

• 药材与资源 •

香鳞毛蕨 4CL 基因密码子使用特性及表达预测分析

李珊珊^{1,2}, 陈玲玲¹, 王鹤萌¹, 胡宝忠^{1*}, 常 纓^{1*}

1. 东北农业大学生命科学学院, 黑龙江 哈尔滨 150030

2. 齐齐哈尔大学生命科学与农林学院, 黑龙江 齐齐哈尔 161006

摘要: 目的 研究蕨类植物中 4CL 基因的表达与分析, 为蕨类植物次级代谢产物的积累和调控奠定理论基础。方法 运用 CHIPS、CUSP 和 CodonW 在线程序对自主克隆的香鳞毛蕨 4CL 基因 (Df4CL) 序列进行分析, 并与大肠杆菌、酵母菌、拟南芥、烟草等基因进行比较, 对了解 Df4CL 基因密码子使用特性, 为其选择合适的表达系统具有重要意义。结果 Df4CL 基因对 ATGC 选择没有偏向性, 并且与小立碗藓的密码子使用偏好性相一致。在密码子使用频率上, Df4CL 基因与拟南芥的差异小于与烟草、大肠杆菌、酵母的差异; 基于 4CL 基因的密码子使用偏性的系统聚类分析表明, 香鳞毛蕨与小立碗藓、拟南芥聚为一类。结论 预示 Df4CL 基因更适合在拟南芥中外源表达。

关键词: 香鳞毛蕨; 4-香豆酸辅酶 A 连接酶; 密码子使用特性; 基因表达

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2016)10-1753-09

DOI: 10.7501/j.issn.0253-2670.2016.10.021

Analysis on codon using features and expression prediction of 4CL gene in *Dryopteris fragrans*

LI Shan-shan^{1,2}, CHEN Ling-ling¹, WANG He-meng¹, HU Bao-zhong¹, CHANG Ying¹

1. College of Life Sciences, Northeast Agricultural University, Harbin 150030, China

2. College of Life Sciences and Agriculture and Forestry, Qiqihar University, Qiqihar 161006, China

Abstract: Objective Expression and analysis of 4CL gene of ferns, which lays the theoretical foundation for the accumulation and regulation of secondary metabolites of ferns. **Methods** Cloned 4CL gene (Df4CL) sequence from *Dryopteris fragrans* was analyzed by CHIPS, CUSP, and Codon W programs online, and then compared with *Escherichia coli*, yeast, *Arabidopsis thaliana*, and *Nicotiana tabacum* L. It is important to identify the codon usage of Df4CL gene and select appropriate expression systems. **Results** Df4CL gene had not the bias toward the synonymous codons with ATGC at the third codon position, but had the preference consistency with *Physcomitrella patens* codon usage. The differences in codon usage frequency between Df4CL gene and *A. thaliana* were less than those in *N. tabacum*, *E. coli*, and yeast. Based on 4CL gene codon usage bias of cluster analysis, it was shown that *D. fragrans*, *P. patens*, and *A. thaliana* were classed into one group. **Conclusion** The *A. thaliana* expression system may be more suitable for the exogenous expression of Df4CL gene of *D. fragrans*.

Key words: *Dryopteris fragrans* (L.) Schott; 4-coumarate: CoA ligase (4CL); codon using features; gene expression

香鳞毛蕨 *Dryopteris fragrans* (L.) Schott 是鳞毛蕨科 (Dryopteridaceae) 鳞毛蕨属 *Dryopteris* Adans 多年生蕨类植物, 其作为民间用药主要用于治疗关节炎及各种皮肤病, 目前国内外研究学者发现, 香鳞毛蕨具有抗肿瘤、抗菌、抗氧化、驱虫等作用^[1-7], 由于其独特的药用价值而引起人们的广泛关注。近年来对香鳞毛蕨的研究主要集中在对其化学成分、功能及多样性的研究^[8-10], 而鲜少有关于次级代谢产物相关基因的

报道。4-香豆酸辅酶 A 连接酶 (4CL, EC 6.2.1.12) 是苯丙氨酸代谢途径的关键酶, 其途径产生的类黄酮、木质素等次生物质在植物抗逆境、抗病害上有非常重要的作用^[11-13]。因此, 深入研究 4CL 基因对次级代谢产物的积累具有至关重要的意义。

为深入研究 4CL 基因的功能及蛋白质的特性常常需要基因的外源表达, 而基因的外源表达受到很多因素的影响, 其中密码子偏好性是影响外源基

收稿日期: 2015-09-06

基金项目: 国家自然科学基金资助项目 (31070291); 齐齐哈尔市科学技术项目 (NYGG-201206-4)

作者简介: 李珊珊 (1983—), 女, 讲师, 博士研究生, 主要从事植物分子生物学研究。E-mail: lishanshan83@163.com

*通信作者 胡宝忠 (1962—), 男, 教授, 博士, 主要从事植物分子生物学。E-mail: bzhu@neau.edu.cn

常 纓 (1970—), 女, 教授, 博士, 主要从事植物分子生物学。E-mail: changying@neau.edu.cn

因表达的重要因素之一。已知物种对密码子的使用不是随机的，往往倾向于使用一种或几种特定的同义密码子，即存在密码子使用偏好性 (codon usage bias) [14]。密码子偏性的分析有助于揭示有关物种间或某一物种的基因家族间的基因进化规律；了解转录和翻译进程中的调控机制；在预测外源基因的最适宿主和通过改良外源基因以提高其表达水平等方面具有重要意义 [15]。

本研究利用自主克隆的 Df4CL 基因序列，通过 CHIPS、CUSP 和 CodonW 在线程序在线程序分析 Df4CL 基因密码子使用偏好性，并且与大肠杆菌、酵母菌、拟南芥、烟草等基因组密码子偏好性进行比较，并进行聚类分析，以期为该基因实现异源表达选择合适的宿主表达系统，或者通过优化密码子

提高基因的表达式，为深入了解该基因及表达调控机制提供理论基础，为进一步研究 4CL 蛋白质的结构和功能奠定基础。

1 材料和方法

1.1 序列数据

Df4CL 基因为前期试验所得，并提交到美国国立生物技术信息中心 NCBI，基因序列号分别为 Df4CL1 (KF801576)、Df4CL2 (KF836752)、Df4CL3 (KF836753) 和 Df4CL4 (KF836754)。其他序列通过 NCBI 收集获得 (表 1)。小立碗藓 *Physcomitrella pAtens* Bruch、火炬松 *Pinus teada* L.、拟南芥 *Arabidopsis thaliana* (L.) Heynh.、烟草 *Nicotiana tabacum* Linn.、水稻 *Oryza sativa* L.、黑麦草 *Lolium perenne* L. 的序列通过 NCBI 收集获得 (表 1)。

表 1 4CL 基因的完整 cDNA 和编码氨基酸序列

Table 1 Complete cDNA and coded amino acid sequences of 4CL gene

植物	物种	基因编号	数据库 cDNA 编号	CDS 长度/bp	数据库蛋白编号
双子叶植物	拟南芥	At4CL1	At1g51680	1 686	NP175579
		At4CL2	At3g21240	1 671	NP188761
		At4CL3	At1g65060	1 686	NP176686
		At4CL4	At3g21230	1 713	NP188760
	烟草	Nt4CL1	D43773	1 629	BAA07828
		Nt4CL2	U50845	1 644	AAB18637
		Nt4CL3	U50846	1 629	AAB18638
单子叶植物	水稻	Os4CL1	NM001052604	1 665	NP001046069
		Os4CL2	NM001054354	1 710	NP001047819
		Os4CL4	NM001064787	1 680	NP001058252
	黑麦草	Lp4CL1	AF052221	1 713	AAF37732
		Lp4CL2	AF052222	1 671	AAF37733
		Lp4CL3	AF052223	1 674	AAF37734
裸子植物	火炬松	Pt4CL1	U12012	1 614	AAA92668
		Pt4CL2	U12013	1 614	AAA92669
苔藓植物	小立碗藓	Pp4CL1	EU167552	1 758	ABV60447
		Pp4CL2	EU167553	1 758	ABV60448
		Pp4CL3	EU167554	1 731	ABV60449
		Pp4CL4	EU167555	1 713	ABV60450
石松类植物	江南卷柏	Sem4CL1Sem4CL2	XM_002979027XM_002986499	1 614	XP_002969881
				1 635	XP_002979073

1.2 方法

1.2.1 4CL 基因家族密码子使用性参数分析 利用 CodonW 软件分析 Df4CL 基因及拟南芥、烟草、水稻、黑麦草、火炬松、小立碗藓、江南卷柏等植物 4CL 基因家族密码子使用特性，参数包括：A3s、G3s、C3s、T3s (同义密码子在第 3 位上相应碱基

的出现频率)、密码子适应指数 (CAI)、密码子偏爱指数 (CBI)、最优密码子使用频率 (FOP)、密码子有效数 (NC)、密码子第 3 位的 (G+C) 量 (GC3s)、基因的 (G+C) 量 (GC)、同义氨基酸数 (L-xym)、氨基酸数 (L-aa)、平均亲水性值 (Gravy) 和芳香性值 (Aroma) 等。

1.2.2 4CL 基因家族同义密码子相对使用度 (RSCU) 分析 RSCU 是某一特定的密码子在编码对应氨基酸的同义密码子中的相对概率。本研究应用 MEGA 5.0 软件计算各基因 CDS 序列的 RSCU 值。

1.2.3 Df4CL 基因家族表达预测 密码子使用频率分别采用 EMBOSS 在线程序 (<http://vm-bioinfo.toulouse.inra.fr/emboss>) 中的 CHIPS (codon heterozygosity in a protein coding sequence) 和 CUSP (create a codon usage table) 进行分析。本实验所采用的大肠埃希菌、酵母、拟南芥、烟草基因组的密码子使用频率数据来源于 Codon Usage Database (<http://www.kazusa.or.jp/codon>)。

1.2.4 Df4CL 基因家族聚类分析 对植物中基因家族密码子使用偏好性的聚类采用 SPSS 19 系统多元分析软件。在此过程中, 将每一条基因作为一个对象, 密码子的 RSCU 值作为变量。去除值始终为 1 的编码 Trp 和 Met 的密码子 UGG 和 AUG, 再除去 3 个不编码氨基酸的终止密码子, 取余下的 59 个密码子的值对基因的密码子使用偏好性进行聚类分析。

2 结果与分析

2.1 4CL 基因家族密码子使用性参数分析

利用 CodonW 软件分析江南卷柏、小立碗藓、火炬松、拟南芥、烟草、大豆、水稻、黑麦草等植物中的 4CL 基因家族密码子使用情况 (表 2)。在

表 2 不同植物中 4CL 基因家族密码子使用特性

Table 2 Codon using feature of 4CL gene in various plants

基因	T3s	C3s	A3s	G3s	CAI	CBI	FOP	NC	GC3s	GC	L-sym	L-aa	Gravy	Aroma
Df4CL1	0.355 0	0.280 7	0.262 8	0.360 2	0.223	-0.072	0.368	54.15	0.494	0.487	516	528	0.198 106	0.075 758
Df4CL2	0.325 2	0.331 3	0.242 6	0.344 9	0.233	-0.021	0.400	56.01	0.532	0.506	587	605	0.010 744	0.061 157
Df4CL3	0.340 9	0.398 4	0.245 6	0.226 0	0.249	0.105	0.470	56.93	0.509	0.495	566	585	0.029 060	0.066 667
Df4CL4	0.353 8	0.314 9	0.280 5	0.258 1	0.214	0.002	0.411	57.59	0.466	0.491	569	589	0.135 654	0.076 401
Sm4CL1	0.107 9	0.570 5	0.074 4	0.470 6	0.270	0.181	0.513	37.59	0.845	0.629	524	537	0.181 750	0.057 728
Sm4CL2	0.152 8	0.502 2	0.171 7	0.401 5	0.256	0.142	0.486	45.20	0.730	0.593	533	544	0.094 669	0.068 015
Pp4CL1	0.322 2	0.329 8	0.289 9	0.306 2	0.201	0.018	0.420	59.35	0.501	0.495	788	813	-0.150 923	0.065 191
Pp4CL2	0.262 7	0.394 9	0.259 7	0.322 6	0.211	0.031	0.424	57.22	0.572	0.522	705	726	-0.048 623	0.064 738
Pp4CL3	0.271 0	0.363 6	0.280 1	0.316 0	0.194	0.016	0.420	59.11	0.545	0.510	672	698	-0.252 006	0.080 229
Pt4CL1	0.255 8	0.428 6	0.214 1	0.364 8	0.257	0.094	0.467	49.55	0.617	0.530	527	537	0.099 814	0.068 901
Pt4CL2	0.257 5	0.423 0	0.207 1	0.375 3	0.258	0.100	0.471	49.60	0.620	0.532	527	537	0.105 587	0.068 901
At4CL1	0.313 6	0.366 2	0.286 7	0.293 8	0.262	0.110	0.477	56.69	0.514	0.481	547	561	0.018 360	0.065 954
At4CL2	0.289 2	0.384 1	0.281 7	0.302 1	0.219	0.035	0.431	55.41	0.536	0.484	541	556	-0.012 770	0.064 748
At4CL3	0.351 8	0.341 2	0.341 7	0.192 0	0.228	0.075	0.451	52.67	0.429	0.460	552	561	-0.017 290	0.065 954
At4CL4	0.329 0	0.324 7	0.290 3	0.313 1	0.250	0.068	0.455	58.24	0.496	0.477	552	570	-0.032 807	0.068 421
Nt4CL1	0.445 5	0.215 9	0.343 0	0.258 8	0.197	-0.090	0.359	50.61	0.361	0.423	529	542	0.029 520	0.075 646
Nt4CL2	0.443 2	0.204 5	0.323 0	0.291 0	0.222	-0.071	0.371	52.31	0.377	0.429	531	547	0.037 294	0.073 126
Nt4CL3	0.445 5	0.227 3	0.337 4	0.253 4	0.208	-0.060	0.377	51.85	0.366	0.423	528	542	0.022 694	0.075 646
Gm4CL1	0.290 1	0.389 0	0.274 6	0.280 5	0.238	0.061	0.448	57.44	0.534	0.500	534	546	0.077 473	0.069 597
Gm4CL2	0.320 8	0.429 2	0.239 0	0.264 2	0.226	0.040	0.436	51.45	0.546	0.497	535	547	-0.120 110	0.069 470
Gm4CL3	0.249 5	0.398 7	0.252 8	0.335 7	0.238	0.081	0.457	56.91	0.588	0.523	558	570	0.068 772	0.068 421
Gm4CL4	0.245 7	0.428 9	0.237 4	0.323 5	0.238	0.083	0.457	55.41	0.603	0.526	549	562	0.097 153	0.069 395
Os4CL1	0.080 8	0.615 7	0.057 6	0.485 0	0.254	0.146	0.500	35.69	0.885	0.646	538	554	0.061 011	0.064 982
Os4CL2	0.329 0	0.324 7	0.290 3	0.313 1	0.250	0.068	0.455	58.24	0.496	0.477	552	570	-0.032 807	0.068 421
Os4CL4	0.118 3	0.548 4	0.090 7	0.477 8	0.255	0.172	0.513	40.13	0.825	0.630	544	559	0.075 313	0.064 401
Lp4CL1	0.033 2	0.580 9	0.037 1	0.565 4	0.275	0.214	0.535	32.55	0.941	0.692	555	570	0.201 754	0.057 895
Lp4CL2	0.110 6	0.597 9	0.076 6	0.429 6	0.294	0.237	0.550	37.91	0.842	0.644	544	556	0.167 446	0.061 151
Lp4CL3	0.038 7	0.651 6	0.047 7	0.496 3	0.281	0.215	0.537	32.64	0.928	0.662	544	557	0.085 637	0.059 246

这些参数中, CAI、CBI 和 NC 常用于预测基因表达水平^[15]。一般 CAI 和 CBI 呈现正相关性, 值在 0~1, 越接近 1 表示密码子偏好性越强, 基因表达水平越高^[16-17], 而与 NC 则呈现负相关, NC 数值一般在 20~61, 越靠近 20 表示同义密码子偏性越大, 越靠近 61 表示同义密码子偏性越小^[18]。高表达基因其密码子偏爱程度大, CAI 和 CBI 值偏大而 NC 值较小; 低表达基因则含有较多种类的稀有密码子, 故偏爱程度低, CAI、CBI 值较小而 NC 值较大, 其值会受到基因氨基酸组成和基因长短的影响^[19-20]。从数据上看, 黑麦草 Lp4CL1、Lp4CL2 和 Lp4CL3, 水稻 Os4CL1, 小立碗蕨 Sm4CL1 的 NC 值在 35 左右, 表达水平较高, 而小立碗蕨 Pp4CL1-4, 拟南芥 At4CL2、At4CL4, 大豆 Gm4CL1、Gm4CL3-4, 水稻 Os4CL2 的 NC 值在 55 以上, 表达水平较低, 剩下的 4CL 基因表达水平一般。一般认为 GC 量和 GC3s 这 2 个因素对基因的密码子选择有重要影响。在这 9 种植

物 28 条基因序列中 13 条 4CL 基因 GC 量小于 0.5, 预示 4CL 基因对 GC 没有明显的偏好性。GC3s 值大于 0.5 的占 28%, 说明大部分 4CL 基因偏好以 A/T 结尾的密码子。

2.2 4CL 基因家族 RSCU 分析

RSCU 是一个氨基酸所对应的密码子实际运用个数与理论运用个数之间的比值, 能直观地反映出密码子使用的偏好性程度^[21], 与基因长度和氨基酸丰度无关。当实际运用个数与理论运用个数相同时, RSCU=1, 此时密码子不存在偏好性; 当 RSCU>1 时, 则说明该密码子出现的频率比其他同义密码子高, 为基因偏好使用的密码子; 当 RSCU<1 时, 则密码子出现的相对频率较低, 是基因较少使用的密码子。通过比较发现 (表 3), 在烟草中, RSCU>1 的密码子有 29 个, 水稻中有 28 个, 黑麦草中有 25 个, 拟南芥中有 31 个, 火炬松中有 28 个, 江南卷柏中有 28 个, 这些密码子即为这些植物的最常使用的密码子, 即最优密码子。

表 3 香鳞毛蕨与其他植物中 4CL 基因家族 RSCU

Table 3 Relatively synonymous codon usage (RSCU) of 4CL gene in *D. fragrans* and other plants

氨基酸	密码子	个数								RSCU							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Phe	UUU	8.8	18.7	1.0	1.3	10.3	14.0	8.5	0.0	0.93	1.27	0.36	0.38	1.11	1.02	0.69	0.00
	UUC	10.0	10.7	7.0	5.7	8.3	13.5	16.0	6.0	1.07	0.73	1.83	1.62	0.89	0.98	1.31	2.00
Leu	UUA	3.3	7.3	0.3	0.0	4.8	7.5	2.0	0.0	0.35	8.30	0.06	0.00	0.70	0.75	0.36	0.00
	UUG	13.8	11.3	0.3	1.0	5.8	12.8	8.0	6.5	1.47	1.66	0.06	0.16	0.85	1.27	1.43	1.10
	CUU	11.8	9.0	8.0	7.7	9.3	9.8	7.0	3.0	1.25	1.32	1.47	1.20	1.36	0.97	11.50	0.51
	CUC	8.3	5.3	10.3	14.7	12.8	13.3	4.5	14.0	0.88	0.78	1.90	2.30	1.88	1.32	9.00	2.37
	CUA	8.0	1.3	4.3	7.3	5.8	5.8	4.0	5.0	0.85	0.20	0.80	1.15	0.85	0.57	0.72	0.85
	CUG	11.3	6.7	9.3	7.7	2.5	11.0	8.0	7.0	1.20	0.98	1.71	1.20	0.37	1.10	1.43	1.18
Ile	AUU	10.8	17.3	4.0	3.3	8.0	9.8	9.5	8.0	1.13	1.64	0.80	0.75	1.04	1.13	1.30	1.26
	AUC	10.8	5.7	2.7	1.7	11.5	9.3	9.5	10.0	1.13	0.54	1.00	1.00	1.5	1.07	1.30	1.58
	AUA	7.0	8.7	1.0	0.7	3.5	7.0	30.0	10.0	0.74	0.82	0.20	0.15	0.46	0.81	0.41	0.16
Met	AUG	8.0	12.3	2.7	1.7	5.0	12.8	8.0	3.0	1.00	1.00	1.00	1.00	1.00	1.00	2.00	1.00
Val	GUU	7.5	13.3	11.7	8.3	14.8	12.0	3.0	3.0	1.26	1.44	1.22	1.01	1.92	1.31	0.75	0.67
	GUC	5.3	5.3	15.0	15.3	6.5	9.8	80.0	5.0	0.88	0.58	1.57	1.86	0.85	1.06	2.00	1.11
	GUA	3.5	5.7	4.0	1.7	2.8	5.8	3.0	0.0	0.59	0.61	0.42	0.2	0.36	0.63	0.75	0.00
	GUG	7.5	12.7	7.7	7.7	6.8	9.3	20.0	10.0	1.26	1.37	0.80	0.93	0.88	1.01	0.50	2.22
Ser	UCU	12.3	11.0	1.0	1.3	12.3	16.5	11.0	7.0	1.05	1.68	0.36	0.49	1.41	1.16	1.18	0.88
	UCC	12.8	7.3	6.3	5.0	8.5	14.3	8.0	12.0	1.09	1.12	2.28	1.84	0.98	1.00	0.86	1.52
	UCA	10.8	8.3	1.3	1.3	9.5	13.8	10.0	5.0	0.92	1.27	0.48	0.49	1.10	0.96	1.07	0.63
	UCG	10.0	3.7	5.0	5.3	4.0	17.5	10.0	10.5	0.85	0.56	1.80	1.96	0.46	1.23	1.07	1.33

续表 3

氨基酸	密码子	个数								RSCU							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Pro	CCU	9.8	9.0	14.3	11.7	9.3	13.3	11.5	13.0	1.12	1.23	1.15	0.97	1.18	1.09	1.03	0.95
	CCC	7.8	7.7	12.7	11.0	6.5	8.8	9.0	13.0	0.89	1.05	1.02	0.92	0.83	0.72	0.81	0.95
	CCA	10.8	7.7	9.3	9.3	9.0	16.3	13.0	21.5	1.24	1.05	0.75	0.78	1.31	1.34	1.71	1.58
	CCG	6.5	5.0	13.3	16.0	6.5	10.3	11.0	7.0	0.75	0.68	1.07	1.33	0.83	0.85	0.99	0.51
Thr	ACU	4.3	8.0	1.7	2.0	9.0	11.3	3.0	4.0	0.83	1.23	0.54	0.75	1.31	0.98	0.67	0.94
	ACC	5.0	2.7	3.7	5.3	7.0	10.3	2.0	1.0	0.98	0.41	1.19	2.00	1.02	0.90	0.44	0.24
	ACA	8.8	10.7	2.7	1.0	8.3	17.5	9.0	7.0	1.71	1.64	0.86	0.38	1.20	1.53	2.00	1.65
	ACG	2.5	4.7	4.3	2.3	3.3	6.8	4.0	5.0	0.49	0.72	1.41	0.88	0.47	0.59	0.89	1.18
Ala	GCU	9.0	14.3	20.3	26.0	11.0	15.3	12.0	19.0	0.82	1.91	1.44	1.54	1.60	1.21	1.37	1.23
	GCC	12.5	4.7	14.3	2.0	1.14	15.8	6.0	13.0	1.70	0.62	1.02	1.28	1.02	1.25	0.69	0.84
	GCA	16.8	9.0	9.3	10.0	7.8	13.0	11.0	16.0	1.53	1.20	0.66	0.59	1.13	1.03	1.26	1.03
	GCG	5.5	2.0	12.3	10.0	0.5	6.3	6.0	14.0	1.80	0.27	0.88	0.59	0.25	0.50	0.69	0.90
Tyr	UAU	8.3	8.7	1.7	0.7	6.3	5.3	4.0	1.0	1.14	0.98	0.91	0.44	1.39	0.93	1.00	0.67
	UAC	6.3	9.0	2.0	2.3	2.8	6.0	4.0	2.0	0.86	1.02	1.09	1.56	0.61	1.07	1.00	1.33
His	CAU	14.3	6.3	15.0	16.3	13.0	18.3	0.5	21.0	1.36	1.15	1.06	1.07	1.06	1.23	0.13	1.33
	CAC	6.8	4.7	13.3	14.3	11.5	11.5	2.0	10.5	0.64	0.85	0.94	0.93	0.94	0.77	0.50	0.67
Gln	CAA	11.5	9.7	18.0	17.3	16.3	19.8	9.0	13.5	1.24	1.18	1.24	1.37	1.59	1.20	1.29	1.06
	CAG	7.0	6.7	11.0	8.0	4.3	13.3	5.0	12.0	0.50	0.82	0.76	0.63	0.41	0.80	0.71	0.94
Asn	AAU	8.8	12.3	1.7	0.7	9.8	14.3	13.0	1.0	1.15	1.32	0.71	0.33	1.07	1.18	1.18	1.00
	AAC	6.5	6.3	3.0	3.3	8.5	10.0	9.0	1.0	0.85	0.68	1.29	1.67	0.93	0.82	0.82	1.00
Lys	AAA	9.5	21.7	2.7	2.7	15.5	10.3	7.0	1.0	0.92	1.24	0.52	0.43	1.41	0.78	1.56	0.29
	AAG	11.3	13.3	7.7	9.7	6.5	16.0	2.0	6.0	1.08	0.76	1.48	1.57	0.59	1.22	0.44	1.71
Asp	GAU	11.0	18.7	12.0	14.3	23.0	14.5	9.0	8.0	1.07	1.42	0.75	0.88	1.50	1.00	0.90	0.78
	GAC	9.5	7.7	20.0	18.3	7.8	14.5	11.0	12.5	0.93	0.58	1.25	1.12	0.50	1.00	1.10	1.22
Glu	GAA	10.8	13.3	12.0	11.3	17.3	18.0	18.5	7.0	1.01	1.25	0.75	0.96	1.16	1.01	1.57	1.17
	GAG	10.5	8.0	14.7	12.3	12.5	17.5	5.0	5.0	0.99	0.75	1.10	1.04	0.84	0.99	0.43	0.83
Cys	UGU	9.5	8.7	2.7	2.0	7.0	15.5	7.0	6.0	0.86	0.76	0.52	0.46	0.90	1.02	0.78	0.86
	UGC	12.5	14.0	7.7	6.7	8.5	15.0	11.0	8.0	1.14	1.24	1.48	1.54	1.10	0.98	1.22	1.14
Arg	CGU	3.5	3.0	20.0	20.3	15.0	9.5	16.5	16.0	0.58	0.53	1.40	1.31	1.42	0.82	1.14	1.18
	CGC	3.8	2.3	28.3	33.0	9.0	6.5	6.0	16.5	0.63	0.42	1.98	2.31	0.85	0.56	0.41	1.21
	CGA	4.3	6.7	17.7	18.3	13.5	16.5	18.0	12.0	0.71	1.19	1.24	1.18	1.28	1.43	0.88	0.88
	CGG	5.5	2.0	14.0	14.7	9.8	9.3	15.5	19.0	0.92	0.36	0.98	0.95	0.92	0.80	1.07	1.40
Ser	AGU	9.5	4.3	0.3	1.0	9.0	13.0	4.0	3.0	0.81	0.66	0.12	0.37	1.04	0.91	0.43	0.38
	AGC	15.0	4.7	2.7	2.3	8.8	10.5	13.0	10.0	1.28	0.71	0.96	0.86	1.01	0.74	1.39	1.26
Arg	AGA	9.5	12.0	1.3	1.7	11.0	13.8	13.0	8.0	1.58	2.14	0.12	0.09	1.04	1.19	0.90	0.59
	AGG	9.5	7.7	4.0	5.3	5.0	13.8	18.0	10.0	1.58	1.37	0.28	0.34	0.47	1.19	1.24	0.74
Gly	GGU	9.8	7.0	18.3	17.3	12.3	12.3	6.0	8.0	1.20	0.98	0.92	0.91	1.46	1.24	0.91	0.65
	GGC	7.5	5.0	26.3	27.7	6.8	7.5	9.0	17.5	0.92	0.70	1.33	1.46	0.81	0.76	1.36	1.41
	GGA	8.3	10.7	20.0	18.0	10.5	12.5	5.0	11.0	1.02	1.49	1.01	0.95	1.25	1.27	0.75	0.89
	GGG	7.0	6.0	13.0	14.7	4.0	7.3	6.5	13.0	0.86	0.84	0.74	0.68	0.48	0.73	0.98	1.05

1-香磷毛厥 2-烟草 3-水稻 4-黑麦草 5-拟南芥 6-小立碗藓 7-火炬松 8-江南卷柏
 1-*D. fragrans* 2-*N. tabacum* 3-*O. sativa* 4-*L. perenne* 5-*A. thaliana* 6-*P. patens* 7-*P. taeda* 8-*S. moellendorffii*

2.3 Df4CL 基因家族密码子使用分析

通过表 1 可以看出, 4CL 基因在不同物种中一般以基因家族的形式出现, 而本研究中前期获得的 Df4CL 基因, 通过序列比对也确定为一个家族。利用 CodonW 软件分析 Df4CL 基因家族密码子使用情况 (表 2), 发现香鳞毛蕨 Df4CL2-4 的 NC 值在 55 以上, 表达水平较低, 而 DF4CL1 的 CAI 值最大, NC 值相对较小, 推测在无外界胁迫下, Df4CL 基因家族内部更偏向于使用 DF4CL1。Df4CL 家族内部的 GC 量平均为 0.495, GC3s 平均量为 0.500, 说明香鳞毛蕨对 ATGC 选择没有偏向性。总体来看, 香鳞毛蕨与小立碗藓的密码子使用情况基本一致。

通过对 4CL 基因家族 RSCU 分析 (表 3), 在香鳞毛蕨中, 最优密码子 RSCU>1 的密码子有 29 个, 其中 22 个与拟南芥最优密码子使用情况相一致, 20 个与烟草最优密码子使用情况相一致, 密码子相对使用度越一致越容易实现基因的外源表达。而其余密码子略有不同, 例如 Phe 在香鳞毛蕨中密码子应用 UUC, 而拟南芥和烟草密码子使用 UUU 等。

2.4 Df4CL 基因家族表达预测

鉴定基因的功能、蛋白特性及转基因经常需要

基因的异源表达, 而这个过程常常受到各种因素的限制, 其中基因密码子与宿主密码子使用偏性的差异是影响表达的重要因素之一。如果外源基因含有大量与宿主表达系统差异较大的罕见密码子, 特别是当罕见密码子成连续分布的时候, 就会造成表达量极低或翻译提前终止。所以 4CL 基因在异源宿主内表达时必须要考虑密码子偏好性。基因密码子与宿主密码子使用频率的比值在 0.5~2.0, 表示二者的密码子偏好性较为接近, 比值 ≤0.5 或比值 ≥2.0, 表示偏好性差异较大。本研究将 Df4CL 基因家族与大肠杆菌、酵母、拟南芥、烟草密码子使用频率进行比较。从表 4 中不难发现, 在 Df4CL1 基因与大肠杆菌、酵母、拟南芥和烟草的使用频率比值中, 分别有 20、21、16 和 17 个; 分析 Df4CL2 基因与大肠杆菌、酵母、拟南芥和烟草的使用频率比值中, 分别有 19、22、16 和 20 个; 在 Df4CL3 中与大肠杆菌、酵母、拟南芥和烟草的使用频率比值中, 分别有 18、18、14 和 20 个, 在 Df4CL4 中与大肠杆菌、酵母、拟南芥和烟草的使用频率比值中, 分别有 26、18、14 和 12 个, 说明 Df4CL1、Df4CL2、Df4CL3 基因更适合在拟南芥中外源表达; 而 Df4CL4 基因更适合在拟南芥和烟草中表达。

表 4 香鳞毛蕨 4CL 基因与其他植物基因组的密码子出现频率 (f)

Table 4 Codon frequency and ratios of Df4CL gene in *D. fragrans* and various plants genome

密码子	氨基酸	香鳞毛蕨 ($f \times 10^{-3}$)				大肠杆菌 ($f \times 10^{-3}$)				酵母 ($f \times 10^{-3}$)				拟南芥 ($f \times 10^{-3}$)				烟草 ($f \times 10^{-3}$)			
		4CL1	4CL2	4CL3	4CL4	4CL1	4CL2	4CL3	4CL4	4CL1	4CL2	4CL3	4CL4	4CL1	4CL2	4CL3	4CL4	4CL1	4CL2	4CL3	4CL4
GCA	A	38.889	21.452	23.891	23.729	1.888	1.041	1.160	1.152	<u>2.415</u>	1.332	1.484	1.474	<u>2.222</u>	1.226	1.365	1.356	1.684	0.929	1.034	1.027
GCC	A	20.370	23.102	13.652	30.508	0.799	0.906	0.535	1.196	1.630	1.848	1.092	<u>2.441</u>	1.978	<u>2.243</u>	1.325	<u>2.962</u>	1.630	1.848	1.092	<u>2.441</u>
GCG	A	14.815	14.851	10.239	8.475	<u>0.467</u>	<u>0.468</u>	<u>0.323</u>	<u>0.267</u>	<u>2.429</u>	<u>2.435</u>	1.679	1.389	1.646	1.650	1.138	0.942	2.554	2.561	1.765	1.461
GCT	A	27.778	24.752	22.184	30.508	1.781	1.587	1.422	1.956	1.316	1.173	1.051	1.446	0.982	0.875	0.784	1.078	0.890	0.793	0.711	0.978
TGC	C	9.259	11.551	5.119	6.780	1.342	1.674	0.742	0.983	1.970	<u>2.458</u>	1.089	1.443	1.286	1.604	0.711	0.942	1.286	1.604	0.711	0.942
TGT	C	11.111	4.950	1.706	3.390	<u>2.020</u>	0.900	<u>0.310</u>	0.616	1.389	0.619	<u>0.213</u>	<u>0.424</u>	1.058	<u>0.471</u>	<u>0.162</u>	<u>0.323</u>	1.134	0.505	<u>0.174</u>	<u>0.346</u>
GAC	D	11.111	19.802	23.891	20.339	0.597	1.065	1.284	1.093	0.550	0.980	1.183	1.007	0.646	1.151	1.389	1.183	0.657	1.172	1.414	1.203
GAT	D	37.037	31.353	22.184	20.339	1.154	0.977	0.691	0.634	0.980	0.829	0.587	0.538	1.012	0.857	0.606	0.556	1.004	0.850	0.601	0.551
GAA	E	20.370	21.452	22.184	23.729	0.533	0.562	0.581	0.621	<u>0.420</u>	<u>0.442</u>	<u>0.457</u>	<u>0.489</u>	0.594	0.625	0.647	0.692	0.566	0.596	0.616	0.659
GAG	E	55.556	44.554	18.771	27.119	<u>3.139</u>	<u>2.517</u>	1.061	1.532	<u>2.909</u>	<u>2.333</u>	0.983	1.420	1.725	1.384	0.583	0.842	1.890	1.515	0.638	0.922
TTC	F	16.667	23.102	27.304	13.559	0.986	1.367	1.616	0.802	0.916	1.269	1.500	0.745	0.805	1.116	1.319	0.655	0.926	1.283	1.517	0.753
TTT	F	31.481	11.551	15.358	30.508	1.357	<u>0.498</u>	0.662	1.315	1.206	<u>0.443</u>	0.588	1.169	1.444	0.530	0.704	1.399	1.254	<u>0.460</u>	0.612	1.215
GGA	G	9.259	11.551	17.065	15.254	1.029	1.283	1.896	1.695	0.849	1.060	1.566	1.399	<u>0.383</u>	<u>0.477</u>	0.705	0.630	<u>0.399</u>	<u>0.498</u>	0.736	0.658
GGC	G	12.963	23.102	22.184	30.508	<u>0.465</u>	0.828	0.795	1.093	1.336	<u>2.382</u>	<u>2.287</u>	<u>3.145</u>	1.409	<u>2.511</u>	<u>2.411</u>	<u>3.316</u>	1.157	<u>2.063</u>	1.981	<u>2.724</u>
GGG	G	22.222	13.201	6.826	15.254	1.967	1.168	0.604	1.350	<u>3.704</u>	<u>2.200</u>	1.138	<u>2.542</u>	<u>2.179</u>	1.294	0.669	1.495	<u>2.116</u>	1.257	0.650	1.453
GGT	G	12.963	18.152	15.358	27.119	0.531	0.744	0.629	1.111	0.540	0.756	0.640	1.130	0.584	0.818	0.692	1.222	0.581	0.814	0.689	1.216
CAC	H	14.815	14.851	13.652	10.169	1.512	1.515	1.393	1.038	1.924	1.929	1.773	1.321	1.703	1.707	1.569	1.169	1.703	1.707	1.569	1.169
CAT	H	9.259	23.102	25.597	20.339	0.681	1.699	1.882	1.496	0.676	1.686	1.868	1.485	0.671	1.674	1.855	1.474	0.691	1.724	1.91	1.518
ATA	I	25.926	6.601	3.413	18.644	<u>4.801</u>	1.222	0.632	<u>3.453</u>	1.457	<u>0.371</u>	<u>0.192</u>	<u>1.047</u>	<u>2.058</u>	0.524	<u>0.271</u>	1.480	1.852	<u>0.472</u>	<u>0.244</u>	1.332

2.5 香鳞毛蕨 4CL 基因家族聚类分析

为进一步了解所列物种 4CL 基因密码子使用差异的大小,根据表 3 中基于密码子使用频率的聚类树状图(图 1)分析,可以看出,火炬松密码子使用频率与其他植物相比具有明显的差异性,而香鳞毛蕨与小立碗蕨聚为一类,这表明亲缘关系较近的物种,密码子使用偏好性也比较相近,但它们与拟南芥也聚为一类,说明某些物种也存在一定的差异,但也进一步说明,香鳞毛蕨基因在拟南芥植物中表达是可行的。

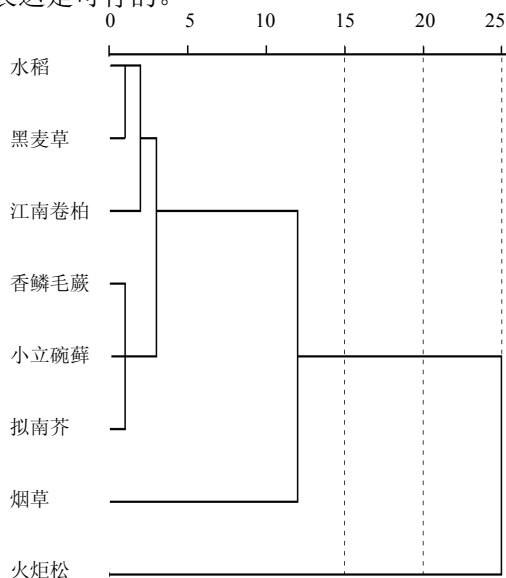


图 1 基于 4CL 基因的 RSCU 的聚类树状图

Fig. 1 Cluster dendrogram of RSCU values based on 4CL gene

3 讨论

苯丙烷代谢途径是植物三大代谢途径之一,可产生植物体中约有五分之一的次级代谢产物,而这些次级代谢产物不但对于植物本身具有不可替代的作用,而对于人们的健康也具有非凡的意义。黄酮类物质就是其中一个,为了更好地发挥药效,以期次级代谢产物更多的积累,而这条代谢途径中的关键基因将影响次级代谢产物的积累和调控。本研究中主要研究苯丙烷代谢途径中的关键调节酶 4CL 的基因,通过密码子使用特性预测其表达体系,为更好地表达蛋白提供重要的理论依据。

通过前期实验获得 Df4CL 基因 4 条,通过序列比对、保守区确定获得的 4 条 4CL 基因因为一个基因家族。将 Df4CL 基因密码子使用频率与代表生物大肠杆菌、酵母和拟南芥、烟草基因组密码子使用频率进行比较,结果显示 Df4CL 基因密码子与大肠杆

菌、酵母、烟草的差异较大,要想在其内表达,还需经过密码子优化,Df4CL 基因更适合在拟南芥中表达。这一结果可以为今后 Df4CL 基因在宿主中实现较高水平表达的研究确立方向,也可为通过优化密码子提高基因的表达量提供依据。

参考文献

- [1] 沈志滨, 马英丽, 江蔚新, 等. 香鳞毛蕨对真菌的抑制作用 [J]. 中草药, 2005, 36(5): 735-736.
- [2] 张彦龙, 曾伟民, 王慧荣. 香鳞毛蕨中木脂素类抗氧化活性成分的研究 [J]. 中草药, 2008, 39(3): 343-346.
- [3] 李红枝, 沈志滨, 赵琪, 等. 香鳞毛蕨软膏对大鼠实验性体癣的治疗作用 [J]. 中药材, 2005, 28(10): 82-83.
- [4] 张彦龙. 香鳞毛蕨治疗类风湿性关节炎有效部位的药理作用和化学成分的研究 [D]. 哈尔滨: 黑龙江中医药大学, 2006.
- [5] 沈志滨, 金哲雄, 张德连, 等. 香鳞毛蕨治疗银屑病的药理作用研究 [J]. 中草药, 2002, 33(5): 67-68.
- [6] 沈志滨, 金哲雄, 张德连. 香鳞毛蕨的生药学研究 [J]. 中草药, 2002, 33(7): 661-663.
- [7] Ito H, Muranaka T, Mori K, et al. Ichthyotoxic phloroglucinol derivatives from *Dryopteris fragrans* and their anti-tumor promoting activity [J]. *Chem Pharm Bull*, 2000, 48: 1190-1195.
- [8] Li X J, Fu Y J, Luo M, et al. Preparative separation of dryofragin and aspidin BB from *Dryopteris fragrans* extracts by macroporous resin column chromatography [J]. *J Pharm Biomed Anal*, 2012, 61: 199-206.
- [9] Li X J, Wang W, Luo M, et al. Solvent-free microwave extraction of essential oil from *Dryopteris fragrans* and evaluation of antioxidant activity [J]. *Food Chem*, 2012, 133: 437-444.
- [10] 黄庆阳, 樊锐锋, 常纓. 香鳞毛蕨种质资源遗传多样性的 AFLP 分析 [J]. 中草药, 2010, 41(6): 971-974.
- [11] Bauer N, Kiseljak D, Jelaska S. The effect of yeast extract and methyl jasmonate on rosmarinic acid accumulation in *Coleus blumei* hairy roots [J]. *Biol Plant*, 2009, 53(4): 650-656.
- [12] Sullivan M L. Phenylalanine ammonia lyase genes in red clover: expression in whole plants and in response to yeast fungal elicitor [J]. *Biol Plant*, 2009, 53(2): 301-306.
- [13] Gutiérrez-Carbajal M G, Monforte-González M, de L Miranda-Ham M, et al. Induction of capsaicinoid synthesis in *Capsicum chinense* cell cultures by salicylic acid or methyl jasmonate [J]. *Biol Plant*, 2010, 54(3): 430-434.
- [14] 刘庆慧, 黄健, 韩文君. WSSV3 编码蛋白的基因密

- 码子偏爱性分析 [J]. 海洋水产研究, 2005, 26(4): 1-7.
- [15] 吴宪明, 吴松峰, 任大明, 等. 密码子偏性的分析方法及相关研究进展 [J]. 遗传, 2007, 29(4): 420-426.
- [16] Sharp P M, Li W H. The codon Adaptation Index: a measure of directional synonymous codon usage bias and its potential applications [J]. *Nucl Acids Res*, 1987, 15(3): 1281-1295.
- [17] Morton B R. Chloroplast DNA codon use: evidence for selection at the psb A locus based on tRNA availability [J]. *J Mol Evol*, 1993, 37(3): 273-280.
- [18] 尚明照, 刘 方, 华金平, 等. 陆地棉叶绿体基因组密码子使用偏性的分析 [J]. 中国农业科学, 2011, 44(2): 245-253.
- [19] Marashi S A, Najafabadi H S. How reliable readjustment is: correspondence regarding A. fugsang, "The 'effective number of codons' revisited" [J]. *Biochem Biophys Res Communicat*, 2004, 324(1): 1-2.
- [20] Wright F. The effective number of codons used in a gene [J]. *Gene*, 1990, 87(1): 23-29.
- [21] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [J]. *J Mol Evol*, 1986, 24(1/2): 28-38.