

DOI: 10.3724/SP.J.1005.2012.01614

茶树 *CBF1* 基因密码子使用特性分析

郭秀丽, 王玉, 杨路成, 丁兆堂

青岛农业大学茶叶研究所, 青岛 266109

摘要: 转录因子 *CBF*(C-repeat-binding factor)广泛存在于各种植物中, 是植物抗逆过程中一个重要的调节因子。

文章运用 CHIPS、CUSP 和 CodonW 在线程序对茶树(*Camellia sinensis*)*CBF1* 基因(*CsCBF1*)序列进行分析, 并与茶树基因、模式植物基因组和其他植物 *CBF* 基因进行比较, 对了解 *CsCBF1* 基因密码子使用特性, 并为其选择合适的表达系统具有重要意义。结果表明: *CsCBF1* 基因与 70 个茶树基因对密码子的使用有明显的差异, *CsCBF1* 基因偏好使用以 G/C 结尾的密码子, 而筛选的 70 个茶树基因偏好使用以 A/T 结尾的密码子。在密码子使用频率上, *CsCBF1* 基因与拟南芥(*Arabidopsis thaliana*)、烟草(*Nicotiana tabacum*)的差异小于与小麦(*Triticum aestivum*)、玉米(*Zea mays*)的差异; 因此, 拟南芥、烟草更适合作为 *CsCBF1* 基因的外源表达宿主。通过分析 40 种植物 *CBF* 基因编码特点可知大部分 *CBF* 基因偏好使用以 G/C 结尾的密码子, 这可能与基因的特殊功能有关。

关键词: 茶树; *CBF1* 基因; 密码子偏好性

Analysis of codon use features of *CBF* gene in *Camellia sinensis*

GUO Xiu-Li, WANG Yu, YANG Lu-Cheng, DING Zhao-Tang

Tea Research Institute, Qingdao Agricultural University, Qingdao 266109, China

Abstract: *CBF* (C-repeat-binding factor) transcription factor exists widely in all kinds of plants. It is an important regulative factor in the process of plant resistance adversity. In this paper, *Camellia sinensis CBF1* gene sequence was analyzed by Codon W, CHIPS, and CUSP programs online, and then compared with *C. sinensis* genes, genomes in other species, and *CBF* genes from 39 plant species. It is important to identify the codon usage of *CsCBF1* gene and select appropriate expression systems. The results showed that *CsCBF1* gene and selected 70 *C. sinensis* genes had distinct usage differences. *CsCBF1* gene was bias toward the synonymous codons with G and C at the third codon position, but 70 *C. sinensis* genes were bias toward the synonymous codons with A and T. The differences in codon usage frequency between *CsCBF1* gene and dicotyledons such as *Arabidopsis thaliana* and *Nicotiana tabacum* were less than monocotyledons such as wheat (*Triticum aestivum*) and corn (*Zea mays*). Therefore, *A. thaliana* and *N. tabacum* expression systems may be more suitable for the expression of *CsCBF1* gene. The analysis results of *CBF* genes from 40 plant species also showed that most of the

收稿日期: 2012-07-02; 修回日期: 2012-07-25

基金项目: 国家科技支撑计划项目(编号: 2011BAD01B01), 山东省科技发展计划(编号: 2011GNC11010)和山东省自然科学基金项目(编号: ZR2010CL021、2009ZRB01654)资助

作者简介: 郭秀丽, 硕士研究生, 专业方向: 茶树分子生物学。E-mail: gxltea@126.com

通讯作者: 丁兆堂, 教授, 博士, 研究方向: 茶树遗传育种与分子生物学研究。E-mail: dztea@163.com

网络出版时间: 2012/9/25 15:17:47

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20120925.1517.001.html>

CBF genes were bias toward the synonymous codons with G and C at the third codon position. The reason of this phenomenon is possible due to special functions of these genes.

Keywords: *Camellia sinensis*; *CBF1* gene; codon bias

核酸是蛋白质合成的模板, 编码天然蛋白质 20 种基本氨基酸的密码子共 61 种, 每种氨基酸可由 1~6 个密码子编码, 这些密码子称为同义密码子。研究表明基因在同义密码子的使用上, 会优先使用其中的一些密码子, 即存在密码子使用偏性^[1]。不同物种之间或同一物种的基因家族间的密码子使用特点可能存在差异, 表明这些基因组以及基因进化过程中承受的压力不同。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^[2]。对于密码子偏性现象, 目前主要有 2 种理论: 中性理论和“选择-突变-漂变”学说^[3]。中性理论认为: 密码子第三位核苷酸的突变是中性选择的结果, 同义密码子的突变并不影响生存适合度, 这种对密码子的选择仅仅和突变有关, 并不受自然选择压力影响; “选择-突变-漂变”学说则认为: 突变的发生是有方向的, 同义密码子的使用偏好性反映了对最优密码子的选择和对非同义密码子的突变-漂变两方面的作用^[4]。研究发现密码子的使用除了受到选择和突变的影响外, 还受其他因素影响^[5], 如碱基组成^[6]、基因表达水平^[7]、tRNA 丰度^[8]、基因长度^[9]、mRNA 二级结构^[10,11]、蛋白质的疏水性、氨基酸保守性^[6]以及密码子-反密码子间结合强度^[12]。

CBF(C-repeat-binding factor)是低温响应基因表达调控网络的重要转录激活因子, 被认为是植物冷驯化过程中基因表达的“主开关”。目前茶树(*Camellia sinensis*)中已经发现的与抗寒调控途径有关的基因很少, 只有 *CBF1*、*ICE1*^[13]、*RAV*^[14]、*ERF*^[15]、*CsCOR*^[16] 5 个。其中对茶树 *CBF1*(*CsCBF1*) 基因的研究较为深入, 已经完成克隆和在低温条件下的表达模式分析。但由于缺乏密码子偏好性的分析, *CsCBF1* 基因还无法在表达水平进行预测, 也没有建立合适的外源表达系统。本研究分析了 *CsCBF1*

基因的密码子偏好性, 并将其与不同物种的密码子偏好性进行比较研究, 对 *CsCBF1* 基因选择合适的表达系统以及提高 *CsCBF1* 基因在外源表达系统中的表达具有重要意义, 并为通过将 *CsCBF1* 基因转入模式植物中进行功能验证奠定基础。

1 材料和方法

1.1 目的基因的获取

本研究所分析的 *CsCBF1* 基因是本实验室自主克隆所得, 其 cDNA 全长 780 bp, 编码 259 个氨基酸, GenBank 序列号为 JX028828。拟南芥(*Arabidopsis thaliana*)、扁桃(*Prunus tenella*)等 39 种植物的 *CBF* 基因序列以及筛选的 70 个茶树基因序列均来源于 GenBank。

1.2 目的基因的筛选

筛选 CDS 的原则如下^[17]: (1) 编码蛋白质的基因, 并且编码序列完整; (2) 编码序列长度大于 300 bp; (3) 在细胞质中翻译, 排除细胞器中翻译的基因; (4) 排除位于质粒、转座子、病毒上的基因; (5) 对多拷贝基因只统计一次。

1.3 有效密码字数(Effective number of codons, ENc)、GC3s 的计算

ENc 是一个基因的密码子使用频率与同义密码子平均使用频率偏差的量化值, 该值的范围在 20 到 61 之间, 越靠近 20 偏好性越强。此值能可以反映同义密码子非均衡使用的偏好程度^[18]。GC3s 则表示密码子的第 3 位碱基中 G+C 出现的频率。本研究运用 CodonW 在线程序(<http://mobyle.pasteur.fr/data/jobs/codonw>)对每条 CDS 序列进行 ENc、GC 含量、GC3s 统计。

1.4 同义密码子相对使用度(Relative synonymous codon usage, RSCU)的计算

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。RSCU 值与基因长度和

氨基酸丰度无关, 它能直观地反映出密码子使用的偏好性程度^[19]。如果某一密码子的RSCU值等于 1, 表明该密码子的使用没有偏好。而当RSCU值大于 1, 则表明该密码子的使用频率较高, 反之亦然^[20]。本研究应用CodonW在线程序计算各基因CDS序列的RSCU值。

1.5 密码子使用频率(Frequency)的计算

Frequency代表该密码子在编码基因总密码子中出现的频率, 即在 1 000 个密码子中出现的次数。密码子使用频率的比值是衡量物种间密码子使用偏好性差异的指标之一。比值在 0.5 ~ 2.0 之间, 表示二者的密码子偏好性较为接近, 反之表示偏好性差异较大^[20]。本研究运用EMBOSS 在线程序中的CUSP 计算*CsCBF1* 基因CDS区的密码子使用频率。拟南芥、烟草等基因组的密码子使用频率数据从Codon Usage Database(<http://www.kazusa.or.jp/codon>)上获得。

1.6 CBF 基因的聚类分析方法

聚类分析是指根据所给样品的特征, 按照模式所在空间的相邻性或相似性对若干样品进行分类的分析方法。我们对基于密码子使用偏性的欧氏平方距离系数进行聚类分析。在聚类过程中, 将单个 mRNA 序列中不同的相对使用概率组成的变量组看成是多维空间中的一个点。不考虑色氨酸 (Trp) 和甲硫氨酸(Met)以及 3 个终止密码子。所以, 这个空间的维数是 59。每一个基因序列可以看成是由 59 个变量组成的空间向量^[21]。计算 2 个基因a和b之间的密码子使用偏性的欧氏距离系数, 公式如下:

$$D_{ab} = \sqrt{\sum_{i=1}^{59} (RSCU_{ai} - RSCU_{bi})^2}$$

我们利用 SPSS18.0 系统多元分析软件来实现对密码子使用偏性的聚类。为更清晰地反映聚类结果, 事先设定聚类类数为 8, 这样聚类后, 各样本序列将按照各自的密码子使用偏性分布到 8 个大类中。而基于基因 CDS 的聚类分析采用 MEGA4 分析软件。

2 结果与分析

2.1 *CsCBF1* 基因密码子偏好性分析

2.1.1 有效密码子数(ENc)和 GC 含量分析

本研究通过在线程序计算 *CsCBF1* 基因的 ENc

值、GC 含量和 GC3s 值分别为 51.37、0.523、0.519, 发现 ENc 值偏大, 说明 *CsCBF1* 基因中各密码子没有明显的偏好性; *CsCBF1* 基因编码区 GC>0.5, 表明 *CsCBF1* 基因在整个编码区序列中 G+C 含量大于 A+T; GC3s>0.5, 表明 *CsCBF1* 基因偏好使用以 G/C 结尾的密码子。

2.1.2 与茶树其他基因密码子偏好性进行比较

CodonW 在线程序计算结果(表 1)表明, 在 *CsCBF1* 基因的密码子中, 有 22 种密码子的RSCU值

1.2, 为*CsCBF1* 基因的偏好性密码子; 其中偏好性较强的有TTG、CAG、AGG、TCA(RUSC>2)。以 G/C 结尾的占多数, 为 12 种。而编码Phe的密码子 TTC、TTT的RSCU值等于 1, 说明在*CsCBF1* 基因中 Phe的使用没有偏好性。从GenBank中筛选了 70 个茶树基因, 共有 28291 个密码子, 其中 15 种密码子的RSCU值 1.2, 其中以A/T结尾的有 12 种, 占多数。说明这 70 个茶树基因总体偏好使用以A/T结尾的密码子, *CsCBF1* 与之相比存在明显的差异。但Lys在*CsCBF1* 基因和所筛选的 70 个茶树基因中都偏爱使用AAG密码子, 而不是AAA, 这与玉米、小麦等单子叶植物和拟南芥、杨树、柑橘等双子叶植物相同, 可见 Lys 的偏爱密码子在植物中具有较高的保守性^[22-24]。

2.2 与其他物种基因组密码子偏好性的比较

转基因研究常常需要进行基因的异源表达, 虽然影响外源基因在宿主中表达的因素较多, 但是选用合适的密码子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^[25]。由于考虑到*CsCBF1* 基因密码子与宿主密码子使用偏性的差异将影响到表达水平, 所以*CsCBF1* 基因在异源宿主内表达时必须要考虑密码子偏好性。因此本研究将*CsCBF1* 基因与拟南芥、烟草、小麦、玉米、大肠杆菌、酵母菌密码子使用频率进行比较(表 2), 其中 C/A、C/N、C/T、C/Z、C/E、C/Y分别表示*CsCBF1* 基因与拟南芥、烟草、小麦、玉米、大肠杆菌、酵母的每种密码子使用频率的比值。在茶树与双子叶植物拟南芥、烟草的密码子使用频率比值中, 2.0 或 0.5 的密码子分别有 22、20 种; 而与单子叶植物小

表 1 CodonW 程序分析 *CsCBFI* 基因和 70 个茶树基因的密码子偏好性

密码子	氨基酸	个数(<i>CBFI</i> 基因)	同义密码子相对 使用度 (<i>CBFI</i> 基因)	个数(70 个茶树基因)	同义密码子相对 使用度(70 个茶树基因)
GCA	A(Ala)	9	1.16	590	1.17
GCC	A	4	0.52	465	0.92
GCG	A	7	0.90	225	0.45
GCT	A	11	<u>1.42</u>	740	<u>1.47</u>
TGC	C(Cys)	3	<u>1.50</u>	230	0.92
TGT	C	1	0.50	268	1.08
GAC	D(Asp)	12	<u>1.26</u>	540	0.74
GAT	D	7	0.74	928	<u>1.26</u>
GAA	E(Glu)	10	0.95	841	0.86
GAG	E	11	1.05	1122	1.14
TTC	F(Phe)	4	1.00	591	0.97
TTT	F	4	1.00	622	1.03
GGA	G(Gly)	4	0.94	571	1.17
GGC	G	2	0.47	366	0.75
GGG	G	4	0.94	404	0.83
GGT	G	7	<u>1.65</u>	605	<u>1.24</u>
CAC	H(His)	5	<u>1.67</u>	332	0.98
CAT	H	1	0.33	343	1.02
ATA	I(Ile)	2	0.86	285	0.58
ATC	I	1	0.43	501	1.02
ATT	I	4	<u>1.71</u>	684	<u>1.40</u>
AAA	K(Lys)	4	0.57	784	0.81
AAG	K	10	<u>1.43</u>	1157	1.19
CTA	L(Leu)	1	0.32	271	0.62
CTC	L	2	0.63	523	1.19
CTG	L	1	0.32	329	0.75
CTT	L	5	<u>1.58</u>	587	<u>1.33</u>
TTA	L	0	0.00	238	0.54
TTG	L	10	<u>3.16</u>	692	<u>1.57</u>
ATG	M(Met)	9	1.00	665	1.00
AAC	N(Asn)	2	0.50	532	0.92
AAT	N	6	<u>1.50</u>	627	1.08
CCA	P(Pro)	7	<u>1.40</u>	529	<u>1.35</u>
CCC	P	3	0.60	265	0.67
CCG	P	6	<u>1.20</u>	229	0.58
CCT	P	4	0.80	548	<u>1.40</u>
CAA	Q(Gln)	0	0.00	497	1.12
CAG	Q	2	<u>2.00</u>	392	0.88
AGA	R(Arg)	2	0.80	327	<u>1.56</u>
AGG	R	8	<u>3.20</u>	362	<u>1.73</u>
CGA	R	1	0.40	150	0.72
CGC	R	1	0.40	108	0.52
CGG	R	3	<u>1.20</u>	144	0.69

续表 1

密码子	氨基酸	个数(<i>CBF1</i> 基因)	同义密码子相对 使用度 (<i>CBF1</i> 基因)	个数(70 个茶树基因)	同义密码子相对 使用度(70 个茶树基因)
CGT	R	0	0.00	163	0.78
AGC	S(Ser)	1	0.30	342	0.96
AGT	S	1	0.30	311	0.87
TCA	S	8	<u>2.40</u>	428	<u>1.20</u>
TCC	S	2	0.60	316	0.89
TCG	S	4	<u>1.20</u>	232	0.65
TCT	S	4	<u>1.20</u>	506	<u>1.42</u>
ACA	T(Thr)	6	<u>1.50</u>	450	1.18
ACC	T	1	0.25	413	1.09
ACG	T	2	0.50	157	0.41
ACT	T	7	<u>1.75</u>	499	<u>1.31</u>
GTA	V(Val)	3	0.86	236	0.47
GTC	V	2	0.57	405	0.81
GTG	V	5	<u>1.43</u>	639	<u>1.28</u>
GTT	V	4	1.14	724	<u>1.45</u>
TGG	W(Trp)	6	1.00	363	1.00
TAC	Y(Tyr)	2	<u>1.33</u>	440	1.06
TAT	Y	1	0.67	288	0.94

注： 1.2 的 RSCU 值用下划线表示，说明该密码子使用具有较强的偏好性。

表 2 *CsCBF1* 基因与 6 个物种基因组的密码子频率的比值

密码子	氨基酸	C/A	C/N	C/T	C/Z	C/E	C/Y
GCA	A(Ala)	1.98	1.50	<u>2.25</u>	<u>2.07</u>	1.68	<u>2.15</u>
GCC	A	1.49	1.23	<u>0.47</u>	<u>0.49</u>	0.6	1.23
GCG	A	<u>2.99</u>	<u>4.64</u>	1.21	1.16	0.85	<u>4.41</u>
GCT	A	1.49	1.36	<u>2.58</u>	<u>2.01</u>	<u>2.71</u>	<u>2.01</u>
TGC	C(Cys)	1.6	1.60	0.85	0.95	1.67	<u>2.45</u>
TGT	C	<u>0.37</u>	<u>0.39</u>	0.75	0.69	0.7	<u>0.48</u>
GAC	D(Asp)	<u>2.68</u>	<u>2.73</u>	1.59	1.44	<u>2.48</u>	<u>2.28</u>
GAT	D	0.74	0.73	1.57	1.18	0.84	0.71
GAA	E(Glu)	1.12	1.07	<u>2.51</u>	1.93	1.01	0.79
GAG	E	1.31	1.44	1.11	1.04	<u>2.39</u>	<u>2.22</u>
TTC	F(Phe)	0.71	0.85	0.62	0.61	0.91	0.85
TTT	F	0.74	0.61	1.25	1.22	0.66	0.59
GGA	G(Gly)	0.64	0.66	1.03	1.15	1.71	1.41
GGC	G	0.84	0.69	<u>0.25</u>	<u>0.25</u>	<u>0.28</u>	<u>0.79</u>
GGG	G	1.51	1.47	0.86	1	1.36	<u>2.56</u>
GGT	G	1.21	1.21	1.97	1.91	1.1	1.12
CAC	H(His)	<u>2.21</u>	<u>2.21</u>	1.39	1.29	1.96	<u>2.5</u>
CAT	H	<u>0.28</u>	<u>0.29</u>	<u>0.46</u>	<u>0.38</u>	<u>0.28</u>	<u>0.28</u>
ATA	I(Ile)	0.61	0.55	1.15	0.92	1.42	<u>0.43</u>
ATC	I	<u>0.21</u>	<u>0.28</u>	<u>0.16</u>	<u>0.17</u>	<u>0.16</u>	<u>0.23</u>
ATT	I	0.72	0.55	1.31	1.11	0.52	0.51

续表 2

密码子	氨基酸	C/A	C/N	C/T	C/Z	C/E	C/Y
AAA	K(Lys)	<u>0.5</u>	<u>0.47</u>	1.47	1.02	<u>0.46</u>	<u>0.36</u>
AAG	K	1.18	1.15	1.02	0.98	<u>3.59</u>	1.25
CTA	L(Leu)	<u>0.39</u>	<u>0.41</u>	0.52	0.53	0.96	<u>0.29</u>
CTC	L	<u>0.48</u>	0.63	<u>0.28</u>	<u>0.3</u>	0.7	1.42
CTG	L	<u>0.39</u>	<u>0.38</u>	<u>0.17</u>	<u>0.15</u>	<u>0.08</u>	<u>0.37</u>
CTT	L	0.8	0.80	1.56	1.22	1.64	1.59
TTA	L	0	0.00	0	0	0	0
TTG	L	1.84	1.72	<u>3.18</u>	<u>2.96</u>	<u>2.75</u>	1.42
ATG	M(Met)	1.41	1.38	1.42	1.43	1.28	1.66
AAC	N(Asn)	<u>0.37</u>	<u>0.43</u>	<u>0.36</u>	<u>0.35</u>	<u>0.36</u>	<u>0.31</u>
AAT	N	1.03	0.82	<u>2.2</u>	1.71	1.24	0.64
CCA	P(Pro)	1.67	1.36	1.15	1.95	<u>3.17</u>	1.48
CCC	P	<u>2.18</u>	1.75	0.79	0.85	1.99	1.7
CCG	P	<u>2.68</u>	<u>4.62</u>	1.41	1.46	1.06	<u>4.35</u>
CCT	P	0.82	0.82	1.31	1.22	<u>2.11</u>	1.13
CAA	Q(Gln)	0	0.00	0	0	0	0
CAG	Q	0.51	0.51	<u>0.2</u>	<u>0.32</u>	<u>0.26</u>	0.64
AGA	R(Arg)	<u>0.4</u>	<u>0.48</u>	1.15	0.87	<u>2.65</u>	<u>0.36</u>
AGG	R	<u>2.8</u>	<u>2.52</u>	<u>2.37</u>	<u>2.08</u>	<u>16.19</u>	<u>3.34</u>
CGA	R	0.61	0.73	1.28	0.87	0.99	1.28
CGC	R	1.01	0.99	<u>0.3</u>	<u>0.27</u>	<u>0.18</u>	1.48
CGG	R	<u>2.35</u>	<u>3.12</u>	1.3	1.21	1.83	<u>6.79</u>
CGT	R	0	0.00	0	0	0	0
AGC	S(Ser)	<u>0.34</u>	<u>0.39</u>	<u>0.23</u>	<u>0.23</u>	<u>0.24</u>	<u>0.40</u>
AGT	S	<u>0.27</u>	<u>0.29</u>	0.58	<u>0.49</u>	<u>0.4</u>	<u>0.27</u>
TCA	<u>S</u>	1.68	1.75	<u>2.90</u>	<u>2.80</u>	<u>3.94</u>	1.64
TCC	<u>S</u>	<u>0.69</u>	0.75	<u>0.43</u>	<u>0.47</u>	<u>0.86</u>	<u>0.54</u>
TCG	S	1.65	<u>2.90</u>	1.44	1.44	1.77	1.81
TCT	S	0.61	0.77	1.49	1.28	1.77	0.65
ACA	T(Thr)	1.47	1.33	<u>2.54</u>	<u>2.2</u>	<u>2.81</u>	1.3
ACC	T	<u>0.37</u>	<u>0.40</u>	<u>0.2</u>	<u>0.23</u>	<u>0.17</u>	<u>0.31</u>
ACG	T	1	1.71	0.79	0.7	0.52	0.97
ACT	T	1.54	1.33	<u>2.93</u>	<u>2.52</u>	<u>2.96</u>	1.33
GTA	V(Val)	1.17	1.01	<u>2.18</u>	1.80	1.04	0.98
GTC	V	0.6	0.69	<u>0.36</u>	<u>0.37</u>	0.51	0.66
GTG	<u>V</u>	1.11	1.14	0.77	0.75	0.75	<u>1.81</u>
GTT	V	0.57	0.57	1.08	0.98	0.83	0.7
TGG	W(Trp)	1.85	1.89	1.92	1.78	1.52	<u>2.24</u>
TAC	Y(Tyr)	0.56	0.57	<u>0.37</u>	<u>0.4</u>	0.64	0.53
TAT	Y	<u>0.26</u>	<u>0.22</u>	<u>0.46</u>	<u>0.4</u>	<u>0.23</u>	<u>0.2</u>
TAA	*	<u>4.27</u>	<u>3.50</u>	<u>6.41</u>	<u>7.69</u>	1.92	<u>3.85</u>
TAG	*	0	0.00	0	0	0	0
TGA	*	0	0.00	0	0	0	0

*: 终止密码子; 下划线: 表示两个物种密码子比较具有明显偏差(0.5, 2)的分值。

麦、玉米密码子使用频率比值中, 2.0 或 0.5 的密码子分别有 26、24 种。表明虽然 *CsCBF1* 基因与这 4 个物种的密码子使用偏好性存在不同程度的差异, 但与双子叶植物的差异普遍小于与单子叶植物的差异。酵母和大肠杆菌分别作为真核和原核表达系统, 广泛用于基因的表达研究。为了解 *CsCBF1* 基因在这两类表达系统中的表达情况, 有必要进行密码子偏好性比较。结果表明, *CsCBF1* 基因与大肠

杆菌、酵母的密子使用频率比值中, 2.0 或 0.5 的密码子分别有 25、29 个, 说明对 *CsCBF1* 基因而言, 大肠杆菌表达系统可能要优于酵母表达系统。

2.3 与其他植物 *CBF* 基因密码子偏好性的比较

2.3.1 ENc、GC 及 GC3s 含量分析

各物种 *CBF* 基因的 ENc 值、GC 和 GC3s 含量计算结果见表 3。所选的 40 种植物的 *CBF* 基因的

表 3 39 个物种 *CBF* 的 ENc、GC 和 GC3s

物种	序列登录号	ENc	GC	GC3s
玉米(<i>Zea mays</i>)	NM_001158421	35.95	0.713	0.889
拟南芥(<i>Arabidopsis thaliana</i>)	NM_118681	54.43	0.507	0.460
扁桃 (<i>Prunus tenella</i>)	HQ908653	51.85	0.526	0.602
沙冬青(<i>Ammopiptanthus mongolicus</i>)	EU840990	47.18	0.591	0.671
毛果杨(<i>Populus trichocarpa</i>)	EF151456	37.58	0.502	0.404
黄瓜 (<i>Cucumis sativus</i>)	DQ776899	54.45	0.505	0.455
水稻(<i>Oryza sativa</i>)	AF243384	60.54	0.586	0.639
辣椒(<i>Capsicum frutescens</i>)	HM748942	48.87	0.436	0.304
蓝桉(<i>Eucalyptus globulus</i>)	DQ523829	42.63	0.624	0.808
碧桃(<i>Prunus persica</i>)	HM228387	55.75	0.532	0.609
小麦(<i>Tritium aestivum</i>)	AF376136	34.39	0.712	0.925
紫茎泽兰(<i>Ageratina adenophora</i>)	EF413000	54.53	0.519	0.541
柑橘(<i>Citrus trifoliata</i>)	FJ861083	40.21	0.561	0.627
马蔺(<i>Iris lacteal</i>)	DQ131497	55.50	0.513	0.475
甜橙(<i>Citrus Sinensis</i>)	FJ861084	37.13	0.558	0.622
川椒(<i>Capsicum annuum</i>)	AY368482	49.61	0.459	0.372
马铃薯(<i>Solanum tuberosum</i>)	DQ887515	40.43	0.421	0.289
冈尼桉(<i>Eucalyptus gunnii</i>)	DQ241821	41.69	0.624	0.822
甜樱桃(<i>Prunus avium</i>)	AB121674	51.55	0.546	0.647
西红柿(<i>Solanum lycopersicum</i>)	NM_001247194	47.31	0.433	0.337
小盐芥(<i>Thellungiella halophila</i>)	EF520733	50.63	0.578	0.633
沙棘(<i>Hippophae rhamnoides</i>)	EF502044	60.17	0.550	0.557
大豆(<i>Glycine max</i>)	NM_001248578	48.98	0.564	0.628
大白菜(<i>Brassica rapa</i>)	DQ022954	59.01	0.542	0.603
甘蓝型油菜(<i>Brassica napus</i>)	AY437878	57.99	0.540	0.571
黑麦草(<i>Lolium perenne</i>)	AY960831	49.15	0.688	0.637
葡萄(<i>Vitis riparia</i>)	AY390370	56.71	0.499	0.464
梨(<i>Pyrus Pyrifolia</i>)	HQ622623	50.74	0.588	0.643
榛子(<i>Corylus heterophylla</i>)	HQ639413	57.55	0.530	0.526
滨藜(<i>Atriplex halimus</i>)	JF451138	61.00	0.503	0.479
籼稻(<i>Oryza Sativa indica</i>)	HM807364	57.91	0.476	0.429
东方山羊豆(<i>Galega orientalis</i>)	FJ223566	47.34	0.451	0.376
芦荟(<i>Aloe vera</i>)	DQ981484	41.97	0.581	0.726
箬棕(<i>Sabal palmetto</i>)	DQ497730	46.35	0.643	0.769
苜蓿(<i>Medicago sativa</i>)	EU233782	58.09	0.449	0.420
紫衫(<i>Taxus cuspidate</i>)	EU549861	45.13	0.455	0.344
可可椰子(<i>Coco nucifera</i>)	DQ497739	44.72	0.645	0.772
散尾葵(<i>Dypsis lutescens</i>)	DQ497738	51.64	0.635	0.753
洲油棕榈(<i>Elaeis Oleifera</i>)	DQ497734	44.08	0.629	0.740

ENc值介于 34.39 到 61.00 之间, 平均值为 49.52。大部分已知的*CBF*基因均一性较好, 偏好性不强。一般来说, ENc值越小表明该基因的表达水平越高, ENc值小于 30 或大于 55 的基因可被预测为高表达或低表达基因^[19]。从数据看, 扁桃、川椒(*Capsicum annuum*)、甜樱桃(*Prunus avium*)、小盐芥(*Thellungiella halophila*)、黑麦草(*Lolium perenne*)、梨(*Pyrus Pyrifolia*)、散尾葵(*Dypsis lutescens*)、茶树*CBF*基因的ENc值均在 50 左右, 其表达水平一般; 而水稻(*Oryza sativa*)等 11 种植物*CBF*基因的ENc值均大于 55, 表达水平较低; 玉米、毛果杨(*Populus trichocarpa*)、小麦、甜橙(*Citrus Sinensis*)4 种植物的*CBF*基因的

ENc 值在 35 左右, 表达水平较高。在 40 种植物中只有 9 种植物的 GC 含量小于 0.5, 预示大部分 *CBF* 基因对 GC 有一定的偏好性。GC3s 值大于 0.5 的占 66%, 说明大部分 *CBF* 基因偏好以 G/C 结尾的密码子。

2.3.2 基于 *CBF* 基因密码子偏好性的聚类分析

基于 *CBF* 基因密码子偏好性的聚类分析结果将各样本序列将按照各自的密码子使用偏性分布到 8 个大类中。聚类分析结果见图 1。参照图 1 可以看出, 在 40 条样本序列中, 有 29 条被分在第 1 大类和第 2 大类中, 而毛果杨、黑麦草、芦荟(*Aloe vera*)、洲油棕榈(*Elaeis Oleifera*)这 4 个植物各成一类。表

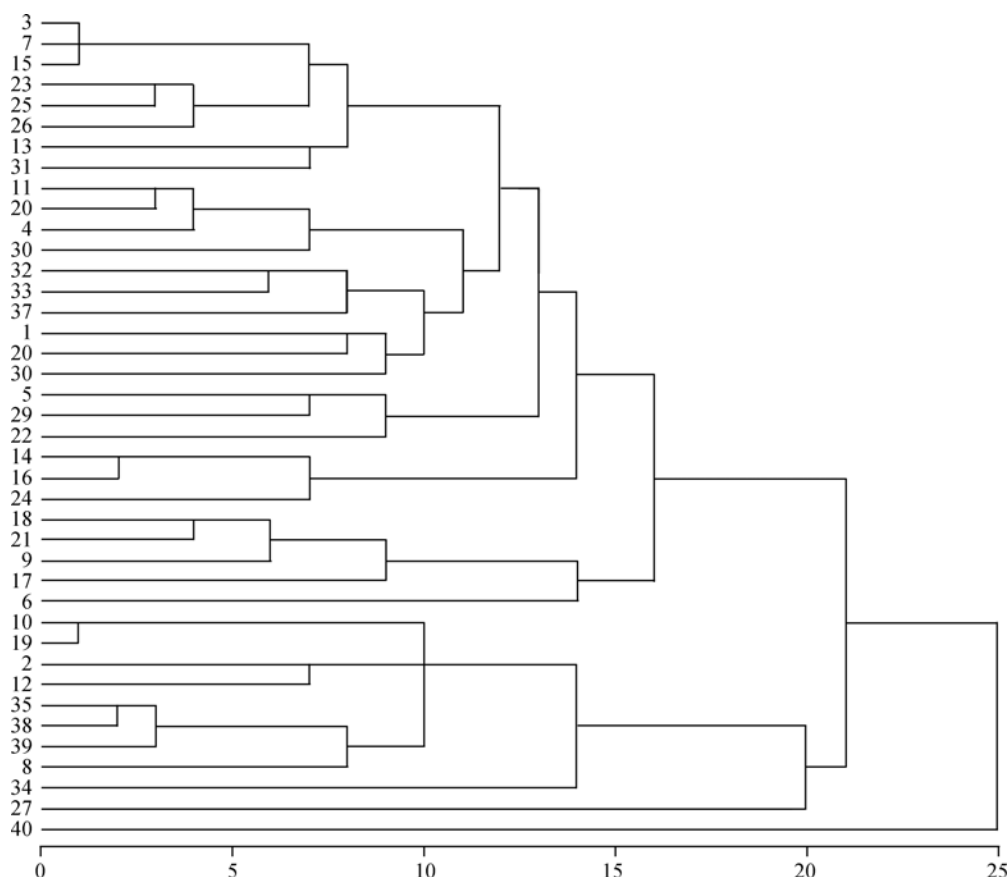


图 1 基于 *CBF* 基因的相对同义密码子使用度的聚类树状图

1 : *Camellia sinensis*; 2 : *Zea mays*; 3 : *Arabidopsis thaliana*; 4 : *Prunus tenella*; 5 : *Ammopiptanthus mongolicus* 6 : *Populus trichocarpa*; 7 : *Cucumis sativus*; 8 : *Oryza sativa*; 9 : *Capsicum frutescens*; 10 : *Eucalyptus globulus*; 11 : *Prunus persica*; 12 : *Triticum aestivum*; 13 : *Ageratina adenophora*; 14 : *Citrus trifoliata*; 15 : *Iris lacteal*; 16 : *Citrus Sinensis*; 17 : *Capsicum annuum*; 18 : *Solanum tuberosum*; 19 : *Eucalyptus gunnii*; 20 : *Prunus avium*; 21 : *Solanum lycopersicum*; 22 : *Thellungiella halophila*; 23 : *Hippophae rhamnoides*; 24 : *Glycine max*; 25 : *Brassica rapa*; 26 : *Brassica napus*; 27 : *Lolium perenne*; 28 : *Vitis riparia*; 29 : *Pyrus Pырifolia*; 30 : *Corylus heterophylla*; 31 : *Atriplex halimus*; 32 : *Oryza Sativa indica*; 33 : *Galega orientalis*; 34 : *Aloe vera*; 35 : *Sabal palmetto*; 36 : *Medicago sativa*; 37 : *Taxus cuspidate*; 38 : *Coco nucifera*; 39 : *Dypsis lutescens*; 40 : *Elaeis Oleifera*.

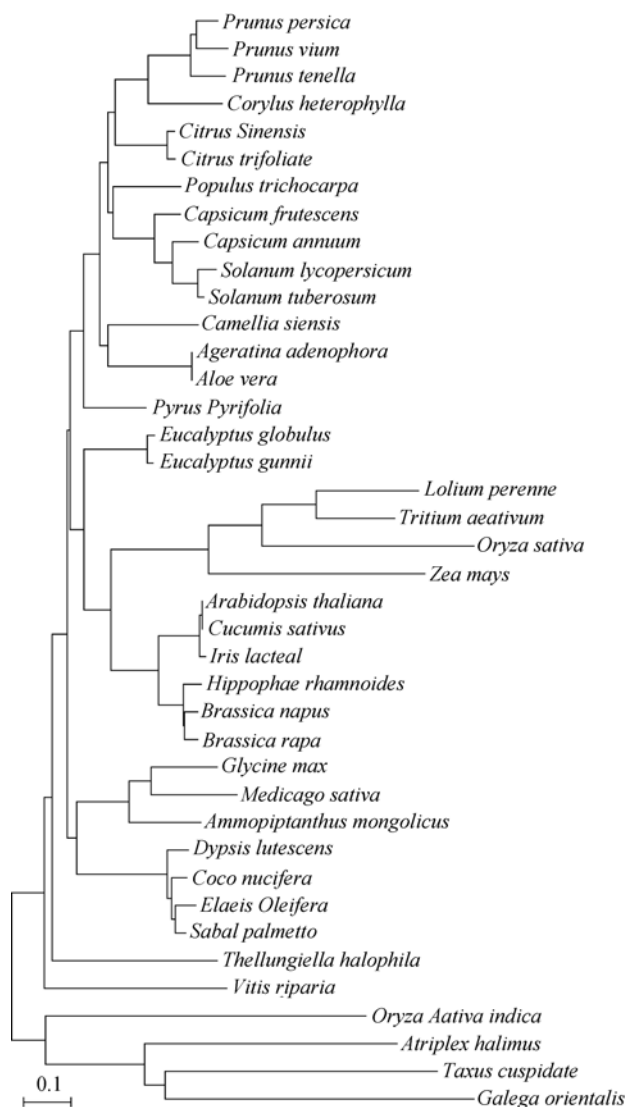


图 2 基于 CBF 基因 CDS 的聚类树状图

明这 4 种植物 CBF 基因在密码子使用上与其他植物 CBF 基因具有明显的差异性。从图 1 中可以看出在第 1、2 大类内部形成了若干个聚类中心，茶树与葡萄(*Vitis riparia*)、苜蓿(*Medicago sativa*)被分成一小类，说明他们对密码子的用法相似。同时也可以看出，基于 RSCU 的聚类没有反映出植物之间的亲缘关系，比如水稻没有和同为禾本科的玉米、小麦单独聚成一小类，而是和箬棕(*Sabal palmetto*)、可可椰子(*Coco nucifera*)、散尾葵首先聚在一起。这一聚类结果与基于 CBF 基因的 CDS 构建的系统进化树(图 2)不一致，基于 CDS 的系统进化树更接近这 40 种植物的真实系统分类，表明 CBF 基因 CDS 的差异大小与物种的亲缘关系远近有一定关系。

3 讨论

同一物种的不同基因对密码子的用法特征都可能存在差异。本研究通过分析 *CsCBF* 基因的密码子用法特征，发现该基因偏好使用以 G/C 结尾的密码子，对 22 种密码子具有偏好性，其中使用频率最高的是 AGG(3.20)，其次是 TTG(3.16)。而将 *CsCBF1* 基因与筛选的 70 个茶树基因密码子用法比较，发现筛选的 70 个茶树基因偏好使用的密码子有 15 种，大多以 A/T 结尾。但没有表现出特别的密码子偏好性。可见 *CsCBF1* 与茶树其他基因在密码子使用上有较大差异，这可能与基因表达水平、蛋白质结构和功能等因素有关，引起 *CsCBF1* 密码子偏好性的因素及各因素所起作用的大小等还需要进一步研究。

转基因研究常常需要进行基因的异源表达，而不同物种基因间密码子用法差异较大，因而，分析外源基因与宿主系统的密码子特性，并对一些密码子进行改造，采用符合宿主基因组的密码子，对提高转录、翻译效率，进而提高表达量有重要意义^[25]。本研究通过将 *CsCBF1* 的密码子用法与 4 种模式植物基因组密码子用法比较，发现 *CsCBF1* 基因与单子叶植物小麦、玉米的密码子偏好性差异大于与同为双子叶植物的拟南芥、烟草的密码子偏好性差异。因此拟南芥和烟草更适合作为 *CsCBF1* 基因的外源表达系统。若 *CsCBF1* 基因要在拟南芥和烟草中高效表达，还需对部分偏好性差异较大的密码子进行改造。

通过比较不同物种的 CBF 基因可知，*CsCBF1* 基因的密码子用法与葡萄、苜蓿的密码子的用法相似，在聚类分析中被聚为一类；而与毛果杨、黑麦草、芦荟、洲油棕榈这 4 种植物的 CBF 基因密码子用法差异显著。同时也发现了大部分 CBF 基因对 G/C 有一定的偏好性，可能与 CBF 基因的独特功能有关。玉米、毛果杨、小麦、甜橙这 4 种植物的 CBF 基因表达水平较高，其他植物的 CBF 基因表达水平一般。密码子偏好性分析既可以为基因表达选择合适的宿主表达系统，同时也可通过改造密码子提高基因的表达量提供依据。但基于基因 RSCU 的聚类分析与基于 CDS 的进化分析结果可能存在一定差异，晁岳恩等^[26]、刘汉梅等^[27]的研究中也曾出现这种现象。

由于单基因在进化上有可能存在较大的突变, 导致其密码子使用偏爱性发生较大变化, 而表现出与真实的分类地位存在差异。Christianson^[28]认为密码子偏好性的聚类并不一定完全和真实的系统发育关系对应, 在较小的分类单元里这一指标比较可靠, 而在分歧较大的分类单元里, 这一方法主要还是作为传统系统树构建方法的补充和参考。

参考文献(References):

- [1] 刘庆慧, 黄健, 韩文君. WSSV3 个编码蛋白的基因密码子偏爱性分析. 海洋水产研究, 2005, 26(4): 1-7. [DOI](#)
- [2] 吴宪明, 吴松锋, 任大明, 朱云平, 贺福初. 密码子偏性的分析方法及相关研究进展. 遗传, 2007, 29(4): 420-426. [DOI](#)
- [3] Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics*, 1991, 129(3): 897-907. [DOI](#)
- [4] Romero H, Musto H, Zavala A. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic Acids Res*, 2000, 28(10): 2084-2090. [DOI](#)
- [5] 周祎. 基于相对密码子频率的茛蒿花叶病毒基因序列的进化分析[学位论文]. 湖南大, 2009. [DOI](#)
- [6] 张文娟. 基于密码子水平的生物信息学分析及进化研究[学位论文]. 复旦大学, 2006. [DOI](#)
- [7] Hiraoka Y, Kawamata K, Haraguchi T, Chikashige Y. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. *Genes Cells*, 2009, 14(4): 499-509. [DOI](#)
- [8] Dutet L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet*, 2000, 16(7): 287-289. [DOI](#)
- [9] 孙铮, 马亮, Murphy R, 张宪省, 黄大卫. *Wolbachia pipientis* wMEL基因组水平上的密码子使用分析. 中国科学 C 辑 (生命科学), 2009, 39(10): 948-953. [DOI](#)
- [10] Gu WJ, Zhou T, Ma JM, Sun X, Lu ZH. Folding type specific secondary structure propensities of synonymous codons. *IEEE Trans Nanobiosci*, 2003, 2(3): 150-157. [DOI](#)
- [11] Gu WJ, Zhou T, Ma JM, Sun X, Lu ZH. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. *Biosystems*, 2004, 73(2): 89-97. [DOI](#)
- [12] 石秀凡, 黄京飞, 梁宠荣, 柳树群, 谢君, 刘次全. 人类基因中同义密码子的偏好与密码子-反密码子间的结合强度密切相关吗? 科学通报, 2000, 45(23): 2520-2525. [DOI](#)
- [13] Wang Y, Jiang CJ, Li YY, Wei CL, Deng WW. *CsICE1* and *CsCBF1*: two transcription factors involved in cold responses in *Camellia sinensis*. *Plant Cell Rep*, 2012, 31(1): 27-34. [DOI](#)
- [14] 陈林波, 李叶云, 王琴, 高永亮, 江昌俊. 茶树冷诱导基因 *RAV* 的克隆与表达特性分析. 植物生理学通讯, 2010, 46(4): 354-358. [DOI](#)
- [15] 陈林波, 房超, 王郁, 李叶云, 江昌俊, 梁名志. 茶树抗逆相关基因 *ERF* 的克隆与表达特性分析. 茶叶科学, 2011, 31(1): 53-58. [DOI](#)
- [16] Wang L, Li XW, Zhao Q, Jing SL, Chen SF, Yuan HY. Identification of genes induced in response to low-temperature treatment in tea leaves. *Plant Mol Biol Rep*, 2009, 27(3): 257-265. [DOI](#)
- [17] 柏锡, 徐建震, 李杰, 郭政, 李琳, 朱延明. 马铃薯密码子用法分析及其在 *t-PA* 基因密码子改造上的应用. 遗传, 2004, 26(1): 75-83. [DOI](#)
- [18] 刘庆坡, 谭军, 薛庆中. 籼稻品种 93-11 同义密码子的使用偏性. 遗传学报, 2003, 30(4): 335-340. [DOI](#)
- [19] Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol*, 1986, 24(1-2): 28-38. [DOI](#)
- [20] 李平, 白云凤, 冯瑞云, 王原媛, 张维锋. 籽粒苋苹果酸酶(NAD-ME)基因密码子偏好性分析. 应用与环境生物学报, 2011, 17(1): 12-17. [DOI](#)
- [21] Mathé C, Peresetsky A, Déhais P, Van Montagu M, Rouze P. Classification of *Arabidopsis thaliana* gene sequences: clustering of coding sequences into two groups according to codon usage improves gene prediction. *J Mol Biol*, 1999, 285(5): 1977-1991. [DOI](#)
- [22] 赵洋, 杨阳, 刘振, 杨培迪. 茶树密码子用法分析. 茶叶科学, 2011, 31(4): 319-325. [DOI](#)
- [23] 胡桂兵, 张上隆, 徐昌杰, 林顺权. 柑橘密码子用法分析. 果树学报, 2006, 23(3): 479-485. [DOI](#)
- [24] 周猛, 童春发, 施季森. 杨树同义密码子用法的初步分析. 植物生理与分子生物学报, 2007, 33(4): 285-293. [DOI](#)
- [25] 赵耀, 刘汉梅, 顾勇, 黄玉碧. 玉米 *waxy* 基因密码子偏好性分析. 玉米科学, 2008, 16(2): 16-21. [DOI](#)
- [26] 晁岳愚, 吴政卿, 杨会民, 何宁, 杨攀. 11 种植物 *psbA* 基因的密码子偏好性及聚类分析. 核农学报, 2011, 25(5): 927-932. [DOI](#)
- [27] 刘汉梅, 赵耀, 顾勇, 黄玉碧, 张怀渝. 几种植物 *waxy* 基因的密码子用法特性分析. 核农学报, 2010, 24(3): 476-481. [DOI](#)

-
- [28] Christianson ML. Codon usage patterns distort phylogenies from or of DNA sequences. *Am J Bot*, 2005, 92(8): 1221–1233. [DOI](#)