

红松查尔酮合成酶基因 *CHS* 密码子偏好性分析

李翔<sup>1</sup> 范作义<sup>2</sup> 王井源<sup>2</sup> 王淇<sup>2</sup> 李喜鹏<sup>2</sup> 王德秋<sup>2</sup>  
孔令远<sup>2</sup> 曹森林<sup>2</sup> 孟庆刚<sup>2</sup> 赵曦阳<sup>1\*</sup>

(1. 林木遗传育种国家重点实验室, 东北林业大学林学院, 哈尔滨 150040 ; 2. 吉林省临江林业局, 临江 134600 )

**摘要** 查尔酮合成酶( Chalcone synthase *CHS* )广泛存在于植物体内, 是花色素形成过程中一种重要的酶, 可以进一步催化生成黄酮类化合物。本研究采用 CodonW 和 EMBOSS 在线软件对红松查尔酮合成酶基因 *CHS* 的密码子使用偏好性进行分析, 并与北美乔松等其他 24 种植物的 *CHS* 基因以及模式植物基因组进行比较, 对认识红松 *CHS* 基因的密码子使用偏好性, 为选择适宜的表达系统奠定了一定的基础。研究表明, 红松 *CHS* 基因编码区的有效密码子数( ENC )和 GC 含量分别为 48.92 和 0.548, C + G 含量高于 A + T 含量, 密码子偏好以 A/T 结尾; 多数植物 *CHS* 基因的 G + C 含量高于 A + T 含量, 且密码子更偏好 C/G 结尾。聚类分析表明, 红松与马尾松和赤松的密码子使用偏好性的相似性较高。密码子使用频率研究发现, 红松 *CHS* 遗传转化与异源表达较优的受体可能是大肠杆菌和拟南芥。

**关键词** 红松 查尔酮合成酶 密码子偏好性 聚类分析

中图分类号 S791.247 文献标志码 A doi :10.7525/j.issn.1673-5102.2020.03.017

Codon Usage Bias of Chalcone Synthase Gene *CHS* in *Pinus koraiensis*

LI Xiang<sup>1</sup> FAN Zuo-Yi<sup>2</sup> WANG Jing-Yuan<sup>2</sup> WANG Qi<sup>2</sup> LI Xi-Peng<sup>2</sup> WANG De-Qiu<sup>2</sup>  
KONG Ling-Yuan<sup>2</sup> CAO Sen-Lin<sup>2</sup> MENG Qing-Gang<sup>2</sup> ZHAO Xi-Yang<sup>1\*</sup>

(1. State Key Laboratory of Tree Genetics and Breeding School of Forestry, Northeast Forestry University, Harbin 150040 ; 2. Linjiang Forestry Bureau of Jilin Province, Linjiang 134600 )

**Abstract** Chalcone synthase( *CHS* ) is widely found in plants and is a key enzyme in anthocyanin synthesis pathway. It can catalyze the binding of malonyl coenzyme A and coumaryl coenzyme A to generate flavonoids. CodonW and EMBOSS were used to analyze the codon usage bias of *CHS* gene in *Pinus koraiensis*. The *CHS* gene of *P. koraiensis* were compared with those of the other 24 plants and model plant genome, which laid a foundation for the selection of appropriate expression system. The effective codon number( ENC ) and GC content of *CHS* gene coding region were 48.92 and 0.351, respectively. The content of C + G was higher than that of A + T, and the codon preferred to end with A/T. The G + C content of most plants *CHS* genes is higher than that of A + T, and the codon preferred to end with C/G. By cluster analysis, the codon usage bias of *P. koraiensis* was similar to *P. massoniana* and *P. densiflora* pdchsx. In addition, it was found that *Arabidopsis thaliana* and *Escherichia coli* might be ideal receptor systems for *CHS* genetic transformation and heterogeneous expression.

**Key words** *Pinus koraiensis* chalcone synthase codon usage bias cluster analysis

基金项目 红松种质资源遗传多样性评价项目( KJZXA2019046 )  
第一作者简介 李翔( 1993— ) 男, 博士研究生, 主要从事红松遗传改良方面的研究。

\* 通信作者 E-mail :zhaoxyphd@163.com  
收稿日期 2019-08-25

Foundation item Genetic Diversity Evaluation Project of *Pinus koraiensis* Germplasm Resources( KJZXA2019046 )  
First author introduction LI Xiang( 1993— ) male doctoral candidate specializing in Genetic improvement of *Pinus koraiensis*.

\* Corresponding author E-mail :zhaoxyphd@163.com  
Received date 2019-08-25

在生物体内,三联体密码子是由信使 RNA 分子中四个碱基组成 3 个核苷酸形成的,编码一个特定的氨基酸,进行遗传信息的传递。一般地,密码子存在简并性,即每个特定的氨基酸可对应 1~6 种密码子<sup>[1]</sup>。在生物的长期进化过程中,由于受到自然选择压力、变异和突变的影响,不同生物体在编码氨基酸时并不是完全随机选择密码子,而是对某些特定的同义密码子存在一定的使用偏好性,即存在密码子使用偏好性现象(codon bias),该现象在自然界中普遍存在。目前,在细菌、哺乳动物和植物等多个物种中,均观察到明显的密码子使用偏好性现象<sup>[2]</sup>。密码子偏好性不但可以影响蛋白质翻译的速度和折叠,还可以影响 mRNA 水平的转录调控及外源基因的表达量<sup>[3-4]</sup>。前人研究发现,在大量组织中普遍表达且表达水平相对恒定的基因的表达量与密码子使用偏好性存在较强的相关性,这些基因受到了明显的翻译选择作用<sup>[5]</sup>。另外,不同物种、不同生物体的基因、基因家族或基因组密码子在使用上存在很大的差异,对功能基因的翻译调控产生重要的影响<sup>[6-7]</sup>,进而会影响基因的功能。即使在一个基因内,密码子的使用偏好性也会有很大的差异。因此,分析同义密码子的使用偏好性将有助于揭示物种的起源和分子进化规律,提高翻译的准确性和效率,为研究生物体进化、基因功能、蛋白质结构,提高外源基因的表达量提供了重要信息,同时,可利用编码区和非编码区的基因组特征差异进行全基因组扫描,发现新基因,对研究基因功能和蛋白质功能具有重要的参考价值<sup>[8]</sup>。

目前,关于密码子使用偏好性的研究表明,密码子偏好性是通过选择、突变和遗传漂变之间的平衡来维持的。密码子偏好性可能与 3 个原因有关,一是避免转录终止,影响基因表达,二是密码子与反密码子的相互作用强度适中,有利于蛋白质生物合成的迅速进行,三是密码子偏好性可提高翻译效率,增加细胞内 tRNA 的含量。此外,密码子偏好性受其他因素影响,如基因的碱基组成、GC 含量、基因长度、表达水平、蛋白质亲水性、蛋白质二级结构、表达水平等<sup>[9]</sup>。近年来,随着高通量测序技术的快速发展,多个物种基因组测序已经完成,将有助于在全基因组水平上揭示密码子使用偏好性。目前,已在多个模式生物和非模式生物中对单基因或基因组密码子使用偏好性进行了深入的研究,如毛果杨(*Populus trichocarpa*)<sup>[10]</sup>

水稻(*Oryza sativa*)<sup>[11]</sup>和苹果(*Malus domestica*)<sup>[12]</sup>等。而红松(*Pinus koraiensis*)基因组测序尚未完成,相关分子研究基础较为薄弱,在红松中还未见对单基因进行密码子偏好性分析的报道。

红松为松科(Pinaceae)松属(*Pinus* Linn.)常绿针叶树,是我国东北地区珍贵的用材树种和经济林树种,具有重要的经济价值和生态价值<sup>[13-14]</sup>。其材质轻软、耐腐蚀性强,树皮富含松多酚,具有极高的开发价值。多酚是一种广泛存在于植物的根、叶、皮等器官中的一种次生代谢化合物,具有多种生理功能,可降低血脂,减缓衰老,现已在材料、食品等多个领域被广泛应用。松多酚是松科植物中特有的多酚类化合物,其含量丰富,具有抗炎镇痛、降血压等功效,科研工作者对松多酚开展了大量的研究<sup>[15]</sup>。研究发现,松多酚中含有类黄酮和花色素,使植物的花、果实等器官呈现不同的颜色<sup>[16]</sup>。关于松多酚的分子研究较为广泛,松多酚合成过程中相关功能基因很多,主要包括查尔酮合成酶(CHS)、黄酮醇合成酶(FLS)和查尔酮异构酶(CHI)等。其中,查尔酮合成酶(CHS)是植物中黄酮、花色素和异黄酮等化合物合成的关键酶之一,在多酚合成过程中发挥着重要的作用,可影响植物的花色及防御反应<sup>[17]</sup>。查尔酮合成酶基因 CHS 于 1983 年首次在荷兰芹(*Petroselinum crispum*)<sup>[18]</sup>中发现,现已在多个植物中克隆得到了 CHS 基因,如甘薯(*Pomoea batatas*)<sup>[19]</sup>、桂花(*Osmanthus fragrans*)<sup>[20]</sup>和樟子松(*Pinus sylvestris*)<sup>[21]</sup>和拟南芥(*Arabidopsis thaliana*)<sup>[22]</sup>等。而关于 CHS 基因密码子偏好性的研究较少,仅有部分植物研究中有报道。前人研究发现,不同植物中的 CHS 密码子偏好性存在一定差异,如蓼科大黄属物种<sup>[23]</sup>和猕猴桃中 CHS 基因<sup>[8]</sup>最优密码子偏爱以 G/C 结尾,而豆科植物 CHS 基因<sup>[24]</sup>对 A 或 U 结尾的密码子偏好性较强。目前,关于红松 CHS 基因方面的报道较少<sup>[25-26]</sup>,相关功能还不清楚,开展 CHS 密码子偏好性分析将有助于研究该基因的遗传特征和分子机理,进而为有效利用松多酚资源奠定一定的基础。笔者从 GenBank 中获得红松查尔酮合成酶基因 CHS 序列,采用 EMBOSS 在线程序、MEGA7.0 软件<sup>[27]</sup>和 CodonW 分析其密码子偏好性,运用 R 语言软件(version 3.5.0)<sup>[28]</sup>和 SPSS (version 23.0)软件<sup>[29]</sup>统计密码子偏好性参数,为 CHS 基因功能和分子进化的深入研究提供理论依据。此外,将 CHS 的密码子偏好性与大肠杆菌、酵

母、烟草和拟南芥的基因组密码子进行比较,以期  
为筛选该基因的理想表达系统奠定基础。

1 材料和方法

1.1 序列来源

本研究采用的红松查尔酮合成酶基因 *CHS* 编  
码区序列来自 NCBI 数据库( <https://www.ncbi.nlm.nih.gov/> ),Genbank 登录号为 KJ796482.1。  
以 Chalcone Synthase 为关键词从 NCBI 中检索并  
下载苹果( *Malus domestica* )、北美乔松( *Pinus stro-*  
*bus* )、大豆( *Glycine max* )、猕猴桃( *Actinidia chinensis* )和玉米( *Zea mays* )等其他 24 种植物 *CHS* 基因  
的编码序列用于本研究,登录号见表 1。

表 1 不同植物 *CHS* 基因完整 CDS 编码序列  
Table 1 Complete CDS sequence coding frame of *CHS*  
gene in different plants

序号 Number	物种 Species	编号 ID	登陆号 Accession number
1	红松 <i>Pinus koraiensis</i>	Pk	KJ796482.1
2	甘蓝 <i>Brassica oleracea</i>	Bo	HQ189776.1
3	北美乔松 <i>Pinus strobus</i>	Ps	AJ004800.1
4	碧桃 <i>Prunus persica</i>	Pp	HM543568.1
5	草莓 <i>Fragaria × ananassa</i>	Fa	AB201758.1
6	大豆 <i>Glycine max</i>	Gm	XM_003553938.4
7	杜鹃花 <i>Rhododendron simsii</i>	Rs	AJ413277.1
8	柑橘 <i>Citrus sinensis</i>	Cs	AB009350.1
9	高粱 <i>Sorghum bicolor</i>	Sb	AF152553
10	菊花 <i>Chrysanthemum × morifolium</i>	Cm	DQ521272.1
11	大麦 <i>Hordeum vulgare</i>	Hv	AY286093
12	猕猴桃 <i>Actinidia chinensis</i>	Ac	KF157394.1
13	拟南芥 <i>Arabidopsis thaliana</i>	At	NM_121396.4
14	苹果 <i>Malus domestica</i>	Md	AB074485.1
15	葡萄 <i>Vitis vinifera</i>	Vv	NM_001281135.1
16	水稻 <i>Oryza sativa</i>	Os	AB000801.2
17	王府海棠 <i>Malus hybrid</i>	Mh	FJ599763.1
18	西洋梨 <i>Pyrus communis</i>	Pc	DQ901397.2
19	烟草 <i>Nicotiana tabacum</i>	Nt	NM_001326166.1
20	玉米 <i>Zea mays</i>	Zm	NM_001148774.1
21	赤松 <i>Pinus densiflora pdchsx</i>	Pd	AB015490
22	毛果杨 <i>Populus trichocarpa</i>	Pt	DQ371804
23	马尾松 <i>Pinus massoniana</i>	Pm	KF303883
24	辐射松 <i>Pinus radiata</i>	Pr	KF704817
25	海岸松 <i>Pinus pinaster</i>	Pp1	AY168850

1.2 分析方法

1.2.1 密码子偏好性参数分析

将 25 种植物的 CDS 序列整理后,利用分析平  
台 Galaxy( version 1.4.4 )( <https://galaxy.pasteur.fr/> )中的 CodonW 模块和 EMBOSS 在线软件( <http://www.bioinformatics.nl/emboss-explorer/> )的  
CUSP 模块与 CHIPS 模块对 *CHS* 进行密码子偏好  
性参数分析,包括同义密码子相对使用度  
( RSCU )、密码子适应指数( CAI )、有效密码子数  
( ENC )、密码子 G 和 C 总含量( GC )、密码子末位  
碱基 G 或 C 的含量( GC<sub>3</sub>s )和密码子使用频率  
( frequency )等。从 Condo Usage Database( <http://www.kazusa.or.jp/codon/> )数据库中获得模式生  
物拟南芥( *Arabidopsis thaliana* )、酵母( *Saccharomy-*  
*ces cerevisiae* )、大肠杆菌( *Escherichia coli* )和烟草  
( *Nicotiana sylvestris* )基因组的密码子使用频率数据。

RSCU 是指某些特定密码子在同义密码子中  
的相对概率,一般地,RSCU 值大于 1,表示该密码  
子使用偏好性强,而 RSCU 值小于 1,表示无使用  
偏好性<sup>[10]</sup>。ENC 值可反映所有密码子中同义密  
码子使用偏好程度,其分布范围为 20 ~ 61,该值是  
反应密码子使用偏离随机选择的程度。ENC 值越  
靠近 20,该基因的密码子偏好性越强,而越靠近  
61,偏好性越弱<sup>[30]</sup>。CAI 值范围为 0 ~ 1,该值越  
接近 1,表明该基因的偏好性越强,而越接近 0,则  
偏好性越弱<sup>[31]</sup>。密码子的第 3 位碱基中 G + C 的  
含量与第 3 位碱基总量的比值即 GC<sub>3</sub>s。GC 含量  
是指目的序列中鸟嘌呤( G )和胞嘧啶( C )所占的  
比率。

1.2.2 聚类分析

基于密码子使用偏好性的聚类分析是以不同  
物种的简并密码子( 去除终止密码子 TAA、TAG 和  
TGA,起始密码子 ATG 和甲硫氨酸 TGG )的 RSCU  
值为变量,采用 R 语言中的 pheatmap 包( <https://cran.r-project.org/web/packages/pheatmap/index.html> )绘制系统聚类热图。基于 CDS 编码区序列  
的聚类分析是采用 MEGA7.0( Neighbour-Joining,  
NJ )构建的系统发育树,并利用在线软件绘图工具  
iTOL( Interactive Tree Of Life, <https://itol.embl.de/> )对进化树进行可视化展示。

2 结果分析

2.1 红松 *CHS* 密码子使用偏好性分析

对红松 *CHS* 基因使用频率和 RSCU 值统计发

玢( 详见表 2 ) ,61 个密码子中使用频率最大的是 AAG 和 GAG( 45.34% ) ,26 个密码子 RSCU 值大于 1 ,是红松 *CHS* 基因的偏好密码子。GCC、TTC、GGA、ATC、CCC、AGA、AGT、CTC、CTG、TAC 共 10 个密码子的 RSCU 值大于 1.5 ,表明是该基因的高频率密码子 ,其中 GGA( 2.06 )、CCC( 2.48 )、AGA ( 2.47 )、CTG( 2.31 )这 4 个密码子的 RACU 值均大于 2 ,表明该基因对这些密码子具有极强的偏好

性。我们发现 ,编码 Leu 的 CTA 的 RSCU 值为 0 ,说明该密码子可能不参与红松 *CHS* 基因的翻译过程。此外 ,红松中 NTA 型密码子( GTA :0.38 ,ATA 0.33 ,TTA 0.15 ,CTA 0 )的 RSCU 值较低 ,这可能有利于 mRNA 的降解以提高蛋白含量 ,NCG 型的密码子 RSCU 值( CCG :0.19 ,GCG :0.48 ,TCG 1.43 ,ACG 0.50 )较低 ,表明红松可能是一种甲基化程度较高的植物。

表 2 红松 *CHS* 基因同义密码子相对使用度

Table 2 RSCU of *CHS* in *P. koraiensis*

密码子 Codon	氨基酸 Amino acid	比例 Fraction	频率 Frequency	数目 Number	相对密码子 使用度 RSCU	密码子 Codon	氨基酸 Amino acid	比例 Fraction	频率 Frequency	数目 Number	相对密码子 使用度 RSCU
GCA	A( Ala )	0.212	17.632	7	0.85	CCA	P( Pro )	0.143	7.557	3	0.57
<u>GCC</u>	A	0.394	32.746	13	1.58	CCC	P	0.619	32.746	13	2.48
GCG	A	0.121	10.076	4	0.48	CCG	P	0.048	2.519	1	0.19
GCT	A	0.273	22.670	9	1.09	CCT	P	0.190	10.076	4	0.76
<u>TGC</u>	Q( Cys )	0.750	15.113	6	1.50	CAA	Q( Gln )	0.263	12.594	5	0.53
TGT	C	0.250	5.038	2	0.50	CAG	Q	0.737	35.264	14	1.47
<u>GAC</u>	D( Asp )	0.609	35.264	14	1.22	AGA	R( Arg )	0.412	17.632	7	2.47
GAT	D	0.391	22.670	9	0.78	AGG	R	0.235	10.076	4	1.41
GAA	E( Glu )	0.250	15.113	6	0.50	CGA	R	0.000	0.000	0	0.00
<u>GAG</u>	E	0.750	45.340	18	1.50	CGC	R	0.118	5.038	2	0.71
<u>TTC</u>	F( Phe )	0.867	32.746	13	1.73	CGG	R	0.176	7.557	3	1.06
TTT	F	0.133	5.038	2	0.27	CGT	R	0.059	2.519	1	0.35
<u>GGA</u>	G( Gly )	0.515	42.821	17	2.06	AGC	S( Ser )	0.238	12.594	5	1.43
<u>GGC</u>	G	0.303	25.189	10	1.21	AGT	S	0.286	15.113	6	1.71
GGG	G	0.061	5.038	2	0.24	TCA	S	0.000	0.000	0	0.00
GGT	G	0.121	10.076	4	0.48	TCC	S	0.190	10.076	4	1.14
<u>CAC</u>	H( His )	0.714	12.594	5	1.43	TCG	S	0.238	12.594	5	1.43
CAT	H	0.286	5.038	2	0.57	TCT	S	0.048	2.519	1	0.29
ATA	I( Ile )	0.111	5.038	2	0.33	ACA	T( Thr )	0.250	15.113	6	1.00
<u>ATC</u>	I	0.667	30.227	12	2.00	ACC	T	0.292	17.632	7	1.17
ATT	I	0.222	10.076	4	0.67	ACG	T	0.125	7.557	3	0.50
AAA	K( Lys )	0.280	17.632	7	0.56	ACT	T	0.333	20.151	8	1.33
<u>AAG</u>	K	0.720	45.340	18	1.44	GTA	V( Val )	0.094	7.557	3	0.38
CTA	L( Leu )	0.000	0.000	0	0.00	GTC	V	0.375	30.227	12	1.50
<u>CTC</u>	L	0.308	30.227	12	1.85	GTG	V	0.281	22.67	9	1.12
<u>CTG</u>	L	0.385	37.783	15	2.31	GTT	V	0.250	20.151	8	1.00
CTT	L	0.154	15.113	6	0.92	TGG	W( Trp )	1.000	10.076	4	1.00
TTA	L	0.026	2.519	1	0.15	TAC	Y( Tyr )	0.778	17.632	7	1.56
TTG	L	0.128	12.594	5	0.77	TAT	Y	0.222	5.038	2	0.44
ATG	M( Met )	1.000	32.746	13	1.00	TAA	TER	1.000	2.519	1	3.00
AAC	N( Asn )	0.545	15.113	6	1.09	TAG	TER	0.000	0.000	0	0.00
AAT	N	0.455	12.594	5	0.91	TGA	TER	0.000	0.000	0	0.00

注 :下划线表示红松 *CHS* 基因该密码子使用频率较高  
Note :The underlines mean codon usage with high frequency in *CHS* of *P. koraiensis*

表 3 物种间 *CHS* 密码子偏好性分析

Table 3 Analysis of codon usage bias of *CHS* among different plants

物种 Species	ENC	CAI	GC <sub>3</sub> s	GC	RSCU > 1	RSCU > 1.5
大麦 <i>Hordeum vulgare</i>	31.72	0.779	0.935	0.659	25	19
高粱 <i>Sorghum bicolor</i>	36.19	0.770	0.878	0.645	23	18
水稻 <i>Oryza sativa</i>	30.54	0.771	0.984	0.685	25	23
玉米 <i>Zea mays</i>	29.05	0.807	0.982	0.695	26	22
红松 <i>Pinus koraiensis</i>	48.92	0.762	0.351	0.548	26	10
甘蓝 <i>Brassica oleracea</i>	51.79	0.698	0.641	0.550	26	12
北美乔松 <i>Pinus strobus</i>	54.38	0.750	0.592	0.535	26	8
碧桃 <i>Prunus persica</i>	53.30	0.714	0.565	0.522	29	10
草莓 <i>Fragaria × ananassa</i>	51.28	0.716	0.553	0.518	29	11
大豆 <i>Glycine max</i>	49.40	0.738	0.705	0.564	24	13
杜鹃花 <i>Rhododendron simsii</i>	51.88	0.726	0.657	0.552	27	9
柑橘 <i>Citrus sinensis</i>	58.89	0.692	0.542	0.517	26	8
菊花 <i>Chrysanthemum × morifolium</i>	57.12	0.647	0.432	0.479	31	8
猕猴桃 <i>Actinidia chinensis</i>	57.21	0.710	0.604	0.533	28	8
拟南芥 <i>Arabidopsis thaliana</i>	47.92	0.719	0.597	0.538	23	14
苹果 <i>Malus × domestica</i>	52.02	0.703	0.604	0.535	27	12
葡萄 <i>Vitis vinifera</i>	52.30	0.744	0.570	0.519	26	12
王府海棠 <i>Malus hybrid</i>	51.62	0.712	0.584	0.532	27	14
西洋梨 <i>Pyrus communis</i>	50.06	0.727	0.604	0.536	26	12
烟草 <i>Nicotiana tabacum</i>	57.52	0.647	0.420	0.473	24	11
赤松 <i>Pinus densiflora pdchsx</i>	50.79	0.751	0.632	0.541	23	10
毛果杨 <i>Populus trichocarpa</i>	54.43	0.703	0.514	0.490	26	8
马尾松 <i>Pinus massoniana</i>	51.70	0.750	0.640	0.542	24	10
辐射松 <i>Pinus radiata</i>	52.99	0.748	0.619	0.536	23	10
海岸松 <i>Pinus pinaster</i>	53.72	0.748	0.604	0.531	25	7
平均 Mean	49.47	0.729	0.632	0.551	25.80	11.96

根据密码子偏好性参数分析结果( 详见表 3 ) ,发现红松 *CHS* 基因编码区的 ENC 值为 48.92 , 远离 20 ,更接近 61 ,表明该基因的密码子使用偏好性较弱 ,蛋白质的表达量可能偏低 ;CAI 值为 0.762 ,表明其密码子偏好性一般 ;红松 *CHS* 基因编码区的 GC 含量为 0.548 ,略大于 0.5 ,而 GC<sub>3</sub>s 值是 0.351 ,小于 0.5 ,表明红松 *CHS* 基因在碱基利用时 ,更倾向于 C + G ,对以 A/T 结尾的密码子偏好性较强。

2.2 物种间 *CHS* 密码子使用偏好性比较

25 种植物的 GC<sub>3</sub>s、GC 含量、CAI 值、RSCU 值和 ENC 值如表 3 所示。25 个物种的 ENC 平均值为 50.29 ,最小的是玉米( 29.05 ) ,表达水平较高 ,最大的是柑橘( 58.89 ) ,表达水平较低。进一步分析 ENC 值发现 ,多数植物 *CHS* 基因的 ENC 值均

大于 35.0 ,表明多数植物 *CHS* 基因的密码子偏好性较弱 ,而禾本科植物 *CHS* 基因的 ENC 值均小于 35( 除高粱外 ) ,表明禾本科植物 *CHS* 基因具有更强的密码子偏好性。25 个物种 *CHS* 基因的 CAI 值介于 0.647 ~0.762 ,平均值为 0.729 ,表明密码子选择偏好性一般。对 GC 含量研究发现 ,GC 含量范围为 0.473 ~0.984 ,平均值为 0.632 ,其中 22 个物种的 GC 含量均大于 0.5 ,而 GC<sub>3</sub>s 值大于 0.5 的物种也有 22 个 ,表明不同物种 *CHS* 密码子选择偏好性存在差异 ,多数物种的 G + C 含量高于 A + T 含量 ,且密码子更偏好 C/G 结尾。通过 RSCU 值对 25 个物种 *CHS* 密码子使用偏好性分析 ,结果表明 ,菊花 RSCU >1 的密码子数目最多为 31 个 ,高粱、拟南芥、赤松、辐射松最少为 23 个 ;水稻 RSCU >1.5 的密码子最多达 23 个 ,海岸松最少有

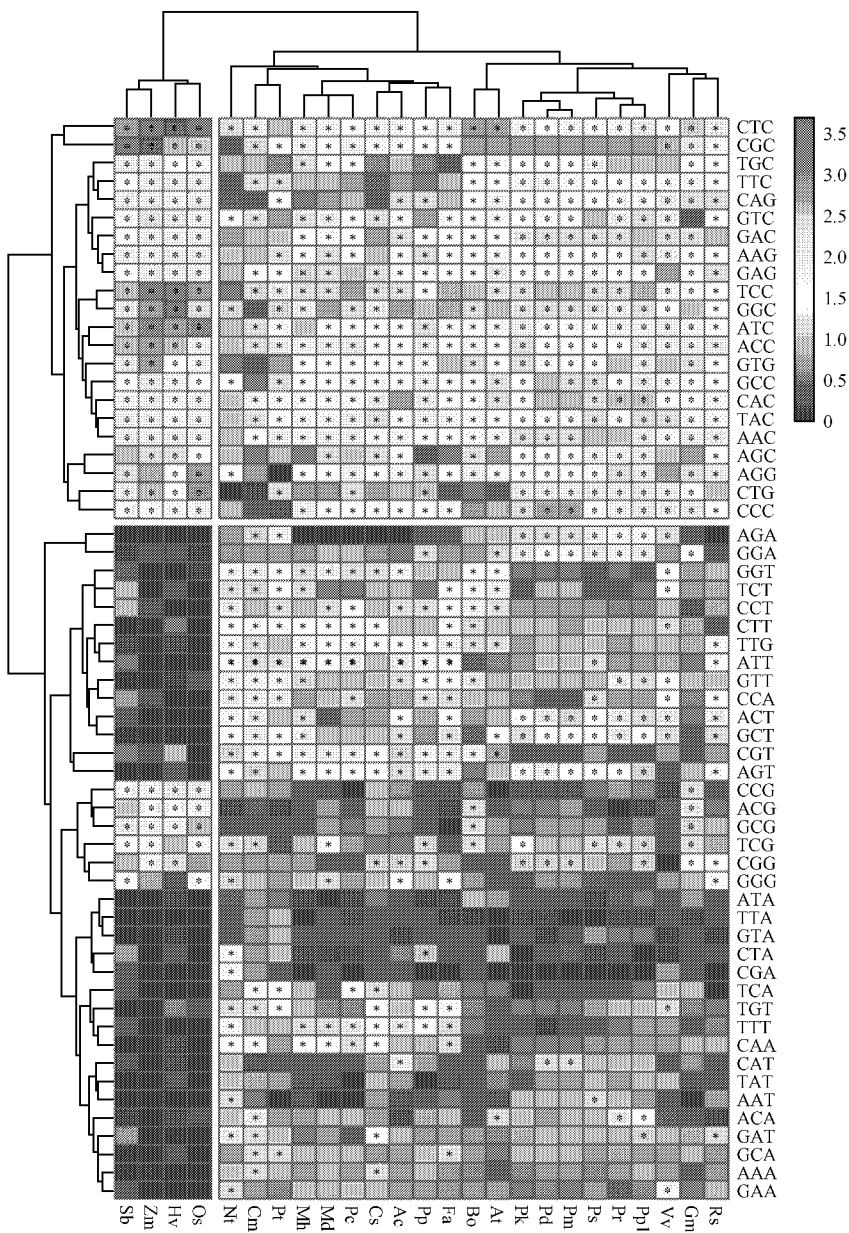


图 1 基于密码子使用频率的聚类分析 \* 表示 RSCU 值大于 1 的密码子

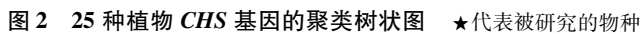
Fig. 1 Cluster analysis heatmap of codon usage frequency \* means the RSCU value is more than 1

7 个。同时,禾本科植物 RSCU > 1.5 的密码子数目要比其他植物多,表明禾本科植物密码子使用偏好性比其他植物强,与基于 ENC 的分析结果相同。除毛果杨外,其他 24 种植物共同偏好使用的密码子是 CTC,该密码子在禾本科植物 *CHS* 基因中使用偏好性明显较其他植物强(见图 1)。

**2.3 *CHS* 基因密码子使用偏好性和 CDS 序列的系统聚类分析**

对 25 种植物的 *CHS* 基因密码子使用偏好性进行聚类分析(见图 1)。结果表明,25 个物种 *CHS* 分为 2 大类,样本数分别为 4 和 21,其中禾本

科植物聚成一大类(水稻、高粱、玉米、大麦) 剩余的植物聚为一大类,说明它们之间的密码子偏好性存在明显的差异,其他分支又分为许多小的聚类中心。其中同属于松科的马尾松、赤松和红松的 *CHS* 基因聚在一个小支上,辐射松、海岸松和北美乔松聚在另一个小支上,这两个小支共聚在一个大支上面,表明具有相近亲缘关系的物种在密码子使用模式上较为相似。进一步分析发现,同属蔷薇科的王府海棠、西洋梨和苹果也聚在一个小支上,碧桃和草莓聚在一个小支上,这两个小支共聚在一个大支上面,说明基于 RSCU 值的聚类



结果与传统分类学很相近。而毛果杨、拟南芥和猕猴桃各单独成为一个小分类,说明毛果杨、拟南芥和猕猴桃的 *CHS* 基因在密码子使用上与其他植物存在差异。同时,拟南芥和红松聚在一个大支上,一定程度上说明拟南芥作为受体植物研究红松 *CHS* 基因更具有优势。此外,不是同一科的烟草、菊花和柑橘聚在一个小支上,说明它们的 *CHS* 基因在密码子使用偏好性上较为相似。因此, RSCU 值是研究植物进化关系一个很好的参数。对 25 种植物的 *CHS* 基因密码子使用频率统计发现, 61 个密码子可分为 4 大类,其中 CTC 密码子在所有物种中均出现,其在单子叶植物中的使用频率明显高于双子叶植物。值得注意的是,AGA、GGA、GGT 和 TCT 等密码子仅在部分双子叶植物中使用频率较高,而在单子叶中使用频率较低,甚至不会出现。运用 MEGA7.0 将不同物种的 CDS 序列进行聚类分析,进一步分析红松 *CHS* 基因与其他物种间的亲缘关系,结果发现(见图 2)禾本科 4 种植物的 CDS 序列同源性较高,聚在一个大的分支上,这一聚类结果与基于 RSCU 值的聚类

密码子使用频率的比值大小是评价不同物种间密码子使用偏好性的重要参数,当该值 $\leq 0.5$ 或 $\geq 2$ 时,表示两个物种间密码子使用偏好性差异较大,反之密码子使用较为一致。将红松 *CHS* 密码子使用频率与烟草、大肠杆菌、拟南芥和酵母进行比较分析(见表 4)。由表 4 可知,红松 *CHS* 与烟草、大肠杆菌、拟南芥和酵母基因组中使用偏好性差异较大的密码子分别有 31、31、28 和 34 个,大肠杆菌、酵母菌密的子使用频率差异较小,其中大肠杆菌表达系统更适用于红松 *CHS* 的异源表达,拟南芥可能为 *CHS* 基因遗传转化试验最佳的受体材料。

表 4 红松 *CHS* 基因与部分模式植物基因组密码子使用偏好性比较

Table 4 Comparison of codon usage preference between *CHS* gene in *P. koraiensis* and the other representative organisms

密码子 Condon	氨基酸 Amino acid	Pk	Sc	Ec	Ns	At	Pk/Sc	Pk/Ec	Pk/Ns	Pk/At
GCA	A	17.6	16.2	23	22.9	17.5	1.09	0.767	0.77	1.01
GCC	A	32.7	12.6	21.6	12.9	10.3	2.60	1.516	2.54	3.18
GCG	A	10.1	6.2	21.1	4.7	9	1.63	0.478	2.14	1.12
GCT	A	22.7	21.2	18.9	34.5	28.3	1.07	1.199	0.66	0.80
TGC	C	15.1	4.8	5.5	7.5	7.2	3.15	2.748	2.02	2.10
TGT	C	5.0	8.1	5.9	8.7	10.5	0.62	0.854	0.58	0.48
GAC	D	35.3	20.2	17.9	16.2	17.2	1.75	1.970	2.18	2.05
GAT	D	22.7	37.6	33.7	35.6	36.6	0.60	0.673	0.64	0.62
GAA	E	15.1	45.6	35.1	35.8	34.3	0.33	0.431	0.42	0.44
GAG	E	45.3	19.2	19.4	27.9	32.2	2.36	2.337	1.63	1.41
TTC	F	32.7	18.4	13.9	17.4	20.7	1.78	2.356	1.88	1.58
TTT	F	5.0	26.1	24.4	23.6	21.8	0.19	0.206	0.21	0.23
GGA	G	42.8	10.9	13.6	26.2	24.2	3.93	3.149	1.63	1.77
GGC	G	25.2	9.8	20.6	14.3	9.2	2.57	1.223	1.76	2.74
GGG	G	5.0	6.0	12.3	11.2	10.2	0.84	0.410	0.45	0.49
GGT	G	10.1	23.9	23.7	31.6	22.2	0.42	0.425	0.32	0.45
CAC	H	12.6	7.8	7.3	8.7	8.7	1.61	1.725	1.45	1.45
CAT	H	5.0	13.6	12.4	10.8	13.8	0.37	0.406	0.47	0.37
ATA	I	5.0	17.8	13.3	10.6	12.6	0.28	0.379	0.48	0.40
ATC	I	30.2	17.2	19.4	13.9	18.5	1.76	1.558	2.17	1.63
ATT	I	10.1	30.1	29.6	26.1	21.5	0.33	0.340	0.39	0.47
AAA	K	17.6	41.9	37.2	28.4	30.8	0.42	0.474	0.62	0.57
AAG	K	45.3	30.8	15.3	32.4	32.7	1.47	2.963	1.40	1.39
CTA	L	0.0	13.4	5.6	7.1	9.9	0.00	0.000	0.00	0.00
CTC	L	30.2	5.4	9.5	11.9	16.1	5.60	3.182	2.54	1.88
CTG	L	37.8	10.5	37.4	7.4	9.8	3.60	1.010	5.11	3.86
CTT	L	15.1	12.3	14.5	25.4	24.1	1.23	1.042	0.60	0.63
TTA	L	2.5	26.2	17.4	12.0	12.7	0.10	0.145	0.21	0.20
TTG	L	12.6	27.2	12.9	23.3	20.9	0.46	0.976	0.54	0.60
ATG	M	32.7	20.9	23.7	24.2	24.5	1.57	1.382	1.35	1.34
AAC	N	15.1	24.8	20.3	16.9	20.9	0.61	0.744	0.89	0.72
AAT	N	12.6	35.7	29.3	22.7	22.3	0.35	0.430	0.55	0.56
CCA	P	7.6	18.3	9.1	23.5	16.1	0.41	0.830	0.32	0.47
CCC	P	32.7	6.8	6.2	6.9	5.3	4.82	5.282	4.75	6.18
CCG	P	2.5	5.3	14.5	4.8	8.6	0.48	0.174	0.52	0.29
CCT	P	10.1	13.5	9.5	20.8	18.7	0.75	1.061	0.48	0.54
CAA	Q	12.6	27.3	14.4	21.5	19.4	0.46	0.875	0.59	0.65
CAG	Q	35.3	12.1	26.7	16.4	15.2	2.91	1.321	2.15	2.32
AGA	R	17.6	21.3	7.1	16.7	19.0	0.83	2.483	1.06	0.93
AGG	R	10.1	9.2	4.0	13.8	11.0	1.10	2.519	0.73	0.92
CGA	R	0.0	3.0	4.8	4.2	6.3	0.00	0.000	0.00	0.00
CGC	R	5.0	2.6	14.0	4.9	3.8	1.94	0.360	1.03	1.33
CGG	R	7.6	1.7	7.9	4.7	4.9	4.45	0.957	1.61	1.54
CGT	R	2.5	6.4	15.9	10.7	9.0	0.39	0.158	0.24	0.28
AGC	S	12.6	9.8	14.3	9.8	11.3	1.29	0.881	1.29	1.11
AGT	S	15.1	14.2	13.2	12.1	14.0	1.06	1.145	1.25	1.08
TCA	S	0.0	18.7	13.1	14.6	18.3	0.00	0.000	0.00	0.00
TCC	S	10.1	14.2	9.7	11.0	11.2	0.71	1.039	0.92	0.90
TCG	S	12.6	8.6	8.2	3.3	9.3	1.46	1.536	3.82	1.35
TCT	S	2.5	23.5	13.1	21.9	25.2	0.11	0.192	0.12	0.10
ACA	T	15.1	17.8	15.1	15.5	15.7	0.85	1.001	0.98	0.96
ACC	T	17.6	12.7	18.9	10.6	10.3	1.39	0.933	1.66	1.71
ACG	T	7.6	8.0	13.6	3.7	7.7	0.94	0.556	2.04	0.98
ACT	T	20.2	20.3	13.1	21.0	17.5	0.99	1.538	0.96	1.15
GTA	V	7.6	11.8	13.1	11.1	9.9	0.64	0.577	0.68	0.76
GTC	V	30.2	11.8	13.1	11.1	12.8	2.56	2.307	2.72	2.36
GTG	V	22.7	10.8	19.9	17.4	17.4	2.10	1.139	1.30	1.30
GTT	V	20.2	22.1	21.6	28.4	27.2	0.91	0.933	0.71	0.74
TGG	W	10.1	10.4	13.4	12.6	12.5	0.97	0.752	0.80	0.81
TAC	Y	17.6	14.8	11.7	14.9	13.7	1.19	1.507	1.18	1.29
TAT	Y	5.0	18.8	21.6	16.9	14.6	0.27	0.233	0.30	0.35
TAA	*	2.5	1.1	2.0	0.7	0.9	2.29	1.260	3.60	2.80
TAG	*	0.0	0.5	0.3	0.5	0.5	0.00	0.000	0.00	0.00
TGA	*	0.0	0.7	1.1	1.2	1.2	0.00	0.000	0.00	0.00



### 3 讨论

同义密码子由于存在简并性,往往不会影响氨基酸的编码,进而不会改变蛋白质的结构功能,因此,不会对细胞功能产生严重的影响。但是,经过长期的进化,有些密码子在基因或测序基因组中比其他密码子出现得更频繁,密码子偏好性是广泛存在且在部分物种中是非常强烈的,将特定的密码子进行的同义突变处理,可使目的基因的表达量显著提高<sup>[32]</sup>,在提高外源基因转化效率方面具有重要的作用。

密码子使用偏好性是物种在漫长的自然选择和进化进程中形成的适应性选择,对研究基因的进化和功能鉴定具有重要的意义。然而,红松中相关研究还未见报道。本研究通过对红松 *CHS* 基因密码子使用模式进行分析,确定了 GCC、TTC、GGA、ATC、CCC、AGA、AGT、CTC、CTG、TAC 共 10 个密码子是该基因的高频率密码子,其中 GGA、CCC、AGA、CTG 这 4 个密码子具有极强的偏好性,可作为该基因的最优密码子,同时也存在 CTA、CGA 和 TCA 等不参与该基因编码过程的密码子,说明红松 *CHS* 进化过程可能较为复杂。红松 *CHS* 与赤松、马尾松、海岸松、北美乔松和辐射松的 ENC 值均大于 48,可见松科植物的密码子使用偏好性相对较弱,对密码子的选择具有较高的随机性。先前的研究发现,单子叶植物基因偏向于 C/G 结尾的密码子,双子叶植物基因偏向于 A/T 结尾的密码子<sup>[33]</sup>,而在裸子植物中还没有相关的结论。本研究表明,在基因密码子碱基选择时,红松 *CHS* 基因编码区 GC<sub>3</sub>s 含量小于 0.5,更偏好于 A/T 结尾的密码子,这可能是由于红松在进化过程中主要以 GC 到 AT 突变方向为主有关,而松科的其他几个物种的 GC<sub>3</sub>s 含量均大于 0.5,AT 到 GC 的突变压力要比 GC 到 AT 的突变压力大,说明红松存在特殊的密码子使用模式和进化形式。此外 4 个 NTA 和 NCG 型密码子的 RSCU 值较低,说明红松的甲基化程度可能较高。

研究表明,基因的密码子使用特性在不同物种之间可能存在明显差异,但在亲缘关系较近的物种之间差异较小,反之差异较大<sup>[34]</sup>。基于密码子使用偏好性和基于 CDS 序列的聚类分析结果存在着一定的差异,这与前人的研究结果一致<sup>[35]</sup>。而将两种方法相结合可在一定程度上客观反应出物种的进化规律,其中基于 CDS 序列的聚类分析

更能反映出系统分类学中物种间的亲缘关系。本研究表明红松与赤松、海岸松、北美乔松、马尾松和辐射松等 6 种松科植物在两种聚类方式中均聚在一个大支上,同属蔷薇科的王府海棠、西洋梨、苹果、碧桃和草莓也聚在一个大支上,进一步说明亲缘关系较近的物种一般具有相似的密码子使用偏好性,这与在大豆(*Glycine max*)<sup>[24]</sup>、籽粒苋(*Amaranthus hypochondriacus*)<sup>[36]</sup>、小麦(*Triticum aestivum*)<sup>[37]</sup>等植物中的研究结果一致。研究中发现禾本科 4 种植物的 CDS 序列同源性较高,这一聚类结果与基于 RSCU 值的聚类分析结果相似,但在更小的分类单元上,水稻和大麦没有聚在一个小分支上,这可能是由于在物种进化过程中由于发生突变而造成 *CHS* 基因密码子使用偏好性存在差异。

在基因工程研究中,外源目的基因在受体细胞中的表达效率常与外源基因的结构密切相关,其中外源目的基因与模式物种基因组密码子使用偏好性是一个重要的指标,尤其是稀有密码子,可在一定程度上影响外源基因的转化、转录表达水平及相应功能<sup>[38]</sup>。本研究表明,拟南芥和大肠杆菌可能作为红松 *CHS* 遗传转化和异源表达较为理想的受体,这与陈义挺等<sup>[10]</sup>对猕猴桃的研究结论一致。通过遗传转化技术改良林木品质是一种强有力的方法,可对 *CHS* 基因的部分密码子进行特殊优化和定向改造,以期获得稳定高效表达的遗传转化体系。随着测序技术的不断发展,越来越多的林木完成了全基因组测序,为林木基因工程的开展和遗传改良奠定了基础。今后,根据林木 *CHS* 基因或基因家族的密码子使用偏好性特征,在全基因组水平上进行相应的遗传改造,不仅能有效促进类黄酮类物质的合成,还对林木花色、抗逆性、转基因育种、基因家族起源及基因复制的研究具有巨大的理论意义。

### 参 考 文 献

- [1] 陈哲,胡福初,王祥和,等. 菠萝密码子使用偏好性分析[J]. 果树学报 2017, 34(8): 946–955.  
Chen Z, Hu F C, Wang X H *et al.* Analysis of codon usage bias of *Ananas comosus* with genome sequencing data [J]. Journal of Fruit Science 2017, 34(8): 946–955.
- [2] Plotkin J B, Kudla G. Synonymous but not the same: the causes and consequences of codon bias[J]. Nature Reviews Genetics 2011, 12(1): 32–42.
- [3] Zhou Z P, Dang Y K, Zhou M *et al.* Codon usage biases

- co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation[ J ]. *eLife* , 2018 , 7 : 33569.
- [ 4 ] Chen S Y , Li K , Cao W Q *et al.* Codon-resolution analysis reveals a direct and context-dependent impact of individual synonymous mutations on mRNA level[ J ]. *Molecular Biology and Evolution* 2017 , 34( 11 ) : 2944 – 2958.
- [ 5 ] Ma L N , Cui P , Zhu J *et al.* Translational selection in human more pronounced in housekeeping genes[ J ]. *Biology Direct* 2014 , 9( 1 ) : 17.
- [ 6 ] Carlini D B , Stephan W. *In vivo* introduction of unpreferred synonymous codons into the drosophila Adh gene results in reduced levels of ADH protein[ J ]. *Genetics* , 2003 , 163( 1 ) : 239 – 243.
- [ 7 ] 金刚 , 王丽萍 , 龙凌云 , 等. 普通野生稻线粒体蛋白质编码基因密码子使用偏好性的分析[ J ]. *植物科学学报* 2019 , 37( 2 ) : 188 – 197.
- Jin G , Wang L P , Long L Y *et al.* Analysis of codon usage bias in the mitochondrial protein-coding genes of *Oryza rufipogon*[ J ]. *Plant Science Journal* 2019 , 37( 2 ) : 188 – 197.
- [ 8 ] Sharp P M , Matassi G. Codon usage and genome evolution [ J ]. *Current Opinion in Genetics & Development* , 1994 , 4( 6 ) : 851 – 860.
- [ 9 ] 陈义挺 , 赖瑞联 , 冯新 , 等. 猕猴桃查尔酮合成酶基因 *CHS* 密码子偏好性分析[ J ]. *福建农业学报* 2017 , 32( 11 ) : 1213 – 1217.
- Chen Y T , Lai R L , Feng X *et al.* Codon usage bias of chalcone synthase gene *CHS* in Kiwifruit[ J ]. *Fujian Journal of Agricultural Sciences* 2017 , 32( 11 ) : 1213 – 1217.
- [ 10 ] Li J , Li H Y , Zhi J K *et al.* Codon usage of expansin Genes in *Populus trichocarpa*[ J ]. *Current Bioinformatics* 2017 , 12( 5 ) : 452 – 461.
- [ 11 ] Liu Q P. Mutational bias and translational selection shaping the codon usage pattern of tissue-specific genes in rice[ J ]. *PLoS One* 2012 , 7( 10 ) : 48295.
- [ 12 ] Li N , Sun M H , Jiang Z S *et al.* Genome-wide analysis of the synonymous codon usage patterns in apple[ J ]. *Journal of Integrative Agriculture* 2016 , 15( 5 ) : 983 – 991.
- [ 13 ] 赵光浩 , 冷伟伟 , 张腾 , 等. 51 个红松子代家系苗期生长变异研究[ J ]. *植物研究* 2018 , 38( 4 ) : 590 – 596.
- Zhao G H , Leng W W , Zhang T *et al.* Variance analysis of growth traits of 51 *Pinus koraiensis* seedling families [ J ]. *Bulletin of Botanical Research* 2018 , 38( 4 ) : 590 – 596.
- [ 14 ] 姜国云 , 蒋路平 , 宋双林 , 等. 红松半同胞家系遗传变异分析及果材兼用优良家系选择[ J ]. *植物研究* , 2018 , 38( 5 ) : 775 – 784.
- Jiang G Y , Jiang L P , Song S L *et al.* Genetic variance analysis and excellent fruit-timber families selection of half-sib *Pinus koraiensis*[ J ]. *Bulletin of Botanical Research* 2018 , 38( 5 ) : 775 – 784.
- [ 15 ] 李波 , 包怡红 , 王振宇. 松多酚生物活性研究进展 [ J ]. *黑龙江农业科学* 2013 ( 10 ) : 133 – 136.
- Li B , Bao Y H , Wang Z Y. Research progress on biological activity of Pine Polyphenol[ J ]. *Heilongjiang Agricultural Sciences* 2013 ( 10 ) : 133 – 136.
- [ 16 ] Hichri I , Barrieu F , Bogs J *et al.* Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway[ J ]. *Journal of Experimental Botany* , 2011 , 62( 8 ) : 2465 – 2483.
- [ 17 ] 赵景梅 , 黄东益 , 张青 , 等. 紫参薯查尔酮合成酶及异构酶基因的克隆与表达分析[ J ]. *热带作物学报* , 2018 , 39( 5 ) : 920 – 925.
- Zhao J M , Huang D Y , Zhang Q *et al.* Cloning and expression analysis of *daCHS* and *daCHI* gene in *Dioscorea alata*[ J ]. *Chinese Journal of Tropical Crops* , 2018 , 39( 5 ) : 920 – 925.
- [ 18 ] Reimold U , Kröger M , Kreuzaler F *et al.* Coding and 3' non-coding nucleotide sequence of chalcone synthase mRNA and assignment of amino acid sequence of the enzyme[ J ]. *The EMBO Journal* , 1983 , 2( 10 ) : 1801 – 1805.
- [ 19 ] 徐靖 , 朱家红 , 王效宁 , 等. 甘薯查尔酮合成酶基因 *IbCHS1* 的克隆和表达分析[ J ]. *分子植物育种* 2018 , 16( 6 ) : 1752 – 1757.
- Xu J , Zhu J H , Wang X N *et al.* Cloning and expression analysis of chalcone synthase gene *IbCHS1* in *Ipomoea batatas*[ J ]. *Molecular Plant Breeding* , 2018 , 16( 6 ) : 1752 – 1757.
- [ 20 ] 徐秀荣. 桂花 *CHS*、*DFR* 基因全长克隆、表达分析及表达载体构建[ D ]. 泰安 : 山东农业大学 2016.
- Xu X R. Cloning and expression analysis of *CHS* and *DFR* gene from *Osmanthus fragrans* and construction of expression vector[ D ]. Tai 'an Shandong Agricultural University 2016.
- [ 21 ] Fliegmann J , Schröder G , Schanz S *et al.* Molecular analysis of chalcone and dihydropinosylvin synthase from Scots pine( *Pinus sylvestris* ) , and differential regulation of these and related enzyme activities in stressed plants [ J ]. *Plant Molecular Biology* , 1992 , 18( 3 ) : 489 – 503.
- [ 22 ] Saslowsky D E , Dana C D , Winkel-Shirley B. An allelic series for the chalcone synthase locus in *Arabidopsis* [ J ]. *Gene* 2000 , 255( 2 ) : 127 – 138.
- [ 23 ] 马盛超 , 余鑫梅 , 姚慧鹏. 蓼科大黄属植物 *CHS* 基因密码子偏爱性分析[ J ]. *分子植物育种* 2017 , 15( 6 ) :

- 2171–2178.
- Ma S C, Yu X M, Yao H P. The codon usage analysis in *CHS* gene within the *Polygonaceae rheum* [ J ]. Molecular Plant Breeding 2017, 15( 6 ) : 2171–2178.
- [ 24 ] 张家琪, 徐悦, 李晓冰, 等. 豆科植物查尔酮合成酶基因密码子偏好性分析 [ J ]. 华北农学报, 2016, 31( 6 ) : 12–20.
- Zhang J Q, Xu Y, Li X B, et al. Analysis of codon bias of *CHS* gene in leguminous plant [ J ]. Acta Agriculturae Boreali-Sinica 2016, 31( 6 ) : 12–20.
- [ 25 ] 杨益森. 红松查尔酮合成酶基因( *PkCHS* )克隆与外源因素对其表达的影响 [ D ]. 哈尔滨: 哈尔滨工业大学, 2013.
- Yang Y M. Cloning and effect of exogenous factors on the expression of *PkCHS* of *Pinus koraiensis* [ D ]. Harbin: Harbin Institute of Technology, 2013.
- [ 26 ] 邓浩. 外源因素对红松查尔酮合成酶基因表达和多酚含量的影响 [ D ]. 哈尔滨: 哈尔滨工业大学, 2014.
- Deng H. Effect of exogenous factors on chalcone synthase gene expression and polyphenol content in Korean pine [ D ]. Harbin: Harbin Institute of Technology, 2014.
- [ 27 ] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets [ J ]. Molecular Biology and Evolution 2016, 33( 7 ) : 1870–1874.
- [ 28 ] Ihaka R, Gentleman R. R: a language for data analysis and graphics [ J ]. Journal of Computational and Graphical Statistics 1996, 5: 299–314.
- [ 29 ] George D, Mallery P. IBM SPSS statistics 23 Step by step: a simple guide and reference [ M ]. New York: Routledge Press, 2016.
- [ 30 ] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [ J ]. Journal of Molecular Evolution 1986, 24( 1–2 ) : 28–38.
- [ 31 ] Sharp P M, Li W H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications [ J ]. Nucleic Acids Research, 1987, 15( 3 ) : 1281–1295.
- [ 32 ] Wang L Y, Xing H X, Yuan Y C, et al. Genome-wide analysis of codon usage bias in four sequenced cotton species [ J ]. PLoS One 2018, 13( 3 ) : e0194372.
- [ 33 ] 薛拥志, 王宝地, 王连荣, 杏扁 *PaCBF1* 基因密码子偏好性分析 [ J ]. 分子植物育种, 2018, 16( 15 ) : 4890–4898.
- Xue Y Z, Wang B D, Wang L R. Analysis of codon bias of *PaCBF1* gene in *Prunus armeniaca* L. [ J ]. Molecular Plant Breeding 2018, 16( 15 ) : 4890–4898.
- [ 34 ] 李蓉, 谢析颖, 王雪晶, 等. 兰科植物 *FNR* 基因的密码子偏好性分析 [ J ]. 热带作物学报, 2018, 39( 6 ) : 1137–1145.
- Li R, Xie X Y, Wang X J, et al. Codon usage bias of ferredoxin-NADP<sup>+</sup> oxidoreductase( *FNR* ) in Orchidaceae [ J ]. Chinese Journal of Tropical Crops 2018, 39( 6 ) : 1137–1145.
- [ 35 ] 赖瑞联, 林玉玲, 钟春水, 等. 龙眼生长素受体基因 *TIR1* 密码子偏好性分析 [ J ]. 园艺学报, 2016, 43( 4 ) : 771–780.
- Lai R L, Lin Y L, Zhong C S, et al. Analysis of codon bias of auxin receptor gene *TIR1* in *Dimocarpus longan* [ J ]. Acta Horticulturae Sinica, 2016, 43( 4 ) : 771–780.
- [ 36 ] 冯瑞云, 闫建俊, 侯雅静, 等. 籽粒苋 *PEPC* 基因密码子使用偏性特征分析 [ J ]. 分子植物育种, 2018, 16( 20 ) : 6646–6656.
- Feng R Y, Yan J J, Hou Y J, et al. Analysis of codon usage bias characteristics of *PEPC* gene in *Amaranthus hypochondriacus* [ J ]. Molecular Plant Breeding, 2018, 16( 20 ) : 6646–6656.
- [ 37 ] 焦义然, 陈文烨, 杨帆, 等. 小麦 *FT* 基因编码蛋白结构及功能的生物信息学分析 [ J ]. 江西农业学报, 2018, 30( 10 ) : 1–6.
- Jiao Y R, Chen W Y, Yang F, et al. Bioinformatics analysis of structure and function of *FT*-coded protein in Wheat( *Triticum aestivum* ) [ J ]. Acta Agriculturae Jiangxi 2018, 30( 10 ) : 1–6.
- [ 38 ] Li X Z, Song H, Kuang Y, et al. Genome-wide analysis of codon usage bias in *Epichloë festucae* [ J ]. International Journal of Molecular Sciences 2016, 17( 7 ) : 1138.