.235. ENC values were higher than 45.00 and the distribution of GC3s value was dispersed. Eleven optimal codons were found, and the number of ending with A, T, G and C were 3, 3, 2 and 3, respectively. Codon analysis indicated that the gene codon usage was weak, and the gene evolution was mainly affected by gene mutation. Purifying selection might play an important role in function maintenance of this family. **【Conclusion】**This was the first comprehensive experimental survey of the TLP gene family in Chinese jujube. TLP family genes would be conservative in evolutionary process, being involved in plant development and stress response regulation. The evolution of Chinese jujube TLP family genes might be mainly influenced by base mutation pressure, and few genes might be affected by natural selection pressure. These investigations would provide theoretical support for functional analysis and genetic breeding application of this gene family in Chinese jujube.

**Key words:** *Ziziphus jujuba*; Thaumatin-like protein; Bioinformatics; Codon usage

枣 (*Ziziphus jujuba* Mill.) 为鼠李科枣属落叶小乔木, 原产中国, 现在世界范围内广泛栽培<sup>[1]</sup>。枣中富含维生素和糖类, 可鲜食、制干或制成蜜饯和果脯, 也是重要的食品工业原料。枣树病虫害主要有枣锈病、枣疯病、炭疽病、黑斑病、轮纹病和枣瘿蚊、椿象等<sup>[2-3]</sup>, 栽培管理、气候和品种等因素决定了枣树病害的发生与流行。果实采摘后, 因自身生理、低温或病原等因素也容易导致果实品质下降或软腐溃烂等<sup>[4]</sup>。多数枣树或果实病害由真菌性病原引起, 不同枣树品种间抗病性有明显差异, 选育抗病品种是防治枣树病害的有效方法。

类甜蛋白 (thaumatin like protein, TLP) 属于病程相关蛋白 (pathogenesis related protein, PR) 第5家族, 能被多种生物或非生物信号诱导表达<sup>[5-7]</sup>, TLP具有抗真菌活性<sup>[8-9]</sup>、抗冷活性<sup>[10-11]</sup>等, 在植物的生长发育和抵御胁迫过程中发挥作用。研究表明, *TLP*基因来自于大约10亿年前的植物、动物和真菌的共同祖先<sup>[12]</sup>。进化过程中, 由于遗传和环境因素导致不同物种中TLP家族基因的数量和结构发生明显的改变, 在植物中的功能也明显不同。基因启动子控制基因表达的起始时间和表达程度, 启动子区顺式作用元件分析能为预测基因表达水平和效率提供参

考,对植物抗胁迫研究有重要的理论和应用价值。基因密码子组成对基因表达具有重要的调节作用,与翻译的效率和准确性有关<sup>[13]</sup>,分析密码子使用性对预测基因表达、功能以及转基因研究具有重要的应用价值<sup>[14-15]</sup>。对 TLP 家族基因特征及密码子组成的研究,将为枣功能基因挖掘利用提供借鉴。

多种植物中的 TLP 蛋白已被鉴定<sup>[16-18]</sup>。目前,枣 (*Z. jujuba*) 基因组数据已公布<sup>[19]</sup>,但有关 TLP 家族基因方面的研究尚未见报道。笔者从全基因组水平上,鉴定枣 TLP 家族基因,并对基因结构、蛋白聚类、启动子区顺式作用元件、密码子使用性进行分析,希望阐明 TLP 家族基因特征,为进一步利用 TLP 家族基因培育优良枣抗性品种奠定基础。

## 1 材料和方法

### 1.1 基因序列获取与鉴定

以拟南芥 TLP 蛋白序列为探针,搜索 GenBank 中枣蛋白数据库,候选蛋白序列在 SMART 数据库 (<http://smart.embl-heidelberg.de/>) 中对蛋白功能域进行确认。蛋白生理生化特征通过 ExPasy (<http://www.expasy.org/tools/>) 预测。

### 1.2 基因结构和蛋白聚类分析

从 GenBank 数据库中下载枣 TLP 蛋白对应的基因序列和编码序列 (coding sequence, CDS)。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。应用 MEGA 5.0 软件,采用邻接法构建系统聚类树,泊松校验法计算距离。

### 1.3 启动子特征分析

通过 GenBank 数据库获取枣 TLP 家族基因转录起始位点上游 1 kb 序列,通过 PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>) 数据库进行基因启动子区顺式作用元件分析。

### 1.4 密码子偏性分析

采用软件 CodonW 对 TLP 家族基因编码序列密码子的密码子适应指数 (codon adaptation index, CAI)、有效密码子数 (effective number of codons, ENC)、密码子的第 3 位的 G+C 含量 (GC3s) 等进行分析,并绘制 ENC 与 GC3s 的关联分布图<sup>[20]</sup>。使用 EMBOSS explorer 网站对同义密码子相对使用度 (relative synonymous codon usage, RSCU) 进行分析。选择 ENC 值前后各 10% 作为低表达和高表达

基因,分别计算 2 组基因中 TLP 基因密码子的 RSCU 值,分析最优密码子 (optimal codon)。当  $\Delta RSCU > 0.3$ ,且在高表达组中  $RSCU > 1$ 、在低表达组中  $RSCU < 1$ ,可确定该密码子为最优密码子<sup>[21]</sup>。

## 2 结果与分析

### 2.1 枣 TLP 家族鉴定

经分析共鉴定到 34 个含有典型 thaumatin 功能域的枣 TLP 家族基因 (表 1),分布在 9 条染色体上,其中除 1 号和 2 号染色体仅有 1 个基因外,其他 7 条染色体上均有 3~5 个基因,另外还包含 5 个未定位到具体染色体上的基因。理化性质分析显示,蛋白质氨基酸个数为 139~422,等电点为 4.19~9.39,55.9% 为酸性氨基酸,70.6% 的蛋白质总平均疏水指数为负值,属于亲水蛋白。

### 2.2 基因结构及聚类分析

基因结构分析 (图 1) 显示,枣 TLP 基因含有 4 种结构类型,基因结构相对较简单,内含子数最多不超过 3 个,其中不含内含子、含 1、2 和 3 个内含子的基因数目分别为 10、11、11、2。聚在一枝上的支持率较高的基因结构类型一致,如编码 XP\_015896205.1、XP\_015896184.1、XP\_015896206.1 的基因均仅含 1 个外显子,且均位于 11 号染色体上;编码 XP\_015898569.1、XP\_015898553.1、XP\_015898612.1、XP\_015898528.1 的基因均含有 1 个内含子,内含子相位均为 1 型,均位于 12 号染色体上,且在染色体上的位置相近;编码 XP\_015876845.1、XP\_015876843.1、XP\_015896956.1 的基因均含有 1 个内含子,内含子相位均为 1 型,基因分别位于 3 号和 11 号染色体上。这些基因可能是染色体间复制的结果,可能来源于共同的祖先基因。

为深入分析枣 TLP 家族与其他物种同源基因的进化关系,分别选取枣 34 个 TLP 基因以及拟南芥、水稻和杨树 TLP 基因各 10 个 (10 个聚类组中每组选取 1 个代表),构建系统聚类树 (图 2)。发现 10 个聚类组中,仅聚类组 4 没有枣 TLP 家族成员,另有编码 XP\_015882090.1、XP\_015882088.1、XP\_015882091.1 的基因未归入任何聚类组中。各聚类组中的成员数目分布很不均衡,聚类组 5 和聚类组 6 中的成员最多,均为 8 个,且同一聚类组中的基因位于相同的染色体上,这 2 组中的基因在进化过程中可能通过染色体间复制发生了特异性扩张,这些基因在植物中

表1 枣 TLP 蛋白家族信息

Table 1 Information of TLP protein family in Chinese jujube

登录号 Accession No.	染色体位置 Chromosome location	氨基酸长度 Sequence length/aa	蛋白分子量 Molecular weight/ku	等电点 pI	总平均疏水指数 GRAVY
XP_015889970.1	Ch 1:1 935 834~1 937 378	234	25.320	8.90	-0.209
XP_015874069.1	Ch 2:9 423 908~9 424 972	251	27.188	7.87	-0.230
XP_015878696.1	Ch 3:26 477 032~26 479 497	369	39.382	4.63	-0.314
XP_015878694.1	Ch 3:26 477 032~26 479 497	400	42.685	4.56	-0.247
XP_015876845.1	Ch 3:7 827 937~7 829 062	245	25.956	8.51	-0.259
XP_015876843.1	Ch 3:7 827 937~7 829 062	283	30.814	8.83	-0.282
XP_015882732.1	Ch 5:11 539 266~11 540 291	244	26.637	9.06	-0.419
XP_015882088.1	Ch 5:6 222 048~6 222 545	165	17.795	5.84	-0.338
XP_015882090.1	Ch 5:6 235 357~6 244 141	286	30.934	5.48	-0.030
XP_015882091.1	Ch 5:6 264 743~6 265 162	139	14.751	9.39	0.088
XP_015885619.1	Ch 6:11 041 723~11 043 451	280	30.429	8.49	0.049
XP_015884662.1	Ch 6:3 958 275~3 960 364	248	25.506	4.55	-0.030
XP_015884848.1	Ch 6:5 099 877~5 102 114	318	33.391	4.46	-0.295
XP_015889659.1	Ch 8:6 195 379~6 198 031	266	27.727	4.60	-0.185
XP_015889658.1	Ch 8:6 195 379~6 198 031	295	30.857	4.67	0.057
XP_015889656.1	Ch 8:6 205 620~6 207 217	305	31.916	4.72	0.064
XP_015893386.1	Ch 9:19 985 470~19 987 790	352	35.961	4.19	0.164
XP_015893387.1	Ch 9:19 985 470~19 987 790	351	35.875	4.19	0.001
XP_015893384.1	Ch 9:19 999 024~20 002 847	311	33.333	8.53	0.012
XP_015893385.1	Ch 9:19 999 024~20 002 847	289	30.948	8.64	-0.076
XP_015893658.1	Ch 9:22 790 288~22 791 952	291	31.217	8.04	-0.128
XP_015896956.1	Ch 11:11 644 960~11 646 228	245	25.938	7.78	-0.084
XP_015896184.1	Ch 11:1 526 958~1 527 743	240	26.117	8.32	-0.097
XP_015896206.1	Ch 11:1 573 406~1 574 311	243	26.519	5.20	-0.177
XP_015896205.1	Ch 11:1 576 574~1 577 595	224	24.294	6.26	0.003
XP_015898569.1	Ch 12:3 962 084~3 963 372	243	25.579	4.60	-0.163
XP_015898553.1	Ch 12:3 967 258~3 968 536	242	25.505	4.62	-0.224
XP_015898612.1	Ch 12:3 979 558~3 980 823	241	25.298	4.93	0.067
XP_015898528.1	Ch 12:3 988 566~3 990 138	250	26.282	5.68	-0.042
XP_015902780.1	Unplaced scaffold:130 584~131 267	227	24.631	8.12	0.080
XP_015867797.1	Unplaced scaffold:36 463~42 410	420	46.109	5.38	-0.264
XP_015902782.1	Unplaced scaffold:37 735~44 885	422	46.574	4.83	-0.057
XP_015902783.1	Unplaced scaffold:73 868~74 772	225	24.503	7.83	-0.441
XP_015902784.1	Unplaced scaffold:92 050~92 822	227	24.808	8.32	-0.025

的功能值得进一步研究。

### 2.3 基因启动子区特征分析

为了解枣 TLP 家族基因的表达潜力及对胁迫的响应能力,使用 PlantCARE 在线数据库对基因启动子区顺式作用元件进行分析,发现多个与增强基因表达、激素响应和胁迫响应相关的元件(表2)。所有基因启动子区均含有多个 CAAT-box 和 TATA-box,部分基因启动子区含有 1~3 个 Py-rich,说明这些基因,尤其含 Py-rich 的基因具有较强的表达潜力。参与植物生命进程的激素主要为生长素、脱落酸、赤霉素、乙烯、水杨酸、茉莉酸等,18 个枣 TLP 家族基因启动子区含有脱落酸响应元件,多数基因含有 2 种或 2 种以上激素响应元件,其中编码 XP\_015889656.1 的基因启动子区含有 5 类激素响应元件,这些基因可能参与了不同的激素信号通路。植

物生长的环境复杂多变,往往面临高温、干旱、低温、病原等多种胁迫,多数枣 TLP 家族基因启动子区含有 1 种以上的胁迫响应元件,启动子区中热激响应元件、受伤和真菌激发子响应元件、防御和胁迫相关响应元件、干旱响应元件和低温响应元件的基因数目分别为 19、14、14、13 和 8 个,这些基因可能在植物抵御胁迫过程中发挥作用。

### 2.4 密码子使用性分析

对枣 TLP 家族基因密码子使用性分析发现,基因的 CAI 均值为 0.235,ENC 值均高于 45(表3)。一般 ENC 值高于 35 表示基因表达对密码子的使用偏性较弱,说明枣 TLP 家族基因密码子使用偏性弱。基因 GC3s 为 0.300~0.600,分布较分散,GC3s 分布反映了基因所受的选择压力,GC3s 分布范围越大,表明基因密码子使用偏性受碱基突变的影响越大,





表 2 枣 TLP 家族基因启动子区顺式作用元件信息

Table 2 Information of putative *cis*-acting elements in the promoter region of TLP genes in Chinese jujube

登录号 Accession No.	个数 Number														
	CAAT- box	TATA- box	Py- rich	TGA- element	TCA- element	CGTCA- motif	ERE	ABRE	GARE-motif	P- box	HSE	LTR	MBS	W- Box	TC- rich
XP_015867797.1	26	67											1	1	2
XP_015874069.1	23	64	3					2	1		1	2			
XP_015876843.1	27	72	2								3		1	1	3
XP_015876845.1	27	72	2		1						3		1	1	2
XP_015878694.1	16	36				2				2		1	1		
XP_015878696.1	16	36				2				2		1	1		
XP_015882088.1	25	33					1	1							
XP_015882090.1	23	102	1				1	1	2		4				
XP_015882091.1	21	38			1				1						
XP_015882732.1	23	33			1					1	1	1	1		2
XP_015884662.1	14	31				2		5	1	2	1	1	1	1	
XP_015884848.1	27	56			1			5						2	
XP_015885619.1	22	60	1		2			1	1	1	2				1
XP_015889656.1	25	29	2	1	1			2	2	1	1	1		1	1
XP_015889658.1	18	49	2	1	1	3								1	
XP_015889659.1	18	49	2	1	1	3								1	
XP_015889970.1	19	102												1	
XP_015893384.1	22	35						1	2						
XP_015893385.1	22	35						1	2						
XP_015893386.1	13	37		2	1					1	3	1	2		1
XP_015893387.1	13	37		2	1					1	3	1	2		1
XP_015893658.1	13	80	1		4	1		1							
XP_015896184.1	18	94	1	1		2	1	1					1	1	
XP_015896205.1	16	53		1				1			2		2		1
XP_015896206.1	21	87			1			1						1	1
XP_015896956.1	26	52	2		1				1		1			1	
XP_015898528.1	18	39	1	1			1	3	1		2				
XP_015898553.1	26	116	1								6				
XP_015898569.1	23	49					1	1					2		3
XP_015898612.1	26	73	1			2					3				2
XP_015902780.1	11	46		1				1			2			2	
XP_015902782.1	25	39			1						1		1		1
XP_015902783.1	28	69	3			1		2			1				1
XP_015902784.1	20	64					1	1			2			2	

注: CAAT-box. 启动增强元件; TATA-box. 转录起始区-30 bp 核心启动子元件; Py-rich. 高转录水平相关顺式作用元件; TGA-element. 生长素响应元件; TCA-element. 水杨酸响应元件; CGTCA-motif. 茉莉酸响应元件; ERE. 乙烯响应元件; ABRE. 脱落酸响应元件; GARE-motif. 赤霉素响应元件; P-box. 赤霉素响应元件; HSE. 热激响应元件; LTR. 参与低温反应的顺式作用元件; MBS. 干旱响应 MYB 结合位点; W-box. 受伤和真菌激发子响应元件; TC-rich. 防御和胁迫相关响应元件。

Note: CAAT-box. Common *cis*-acting element in promoter and enhancer regions; TATA-box. Core promoter element around -30 bp of transcription start; Py-rich. *Cis*-acting element conferring high transcription levels; TGA-element. Auxin-responsive element; TCA-element. *Cis*-acting element involved in salicylic acid responsiveness; CGTCA-motif. *Cis*-acting regulatory element involved in the MeJA-responsiveness; ERE. Ethylene-responsive element; ABRE. *Cis*-acting element involved in the abscisic acid responsiveness; GARE-motif. Gibberellin-responsive element; P-box. Gibberellin-responsive element; HSE. *Cis*-acting element involved in heat stress responsiveness; LTR. *Cis*-acting element involved in low-temperature responsiveness; MBS. MYB binding site involved in drought-inducibility; W-box. Fungal elicitor responsive element; TC-rich. *Cis*-acting element involved in defense and stress responsiveness.

表3 枣 TLP 家族基因密码子使用特性

Table 3 Characterization of codon usage of TLP genes in Chinese jujube

登录号 Accession No.	CAI	ENC	GC3s	登录号 Accession No.	CAI	ENC	GC3s	登录号 Accession No.	CAI	ENC	GC3s
XP_015902783.1	0.265	50.03	0.595	XP_015878696.1	0.175	48.65	0.323	XP_015884662.1	0.253	60.27	0.607
XP_015896205.1	0.275	52.54	0.571	XP_015878694.1	0.181	48.37	0.329	XP_015898553.1	0.273	54.76	0.527
XP_015902784.1	0.216	57.38	0.541	XP_015889656.1	0.227	61.00	0.572	XP_015898528.1	0.289	50.64	0.510
XP_015896184.1	0.258	53.91	0.487	XP_015889659.1	0.260	59.56	0.541	XP_015876845.1	0.214	56.34	0.534
XP_015896206.1	0.244	53.61	0.564	XP_015889658.1	0.246	55.09	0.526	XP_015876843.1	0.198	58.16	0.514
XP_015902780.1	0.229	58.84	0.550	XP_015898569.1	0.273	61.00	0.496	XP_015889656.1	0.191	49.46	0.347
XP_015867797.1	0.233	54.99	0.484	XP_015893658.1	0.186	51.80	0.399	XP_015882090.1	0.214	55.37	0.505
XP_015902782.1	0.246	54.56	0.499	XP_015882732.1	0.246	54.46	0.559	XP_015882088.1	0.219	59.03	0.453
XP_015889970.1	0.224	53.90	0.415	XP_015884848.1	0.231	54.26	0.539	XP_015882091.1	0.204	56.07	0.522
XP_015885619.1	0.217	58.63	0.487	XP_015893386.1	0.240	56.48	0.516	XP_015874069.1	0.350	45.84	0.738
XP_015893384.1	0.209	58.38	0.569	XP_015893387.1	0.242	57.04	0.518				
XP_015893385.1	0.200	57.73	0.585	XP_015898612.1	0.267	51.49	0.506				

注: CAI. 密码子适应指数; ENC. 有效密码子数; GC3s. 密码子的第3位的 G+C 含量。

Note: CAI. Codon adaptation index; ENC. Effective number of codons; GC3s. GC content on the 3rd site of synonymous codon.

显示枣 TLP 家族基因密码子主要受碱基突变的影响。

ENC 与 GC3s 关联分析中, 多数基因分布在标准曲线附近(图3)。GC3s 的分布反映植物所受的选择压力, 枣 TLP 家族基因密码子, GC3 值分布较广泛, ENC 值偏大, 表明枣 TLP 家族基因密码子偏性弱, 主要受碱基突变的影响。

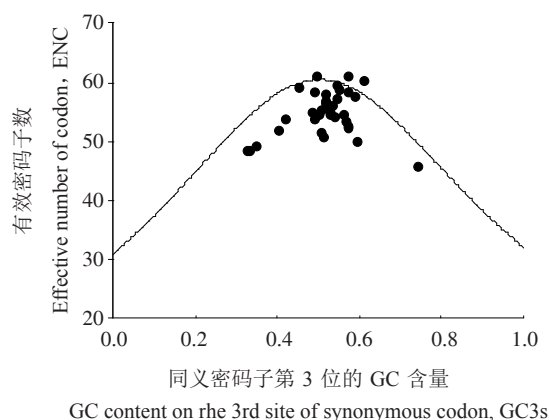


图3 TLP 家族基因 ENC 与 GC3s 的关系  
Fig. 3 Correlative analysis of ENC and GC3s of TLP genes in Chinese jujube

同义密码子相对使用度(RSCU)是同义密码子实际使用数与理论使用数的比值。分析发现, 枣 TLP 家族基因密码子 RSCU 介于 0.42~1.73, 数值变动范围较小, 含有 11 个最优密码子, 以 A、T、G 和 C 结尾的数目分别为 3、3、2 和 3(表4), 说明枣 TLP 家族基因最优密码子使用偏性弱。

### 3 讨论

TLP 家族在植物的生长发育和应对胁迫过程中发挥作用。研究发现, 不同物种中 TLP 家族基因的数目有很大差异, 如拟南芥(*Arabidopsis thaliana*)为 28 个, 水稻(*Oryza sativa*)为 44 个, 胡萝卜(*Daucus carota*)为 32 个, 桃(*Prunus persica*)为 38 个, 杨树(*Populus trichocarpa*)为 55 个, 玉米(*Zea mays*)为 49 个, 火炬松(*Pinus taeda*)甚至高达 87 个<sup>[16-18, 22]</sup>。TLP 家族基因的数目除了与植物基因组进化过程中的染色体重组事件有关外, 与植物生长所面临的环境胁迫也有重要关系<sup>[23]</sup>。笔者分析鉴定了 34 个枣 TLP 家族成员, 相对数量偏少, 在各染色体上分布数量不同, 基因结构相对简单, 多数编码亲水性蛋白, 说明这些基因结构相对稳定, 复制时较难发生可变剪切, 在植物中的功能相对较稳定。

随着基因组学的发展, 越来越多的动植物基因组被公布, 不同物种中同一家族基因的进化、表达及密码子使用性研究, 对植物基因功能的研究和品种改良具有重要的意义。拟南芥、水稻和杨树 TLP 家族基因均归为 10 个聚类组, 且各个聚类组中的成员数目不同<sup>[16, 18]</sup>。分析发现, 枣 TLP 家族在聚类组 5 和聚类组 6 中成员最多, 这与拟南芥、水稻和杨树类似, 说明这些聚类组中的基因可能在植物的生命进程中发挥了重要的作用<sup>[16, 18]</sup>, 聚类组 4 中缺少枣 TLP 家族成员, 可能是该聚类组中的基因发生了丢失现象。同一聚类组中的多数枣 TLP 家族基因内含子数

表4 枣 TLP 家族基因同义密码子使用情况

Table 4 Usage of synonymous codon of TLP in Chinese jujube

氨基酸 Amino acid	密码子 Codon	RSCU	氨基酸 Amino acid	密码子 Codon	RSCU	氨基酸 Amino acid	密码子 Codon	RSCU	氨基酸 Amino acid	密码子 Codon	RSCU
丙氨酸 Ala	GCA**	1.08	组氨酸 His	CAC***	1.01	谷氨酰胺 Gln	CAA**	1.13	苏氨酸 Thr	ACA	1.00
丙氨酸 Ala	GCC	1.02	组氨酸 His	CAT	0.99	谷氨酰胺 Gln	CAG	0.87	苏氨酸 Thr	ACC	1.32
丙氨酸 Ala	GCG	0.57	赖氨酸 Lys	AAA	1.05	精氨酸 Arg	AGA	1.72	苏氨酸 Thr	ACG	0.66
丙氨酸 Ala	GCT	1.34	赖氨酸 Lys	AAG	0.95	精氨酸 Arg	AGG	1.14	苏氨酸 Thr	ACT	1.02
半胱氨酸 Cys	TGC	1.25	亮氨酸 Leu	CTA	0.42	精氨酸 Arg	CGA	1.05	缬氨酸 Val	GTA	0.43
半胱氨酸 Cys	TGT	0.75	亮氨酸 Leu	CTC*	1.73	精氨酸 Arg	CGC	0.68	缬氨酸 Val	GTC**	1.04
天冬氨酸 Asp	GAC	0.92	亮氨酸 Leu	CTG**	0.73	精氨酸 Arg	CGG	0.72	缬氨酸 Val	GTG	1.31
天冬氨酸 Asp	GAT*	1.08	亮氨酸 Leu	CTT	1.05	精氨酸 Arg	CGT	0.70	缬氨酸 Val	GTT	1.21
谷氨酸 Glu	GAA	0.84	亮氨酸 Leu	TTA	0.64	丝氨酸 Ser	AGC	1.24	异亮氨酸 Ile	ATA**	0.83
谷氨酸 Glu	GAG*	1.16	亮氨酸 Leu	TTG	1.42	丝氨酸 Ser	AGT	0.83	异亮氨酸 Ile	ATC	1.24
苯丙氨酸 Phe	TTC	1.25	天冬酰胺 Asn	AAC	1.16	丝氨酸 Ser	TCA	0.91	异亮氨酸 Ile	ATT	0.93
苯丙氨酸 Phe	TTT	0.75	天冬酰胺 Asn	AAT	0.84	丝氨酸 Ser	TCC	1.28			
甘氨酸 Gly	GGA	1.16	脯氨酸 Pro	CCA	1.23	丝氨酸 Ser	TCG	0.83			
甘氨酸 Gly	GGC	1.00	脯氨酸 Pro	CCC	0.71	丝氨酸 Ser	TCT*	0.92			
甘氨酸 Gly	GGG	0.76	脯氨酸 Pro	CCG	0.92	酪氨酸 Tyr	TAC	0.88			
甘氨酸 Gly	GGT	1.07	脯氨酸 Pro	CCT**	1.15	酪氨酸 Tyr	TAT	1.12			

注: \*、\*\*、\*\*\*表示最优密码子,其中\*表示 $\Delta RSCU > 0.3$ ; \*\*表示 $\Delta RSCU > 0.6$ ; \*\*\*表示 $\Delta RSCU > 1$ 。

Note: \*, \*\*, \*\*\* indicate optimal codon, \* indicates  $\Delta RSCU > 0.3$ ; \*\* indicates  $\Delta RSCU > 0.6$ ; \*\*\* indicates  $\Delta RSCU > 1$ .

目和相位一致,这些基因多数来源于同一染色体,少数来自不同染色体,可能发生了基因的染色体间或染色体内复制,是染色体复制事件或进化过程中基因复制的结果。研究还发现,枣 TLP 家族基因启动子区含有大量与激素和胁迫响应有关的顺式作用元件,不同的成员含有的种类和数量不同,这可能与该家族基因功能的复杂性有关,说明该家族基因在枣树生长、果实发育和抵御胁迫过程中发挥了多种功能。

基因密码子具有简并性,由于物种进化和生存环境压力的不同,密码子第3位上碱基使用的概率不同。对密码子使用性分析显示,多数单子叶植物 *actin* 基因 ENC 值相对较低,而多数双子叶植物 *actin* 基因 ENC 值相对较高<sup>[24]</sup>。笔者研究表明,多数枣 TLP 家族基因密码子 ENC 值偏高,与桑树 WRKY 转录因子家族基因类似<sup>[23]</sup>,而明显高于玉米 *waxy* 基因<sup>[25]</sup>,说明物种在很大程度上决定了密码子的使用性。GC3s 数值的分布与基因所受的压力有关,GC3s 分布越集中,表示第3位密码子使用性越一致,基因偏好使用 G 或 C 作为第3位密码子,GC3s 值越高,反之越低。高的 GC3s 值与高的甲基化、高表达、高的基因上游 TATA-box 数和胁迫相关基因数等相关,使基因能够通过甲基化承载更多的基因突变,对物种的进化具有重要意义<sup>[26]</sup>。枣 TLP 家族基因 GC3s 值居于中等水平,且分布分散,表明基因的密码子使用偏性较弱,基因受到的正向选择压力较小,

同义密码子和最优密码子的结果进一步说明,枣 TLP 家族基因进化主要受突变选择压力影响。

#### 参考文献 References:

- [1] LIU M J, LIU P, LIU G N. Advances of research on germplasm resources of Chinese jujube[J]. Acta Horticulturae, 2013, 993: 15-20.
- [2] 王亚萍,王贵禧,李艳菊. 鲜枣贮藏期病害研究进展[J]. 果树学报, 2005, 22(4): 376-380.  
WANG Yaping, WANG Guixi, LI Yanju. Advances in research on fruit diseases of Chinese jujube during storage[J]. Journal of Fruit Science, 2005, 22(4): 376-380.
- [3] 范晓栋,符泽,钟聪慧,刘心心,黄家凤,高峰. 新疆阿拉尔垦区红枣黑斑病发病规律[J]. 北方园艺, 2017(19): 35-41.  
FAN Xiaodong, FU Ze, ZHONG Conghui, LIU Xinxin, HUANG Jiafeng, GAO Feng. Incidence mechanism of black spot disease of jujube in Xinjiang Alar reclamation area[J]. Northern Horticulture, 2017(19): 35-41.
- [4] 肖黎斌,韩军岐,张润光,王小纪,郝剑,杨红军,张有林. 鲜枣采后生理、贮藏病害及保鲜技术研究进展[J]. 陕西农业科学, 2016, 62(4): 87-91.  
XIAO Libin, HAN Junqi, ZHANG Runguang, WANG Xiaoji, HAO Jian, YANG Hongjun, ZHANG Youlin. Research progress on postharvest physiology, storage disease and fresh keeping technology of fresh jujube[J]. Shaanxi Journal of Agricultural Sciences, 2016, 62(4): 87-91.
- [5] KIM S H, LEE J R, KIM S R. Molecular characterization of a fruit-preferential thaumatin-like gene from apple (*Malus domestica* cv. Fuji)[J]. Journal of Plant Biology, 2003, 46(1): 52-58.



- [6] WANG L, YANG L, ZHANG J, DONG J, YU J, ZHOU J, ZHUGE Q. Cloning and characterization of a thaumatin-like protein gene PeTLP in *Populus deltoids* × *P. euramericana* 'Nanlin895' [J]. *Acta Physiologiae Plantarum*, 2013, 35(10): 2985-2998.
- [7] 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.
- [8] RAMOS M V, DE OLIVEIRA R S B, PEREIRA H M, MORENO F B, LOBO M D, REBELO L M, BRANDÃO-NETO J, DE SOUSA J X, MONTEIRO-MOREIRA A C O, FREITAS C D T, GRANGEIRO T B. Crystal structure of an antifungal osmotin-like protein from *Calotropis procera* and its effects on *Fusarium solani* spores, as revealed by atomic force microscopy: insights into the mechanism of action[J]. *Phytochemistry*, 2015, 119: 5-18.
- [9] LIU C, CHENG F, SUN Y, MA H, YANG X. 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.
- [10] D' ANGELI S, ALTAMURA M M. Osmotin induces cold protection in olive trees by affecting programmed cell death and cytoskeleton organization[J]. *Planta*, 2007, 225(5): 1147-1163.
- [11] DAGAR A, FRIEDMAN H, LURIE S. Thaumatin-like proteins and their possible role in protection against chilling injury in peach fruit[J]. *Postharvest Biology and Technology*, 2010, 57(2): 77-85.
- [12] 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 Reports*, 2010, 29(5): 419-436.
- [13] CARLINI D B, CHEN Y, STEPHAN W. The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes *Adh* and *Adhr*[J]. *Genetics*, 2001, 159(2): 623-633.
- [14] PAN L L, WANG Y, HU J H, DING Z T, LI C. Analysis of codon use features of stearoyl-acyl carrier protein desaturase gene in *Camellia sinensis*[J]. *Journal of Theoretical Biology*, 2013, 334(19): 80.
- [15] PEK H B, KLEMENT M, ANG K S, CHUNG B K S, OW D S W, LEE D Y. Exploring codon context bias for synthetic gene design of a thermostable invertase in *Escherichia coli*[J]. *Enzyme and Microbial Technology*, 2015, 75: 57-63.
- [16] SHATTERS R G, BOYKIN L M, LAPOINTE S L, HUNTER W B, WEATHERSBEE A A. Phylogenetic and structural relationships of the PR5 gene family reveal an ancient multigene family conserved in plants and select animal taxa[J]. *Journal of Molecular Evolution*, 2006, 63(1): 12-29.
- [17] 刘潮, 韩利红, 王海波, 宋培兵, 唐利洲. 胡萝卜类甜蛋白家族鉴定与生物信息学分析[J]. *中国蔬菜*, 2017(2): 38-44.  
LIU Chao, HAN Lihong, WANG Haibo, SONG Peibing, TANG Lizhou. Identification and bioinformatics analysis of thaumatin-like protein family in *Daucus carota*[J]. *China Vegetables*, 2017(2): 38-44.
- [18] 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.
- [19] LIU M J, ZHAO J, CAI Q L, LIU G C, WANG J R, ZHAO Z H, LIU P, DAI L, YAN G J, WANG W J, LI X S, CHEN Y, SUN Y D, LIU Z G, LIN M J, XIAO J, CHEN Y Y, LI X F, WU B, MA Y, JIAN J B, YANG W, YUAN Z, SUN X C, WEI Y L, YU L L, ZHANG C, LIAO S G, HE R J, GUANG X M, WANG Z, ZHANG Y Y, LUO L H. The complex jujube genome provides insights into fruit tree biology[J]. *Nature Communications*, 2014, 5: 5315.
- [20] WRIGHT F. The "effective number of codons" used in a gene [J]. *Gene*, 1990, 87(1): 23-29.
- [21] DURET L, MOUCHIROUD D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*[J]. *Proceedings of the National Academy of Sciences*, 1999, 96(8): 4482-4487.
- [22] CAO J, LÜ Y, HOU Z, LI X, DING L. Expansion and evolution of thaumatin-like protein (TLP) gene family in six plants[J]. *Plant Growth Regulation*, 2016, 79(3): 299-307.
- [23] 刘潮, 韩利红, 宋培兵, 王德琴, 王海波, 唐利洲. 桑树 WRKY 转录因子的全基因组鉴定及生物信息学分析[J]. *南方农业学报*, 2017, 48(9): 1691-1699.  
LIU Chao, HAN Lihong, SONG Peibing, WANG Deqin, WANG Haibo, TANG Lizhou. Genome-wide identification and bioinformatics analysis for mulberry WRKY transcription factors [J]. *Journal of Southern Agriculture*, 2017, 48(9): 1691-1699.
- [24] 赵洋, 杨培迪, 刘振, 成杨, 杨阳. 13种植物 *actin* 基因的密码子使用特性分析[J]. *南方农业学报*, 2016, 47(4): 519-523.  
ZHAO Yang, YANG Peidi, LIU Zhen, CHENG Yang, YANG Yang. Characterization of codon usage of actin genes for 13 species of plants [J]. *Journal of Southern Agriculture*, 2016, 47(4): 519-523.
- [25] 赵耀, 刘汉梅, 顾勇, 黄玉碧. 玉米 *waxy* 基因密码子偏好性分析[J]. *玉米科学*, 2008, 16(2): 16-21.  
ZHAO Yao, LIU Hanmei, GU Yong, HUANG Yubi. Analysis of characteristic of codon usage in *waxy* gene of *Zea mays*[J]. *Journal of Maize Sciences*, 2008, 16(2): 16-21.
- [26] TATARINOVA T V, ALEXANDROV N N, BOUCK J B, FELDMANN K A. GC 3 biology in corn, rice, sorghum and other grasses[J]. *BMC Genomics*, 2010, 11(1): 308.