

DOI: 10.11931/guihaia.gxzw201708001

引文格式: 王鹏良, 杨利平, 吴红英, 等. 普通油茶叶绿体基因组密码子偏好性分析 [J]. 广西植物, 2018, 38(2):135-144

WANG PL, YANG LP, WU HY, et al. Condon preference of chloroplast genome in *Camellia oleifera* [J]. *Guihaia*, 2018, 38(2):135-144

## 普通油茶叶叶绿体基因组密码子偏好性分析

王鹏良<sup>1,3\*</sup>, 杨利平<sup>2</sup>, 吴红英<sup>2</sup>, 农有良<sup>1</sup>, 吴双成<sup>1</sup>,  
肖玉菲<sup>3</sup>, 覃子海<sup>3</sup>, 王华宇<sup>1,2</sup>, 刘海龙<sup>3</sup>

(1. 广西北部湾海洋生物多样性养护重点实验室, 钦州学院, 广西 钦州 535011; 2. 钦州市植物生物技术重点实验室, 广西钦州市林业科学研究所, 广西 钦州 535099; 3. 广西特色经济林培育与利用重点实验室, 广西壮族自治区林业科学研究院, 南宁 530002)

**摘要:** 为了利用叶绿体基因工程技术改良普通油茶的重要经济性状, 该研究以普通油茶叶叶绿体全基因组序列为材料, 从中筛选出 51 条长度大于 300 bp 且以 ATG 起始的非重复 CDS (Coding DNA Sequence) 为对象, 利用 CodonW 软件分析其密码子偏好性。结果表明: 密码子第三位 GC 含量为 27.55%, ENC 范围在 35.23~56.67 之间, 平均值为 46.09; RSCU 值大于 1.00 的密码子数目为 30 个, 其中 29 个第三位碱基以 U 或 A 结尾; 中性绘图表明 GC<sub>12</sub> 与 GC<sub>3</sub> 的相关系数为 0.143, 相关性不显著, 回归系数为 0.0573; 频数分布显示, 55% 基因的 ENC 比值集中分布在 0~0.1, 25% 基因的 ENC 比值分布在 0.1~0.2 之间; 对应分析结果表明, 第一向量轴占 10.12% 的差异, 第二向量轴占 9.36% 的差异, 其余两轴分别占 7.97% 和 7.46%, 前 4 轴累计差异为 34.91%。中性绘图、ENC-plot 和对应性分析均表明普通油茶叶叶绿体基因密码子偏好受突变作用, 更多受选择的影响。最终取高表达优越密码子和高频密码子共有的 CUU、AUU、GUU、GUA、UAA、CAA、AAA、GAC、GAA、CCU、ACU、GCU、GCA、UGU、CGU、AGU、UUG、GGU 等 18 个密码子作为最优密码子。该研究结果为利用叶绿体基因工程技术改良普通油茶重要经济性状奠定了基础。

**关键词:** 普通油茶, 叶绿体, 密码子偏好, ENC, RSCU

**中图分类号:** Q943.2 **文献标识码:** A **文章编号:** 1000-3142(2018)02-0135-10

## Condon preference of chloroplast genome in *Camellia oleifera*

WANG Pengliang<sup>1,3\*</sup>, YANG Liping<sup>2</sup>, WU Hongying<sup>2</sup>, NONG Youliang<sup>1</sup>, WU Shuangcheng<sup>1</sup>,  
XIAO Yufei<sup>3</sup>, QIN Zihai<sup>3</sup>, WANG Huayu<sup>1,2</sup>, LIU Hailong<sup>3</sup>

(1. Guangxi Key Laboratory of Beibu Gulf Marine Biodiversity Conservation, Qinzhou University, Qinzhou 535011, Guangxi, China; 2. Qinzhou Key Laboratory of Plant Biotechnology, Qinzhou Forestry Science Institute, Qinzhou 535099, Guangxi, China; 3. Guangxi Key Laboratory of Special Non-Wood Forest Cultivation and Utilization, Guangxi Zhuang Autonomous Region Forestry Science Institute, Nanning 530002, China)

**Abstract:** With the rapid development of the industry and increase of planting area of *Camellia oleifera*, the harvest was still affected by the factors such as the tolerance of disease, insect and drought as well as mating compatibility. Genetic

收稿日期: 2017-08-22

基金项目: 国家自然科学基金 (31460208); 钦州市科学研究与技术开发项目 (201616804) [Supported by the National Natural Science Foundation of China (31460208); the Program of Science and Technology of Qinzhou (201616804)].

作者简介: 王鹏良 (1978-), 男, 浙江新昌人, 博士, 高级工程师, 主要从事遗传育种研究, (E-mail) pengliang\_wang@163.com。

\* 通信作者

engineering of chloroplast was an effective way to improve the important traits of plant. In order to improve the important traits of *C. oleifera*, it was necessary to carry out the analysis on codon preference of its chloroplast genome. The whole genome sequence of chloroplast in *C. oleifera* taken as raw material, 51 CDS who were longer than 300 bps, started with ATG and not repeated were screened for future analysis. The codon preference was conducted by the CodonW software. The results indicated that the GC content of the third base of genetic code was 27.55% and effective number of codons ranged from 35.23 to 56.67 with an average of 46.09 and then 29 codons of 30 whose RSCU were more than 1 ended with A or U. Neutral plot analysis showed the correlation between  $GC_{12}$  and  $GC_3$  was not significant and the coefficient of correlation and regression was 0.143 and 0.057 3, respectively. Frequency analysis showed 55% and 25% of the ENC ratio of genes focused on the region of 0–0.1 and 0.1–0.2. Correspondence analysis indicated the first axis accounted for 10.12 variation and the rested three axes accounted for 9.36%, 7.97% and 7.46% and the first four axes accounted for 34.91% in total. All the methods used in this study showed that codon preference was affected by not only mutation and but also selection. Finally the 18 optimal codons including CUU, AUU, GUU, GUA, UAA, CAA, AAA, GAC, GAA, CCU, ACU, GCU, GCA, UGU, CGU, AGU, UUG, GGU were determined. This work provides plenty of information for improving important traits of *C. oleifera* utilizing the genetic engineering of chloroplast.

**Key words:** *Camellia oleifera*, chloroplast, codon preference, effective number of codon, relative synonymous codon usage

普通油茶 (*Camellia oleifera*) 是山茶科 (Theaceae) 山茶属 (*Camellia*) 一类植物, 是我国南方最重要的木本食用植物油物种之一 (张恩慧等, 2016)。目前, 油茶最具价值的部分为油茶果。茶籽油富含多种对人体健康有益的成分, 对于提高人体的健康水平有很大影响, 尤其是在人体免疫力和预防高血压等方面, 因此是一种非常优质的食用油, 被誉为“东方的橄榄油”; 同时茶籽油还含有多种有益于皮肤的成分, 易于吸收, 无刺激性, 所以常被用于调制药膏和高级化妆品的用油 (庄瑞林, 2007); 其果壳由木质素、半纤维素和纤维素组成, 可以用来制作糠醛、木糖醇、培养基、活性炭等; 茶籽粕可以直接用于制作肥料及提取茶皂素、鞣质 (单宁)、黄酮、生物碱等生物活性物质。

当前由于我国人口规模增长和居民收入水平提高, 我国食用植物油消费总量快速增长, 尽管近年来植物油生产稳步增长, 然而消费增速高于生产增速, 缺口仍然较大 (张雯丽, 2016)。国家主动调整油料物种产业结构, 由于油茶油质好、生长快、寿命长、适应性强、不占用耕地等优点, 油茶产业得到重视, 油茶产业获得快速发展, 新品种数量增加, 推广面积持续扩大。到目前为止, 全国油茶种植面积约为 426.67 万  $hm^2$  (罗晨, 2016)。

油茶产量持续增长, 但油茶生产中还会遇到

以下问题: 首先, 近年来大面积油茶病害的发生, 如炭疽病, 软腐病和叶肿病 (徐丽萍和檀根甲, 2015); 其次, 大规模虫害的出现, 如蚧壳虫, 象鼻虫等; 第三, 由于油茶的抗旱性弱导致低龄油茶植株死亡率高, 引起油茶落花、落果, 降低产量 (左继林等, 2012; 陈永忠等, 2014); 最后, 油茶的交配亲和性也严重影响了油茶的产量 (常维霞和姚小华, 2016; 常维霞等, 2016)。

基因工程技术是改良植物重要性状的有效手段 (王关林和方宏筠, 2014)。叶绿体基因组是理想的转化载体, 高表达优势明显, 同时能将转化基因的扩散控制在最低程度 (Daniell & Chase, 2004)。叶绿体基因组中存在密码子偏好, 而且不同物种间密码子偏好差异显著 (Morton, 1999; Liu & Xue, 2005)。为了更好地设计叶绿体高效表达载体, 进一步利用叶绿体基因工程技术改良油茶重要性状, 有效降低油茶产量损失, 分析普通油茶叶绿体基因组密码子偏好十分必要。

## 1 材料与方法

### 1.1 序列材料

根据已报道的叶绿体基因组的 GenBank 登录号 NC\_023084.1 (Shi et al, 2013), 登陆 NCBI 数据

库(<https://www.ncbi.nlm.nih.gov/>),搜索并下载完整普通油茶叶绿体基因组序列及其 CDS(Coding DNA Sequence),序列总长度为156 971 bp,共包含 87 条编码蛋白的 CDS。由于短序列无法正确计算有效密码子数(Wright, 1990),为了减少样本误差,本研究选择以 ATG 为起始密码子的非重复且长度大于 300 bp 的序列进行分析。

## 1.2 方法

**1.2.1 密码子偏好参数计算** 以选取的 CDS 为研究对象,使用 CodonW 1.4.2 软件和在线程序 CUSP(<http://imed.med.ucm.es/EMBOSS/>)分析密码子使用偏好参数,即有效密码子数(Effective Number of Codon, ENC)、相对同义密码子使用度(Relative Synonymous Codon Usage, RSCU)、GC、GC<sub>1</sub>、GC<sub>2</sub>、GC<sub>3</sub>、GC<sub>3s</sub>。其中,GC、GC<sub>1</sub>、GC<sub>2</sub>、GC<sub>3</sub>分别代表了基因 CDS 序列的 GC 总含量和密码子的对应位置上的 GC 含量,GC<sub>3s</sub>为同义密码子第三位的 GC 含量。

**1.2.2 中性绘图分析** 中性绘图是一种可初步判断影响密码子使用偏好因素的方法。首先计算密码子 GC<sub>1</sub>和 GC<sub>2</sub>的平均值 GC<sub>12</sub>,再以计算出的平均值为纵坐标,以 GC<sub>3</sub>为横坐标作散点图,图中每一个点就代表一个基因的位置。由中性绘图可分析密码子的前两位碱基和第三位碱基的关系,从而判断密码子使用偏好由突变或选择造成。

**1.2.3 ENC-plot 绘图** 以 GC<sub>3</sub>为横坐标,ENC 为纵坐标,构建二维散点图,并在图中构建出 ENC 值的标准曲线。每个基因点分布在标准曲线附近,这代表密码子偏好性只是受到了突变作用的影响,如果各基因点集中在预期曲线下方,表明密码子偏好更多受选择影响。

标准曲线的公式如下(Wright, 1990):

$$ENC = 2 + GC_3 + \frac{29}{GC_3^2 + (1 - GC_3)^2}$$

**1.2.4 对应性分析** 对应性分析是一种利用多元统计探究不同基因密码子变异趋势的方法。利用 CodonW 软件的 RSCU 对应分析功能,将所有基因分布到 59 维向量空间,从而反映出密码子使用偏好特点。通过分析基因位置探究基因向量及基因间的值的变异程度,推测导致密码子偏好发生的原因。

**1.2.5 最优密码子确定** 以 ENC 偏好性为标准排

序,两端各选出 10% 基因,分别建立高低偏性库,取两库中  $\Delta RSCU > 0.08$  的密码子作为高表达优越密码子(杨惠娟等, 2012; 杨国锋等, 2015)。将高表达优越密码子与高频密码子相比较,确定出最优密码子(刘庆坡和薛庆中, 2004; 续晨等, 2010; 罗洪等, 2015; 胡莎莎等, 2016)。

## 2 结果与分析

### 2.1 密码子组成分析

根据密码子使用偏好分析要求,本文筛选到 51 条 CDS,再利用 CodonW 软件和 CUSP 程序对筛选出的 CDS 进行分析。表 1 结果表明,密码子第一位的 GC 平均含量为 46.86%、第二位为 39.28%、第三位为 27.55%,第三位的 GC 含量明显小于前两位的 GC 含量,而且相差较大,说明普通油茶叶绿体基因密码子偏好于 A 和 U 结尾。ENC 范围在 35.23~56.67 之间,平均值为 46.09。ENC 值小于 35 说明密码子偏性强,高于 35 则说明偏性比较弱的标准(Jiang et al, 2008),说明普通油茶叶绿体基因密码子偏性较弱。

相关分析结果(表 2)显示,GC 和 GC<sub>1</sub>、GC<sub>2</sub>、GC<sub>3</sub>的相关性均达到极显著水平,其相关系数分别为 0.829、0.766、0.379。GC<sub>1</sub>和 GC<sub>2</sub>的相关性也达到了极显著水平,其相关系数为 0.445,但 GC<sub>3</sub>与 GC<sub>1</sub>、GC<sub>2</sub>的相关性均未达到显著相关水平,说明密码子前两位的 GC 含量比较相似,第三位 GC 含量和前两位差异较大。ENC 与 GC<sub>3</sub>呈显著相关,其相关系数为 0.350,说明第三位碱基的组成会对密码子的使用偏性有影响。密码子数目 N 与 GC<sub>3</sub>、ENC 显著相关,表明基因序列长度对 GC<sub>3</sub>和密码子的偏好性有一定的影响。

RSCU 分析结果(表 3)表明,RSCU 值大于 1.00 的密码子数目为 30 个。其中,16 个第三位碱基为 U;13 个以 A 结尾;1 个以 G 结尾。这说明普通油茶叶绿体基因组密码子偏爱以 A 或 U(T)结尾,不偏好以 G 或 C 结尾的密码子。

### 2.2 中性绘图分析

普通油茶叶绿体基因中性绘图分析(图 1)显示,GC<sub>12</sub>的取值范围为 0.319 6~0.550 4,GC<sub>3</sub>的范

表 1 普通油茶叶绿体基因组不同位置的 GC 含量

Table 1 GC content of different positions of codon in chloroplast of *Camellia oleifera*

基因 Gene	GC <sub>1</sub> (%)	GC <sub>2</sub> (%)	GC <sub>3</sub> (%)	GC (%)	ENC	基因 Gene	GC <sub>1</sub> (%)	GC <sub>2</sub> (%)	GC <sub>3</sub> (%)	GC (%)	ENC
<i>psbA</i>	49.72	43.50	32.49	40.90	40.60	<i>matK</i>	39.00	32.00	27.60	32.87	46.95
<i>atpA</i>	55.51	40.19	23.82	39.83	42.89	<i>atpF</i>	45.95	34.05	35.68	38.56	43.86
<i>atpI</i>	49.19	37.90	26.61	37.90	44.59	<i>rps2</i>	43.46	42.19	27.85	37.83	47.62
<i>rpoC2</i>	45.75	37.82	28.44	37.33	49.24	<i>rpoC1</i>	50.00	37.72	28.36	38.69	50.15
<i>rpoB</i>	50.14	38.00	27.82	38.66	48.33	<i>psbD</i>	51.69	43.22	31.36	42.09	43.19
<i>psbC</i>	53.16	46.41	31.43	43.67	43.75	<i>rps14</i>	43.56	47.52	31.68	40.92	37.46
<i>psaB</i>	48.84	42.99	30.75	40.86	47.80	<i>psaA</i>	52.20	43.54	31.69	42.48	49.13
<i>yef3</i>	47.93	38.46	28.99	38.46	56.67	<i>rps4</i>	50.00	37.13	25.74	37.62	47.88
<i>ndhJ</i>	50.31	37.74	31.45	39.83	51.41	<i>ndhK</i>	41.34	42.05	25.44	36.28	46.88
<i>ndhC</i>	46.28	33.88	24.79	34.99	45.99	<i>atpE</i>	50.75	38.06	27.61	38.81	47.78
<i>atpB</i>	56.71	41.48	28.06	42.08	44.79	<i>rbcL</i>	58.61	43.70	30.04	44.12	48.16
<i>accD</i>	40.44	36.02	29.38	35.28	47.84	<i>yef4</i>	43.78	41.08	29.19	38.02	46.79
<i>cemA</i>	38.36	26.72	31.47	32.18	49.65	<i>petA</i>	52.34	37.07	28.04	39.15	48.56
<i>rps18</i>	35.29	43.14	26.47	34.97	35.64	<i>rpl20</i>	43.97	41.38	27.59	37.64	47.60
<i>clpP</i>	58.67	37.76	25.51	40.65	49.00	<i>psbB</i>	54.81	46.17	30.84	43.94	48.32
<i>petB</i>	48.61	41.67	30.56	40.28	42.69	<i>petD</i>	50.93	39.13	26.09	38.72	43.64
<i>rpoA</i>	44.64	32.14	25.00	33.93	48.80	<i>rps11</i>	52.52	57.55	20.86	43.65	47.80
<i>rps8</i>	42.65	41.18	27.21	37.01	41.79	<i>rpl14</i>	56.10	36.59	26.02	39.57	44.21
<i>rpl16</i>	51.47	52.21	19.12	40.93	35.23	<i>rps3</i>	47.03	31.51	22.83	33.79	47.33
<i>rpl22</i>	41.03	37.18	25.00	34.40	43.00	<i>rpl2</i>	50.18	47.64	32.36	43.39	54.12
<i>yef2</i>	41.67	34.36	37.10	37.71	53.25	<i>ndhB</i>	41.29	38.55	31.90	37.25	47.36
<i>rps7</i>	51.92	45.51	23.08	40.17	45.81	<i>ndhF</i>	36.72	35.78	22.70	31.73	41.72
<i>ccsA</i>	33.54	36.96	24.53	31.68	47.27	<i>ndhE</i>	39.22	32.35	24.51	32.03	40.93
<i>ndhG</i>	42.37	32.77	22.03	32.39	42.88	<i>ndhI</i>	42.26	37.50	27.98	35.91	49.81
<i>ndhA</i>	42.03	39.01	20.60	33.88	41.23	<i>ndhH</i>	50.76	36.04	24.37	37.06	46.64
<i>yef1</i>	35.07	28.84	24.84	29.58	46.37						
平均值 Average	46.86	39.28	27.55	37.88	46.09						

围为0.191 2~0.371 0。GC<sub>12</sub>与 GC<sub>3</sub>的相关系数为0.143,相关性未达到显著水平,回归系数为0.057 3,说明 GC<sub>12</sub>和 GC<sub>3</sub>的相关性微弱。突变对密码子第一位、第二位和对第三位碱基组成的影响不同。各基因都落在对角线的上方,并且没有对角线上或沿着对角线分布。突变不是密码子偏

性形成的主要原因。因此,普通油茶叶绿体基因密码子的使用更多地受选择的影响。

### 2.3 ENC-plot 绘图

以 GC<sub>3</sub>为 X 轴,ENC 为 Y 轴作图(图 2),图 2 中各点为基因分布情况。代表基因的各点基本都落在了标准曲线的下方。为了能更直接地显示出

表 2 各基因相关参数的相关性分析

Table 2 Correlation analysis of each gene's related parameters

项目 Item	GC <sub>1</sub>	GC <sub>2</sub>	GC <sub>3</sub>	GC	ENC
GC <sub>2</sub>	0.445 **				
GC <sub>3</sub>	0.078	-0.010			
GC	0.829 **	0.766 **	0.379 **		
ENC	0.147	-0.217	0.350 *	0.100	
N	-0.160	-0.235	0.281 *	-0.116	0.291 *

注: \*\* 表示在 0.01 水平上显著相关; \* 表示在 0.05 水平上显著相关。

Note: \*\* means significant correlation ( $P < 0.01$ ); \* means significant correlation ( $P < 0.05$ ).

差异,根据  $\frac{ENC_{exp}-ENC_{obs}}{ENC_{exp}}$  公式,计算 ENC 比值,再

统计 ENC 比值的分布(表 4)。频数分布显示,55%的基因的 ENC 比值集中分布在 0~0.1,25%的基因 ENC 比值分布在 0.1~0.2 的区间,结果表明  $ENC_{exp}$  与  $ENC_{obs}$  有差异。假如  $ENC_{exp}$  与  $ENC_{obs}$  接近时,表明基因的密码子受突变影响;本文中  $ENC_{exp}$  与  $ENC_{obs}$  有差异较大,所以普通油茶叶绿体基因组密码子偏好更多受选择的影响(胡莎莎等,2016)。

## 2.4 对应性分析

RSCU 的对应分析结果表明,第一向量轴占 10.12% 的差异,第二向量轴占 9.36% 的差异,其余两轴分别为 7.97% 和 7.46%,前 4 轴累计差异为 34.91%,说明第一轴对普通油茶叶绿体基因组密码子偏好性造成的影响最大。第一轴与 GC<sub>3</sub> 的相关性为 0.215 和 ENC 的相关性为 0.104,均相关不显著。这说明 GC<sub>3</sub> 是对普通油茶叶绿体基因组密码子偏好性造成影响的不大,密码子的偏好性受到其他多方面因素的影响。以 Axis 1 为横坐标,Axis 2 为纵坐标,从各基因在第一、第二轴的分布(图 3)看出,与遗传结构相关的基因分布相对比较集中,说明两个类基因的密码子使用是比较接近的;而其他类别基因可能是因为密码子使用模式不同,分布比较分散。

## 2.5 最优密码子分析

以密码子的有效数量 ENC 为标准对所有基因的 CDS 序列进行排序,先从排序结果的两端各选取 5 个基因分别建立高表达基因库和低表达基因库;然后使用 CodonW 计算两个表达库中的 RSCU 值,并计算高低库  $\Delta RSCU$  值(表 5);最后确定 22 个密码子(表中带星号的)为普通油茶叶绿体基因组高表达优越密码子,10 个以 U 结尾、6 个以 A 结尾、5 个以 C 结尾和 1 个以 G 结尾。

先将 22 个高表达优越密码子和表 3 得出的 30 个高频密码子相比较,然后选取共有的密码子作为最优密码子,最后确定了 18 个密码子为普通油茶叶绿体基因组最优密码子,分别为 CUU、AUU、GUU、GUA、UAA、CAA、AAA、GAC、GAA、CCU、ACU、GCU、GCA、UGU、CGU、AGU、UUG、GGU,其中有 2 个密码子以 G 或 C 结尾,其余的 16 个密码子均以 A 或 U 结尾。

## 3 讨论

在 20 种氨基酸中,除了甲硫氨酸和色氨酸只有单个密码子外,其他 18 种氨基酸均有多种密码子,即遗传密码子存在简并性。密码子简并性常常表现在第三位密码子上(朱圣庚和徐长发,2016),因此在普通油茶叶绿体基因组碱基组成中发现,GC<sub>1</sub>、GC<sub>2</sub> 显著相关,GC<sub>3</sub> 与 GC<sub>1</sub>、GC<sub>2</sub> 相关不显著,且 GC<sub>3</sub> 小于 GC<sub>1</sub> 和 GC<sub>2</sub>,说明普通油茶叶绿体基因组的碱基组成偏向于 A 和 T,第 3 位密码子通常以 A、T 结尾。这与多数植物,如文心兰(李冬梅等,2012)、甜芥(罗洪等,2015)、苦芥(胡莎莎等,2016)、蝴蝶兰(续晨等,2010)、蕨藜苜蓿(杨国锋等,2015)等物种的特点相一致。

突变和选择是密码子偏好形成的主要原因。ENC-plot 分析结果表明,部分基因落点位于标准曲线附近,基因的实际 ENC 值与理论 ENC 值基本一致,说明这些密码子偏好受到 GC 含量(突变)的影响较大;而大多数基因的落点离标准曲线较远,实际 ENC 值与理论 ENC 值的差异较大,受选择的影响较大。因此,普通油茶叶绿体密码子偏好受到突变的影响,更多的是受到了选择的作用。

表 3 普通油茶叶绿体各氨基酸相对同义密码子使用度  
Table 3 RSCU analysis on protein coding region in *Camellia oleifera*

氨基酸 Amino acid	密码子 Codon	数目 Number	RSCU	氨基酸 Amino acid	密码子 Codon	数目 Number	RSCU
Phe	UUU	798	1.32	Ser	UCU	486	1.81
	UUC	408	0.68		UCC	241	0.90
Leu	UUA	737	2.01	UCA	316	1.18	
	UUG	444	1.21	UCG	136	0.51	
	CUU	453	1.24	Pro	CCU	365	1.65
	CUC	143	0.39		CCC	156	0.70
Ile	CUA	285	0.78	CCA	259	1.17	
	CUG	134	0.37	CCG	106	0.48	
	AUU	899	1.46	Thr	ACU	436	1.65
	AUC	356	0.58		ACC	193	0.73
Met	AUA	595	0.96	ACA	324	1.23	
	AUG	514	1.00	ACG	103	0.39	
Val	GUU	433	1.50	Ala	GCU	531	1.82
	GUC	118	0.41		GCC	181	0.62
	GUA	445	1.55		GCA	340	1.17
	GUG	155	0.54		GCG	113	0.39
Tyr	UAU	653	1.61	Cys	UGU	173	1.52
	UAC	158	0.39		UGC	55	0.48
TER	UAA	27	1.56	TER	UGA	14	0.81
	UAG	11	0.63	Trp	UGG	393	1.00
His	CAU	411	1.57	Arg	CGU	291	1.40
	CAC	113	0.43		CGC	71	0.34
Gln	CAA	584	1.54	CGA	299	1.44	
	CAG	175	0.46	CGG	83	0.40	
Asn	AAU	811	1.57	Ser	AGU	345	1.28
	AAC	224	0.43		AGC	87	0.32
Lys	AAA	863	1.53	Arg	AGA	380	1.82
	AAG	264	0.47		AGG	126	0.60
Asp	GAU	723	1.63	Gly	GGU	478	1.30
	GAC	163	0.37		GGC	161	0.44
Glu	GAA	868	1.54		GGA	590	1.61
	GAG	262	0.46	GGG	239	0.65	

中性绘图和对应性分析均支持普通油茶密码偏好受突变和选择共同作用。这与已报道的文心兰(李冬梅等, 2012)、蝴蝶兰(续晨等, 2010)、银白杨(Zhou et al, 2008)、甜荞(罗洪等, 2015)等有

相似的结论;本文结果也印证了选择对叶绿体密码子使用起显著影响(Morton, 1998)。

本研究采用由高表达优越密码子和高频密码子中选出两者共有的密码子,可以避免由于单方

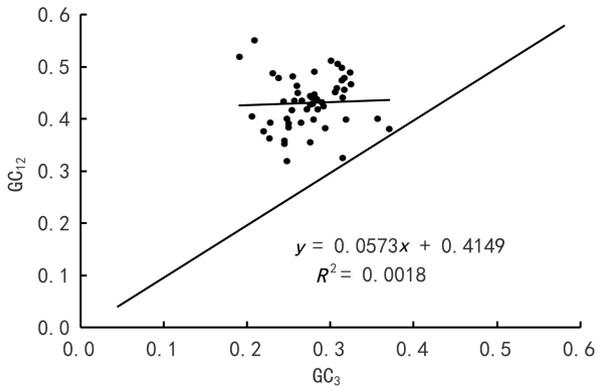


图 1 中性绘图分析

Fig. 1 Neutrality plot analysis

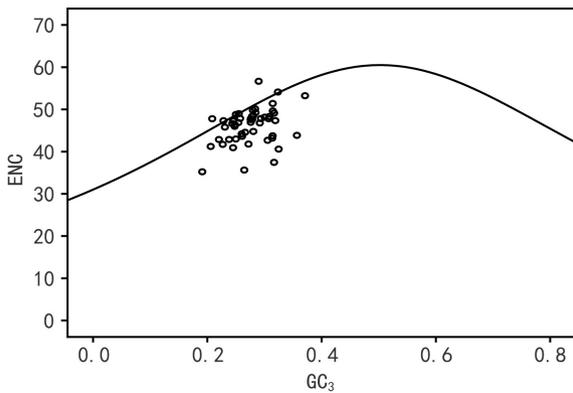


图 2 ENC-plot 绘图分析

Fig. 2 Analysis on ENC and GC<sub>3</sub> relationship

表 4 ENC 比值频数分布

Table 4 Distribution of ENC ration

组限 Class limit	组中值 Class midvalue	频数 Frequency number	组频 Frequency
-0.1~0	-0.05	5	0.10
0~0.1	0.05	28	0.55
0.1~0.2	0.15	13	0.25
0.2~0.3	0.25	5	0.10
总计 Total		51	1

CCU、ACU、GCU、GCA、UGU、CGU、AGU、UUG、GGU 等 18 个密码子为普通油茶叶绿体基因的最优密码子。在此基础上可以通过对目的基因的密码子进行优化,提高该基因的表达效率,从而为改良普通油茶重要性状奠定坚实基础。

### 参考文献:

- CHANG WX, YAO XH, 2016. Study on self-compatibility of *Camellia oleifera* clones [J]. For Res, 29(4): 508-514. [常维霞,姚小华,2016.油茶无性系自交亲和性分析[J].林业科学研究,29(4): 508-514.]
- CHANG WX, YAO XH, LONG W, et al, 2016. Cross-compatibility of four kinds of *Camellia* species [J]. Bull Bot Res, 36(4): 527-534. [常维霞,姚小华,龙伟,等,2016.4个油茶物种杂交亲和性分析[J].植物研究,36(4): 527-534.]
- CHEN YZ, LUO J, CHEN LS, et al, 2014. Drought injury and countermeasure of *Camellia oleifera* in Hunan Province [J]. Nonwood For Res, 32(3): 22-29. [陈永忠,罗健,陈隆升,等,2014.湖南省油茶旱害及防治对策[J].经济林研究,32(3): 22-29.]
- DANIELL H, CHASE C, 2004. Molecular biology and biotechnology of plant organelles [M]. Dordrecht, Springer.
- HU SS, LUO H, WU Q, et al, 2016. Analysis of codon bias of chloroplast genome of *Fagopyrum tataricum* [J]. Mol Plant Breed, 14(2): 309-317. [胡莎莎,罗洪,吴琦,等,2016.苦荞叶绿体基因组密码子偏爱性分析[J].分子植物育种,14(2): 309-317.]
- JIANG Y, DENG F, WANG H, et al, 2008. An extensive analysis on the global codon usage pattern of baculoviruses [J]. Arch Virol, 153(12): 2273-2282.
- LI DM, LÜ FB, ZHU GF, et al, 2012. Analysis on codon usage of chloroplast genome of *Oncidium Gower Ramsey* [J]. Guangdong Agric Sci, 10: 61-65. [李冬梅,吕复兵,朱根发,等,2012.文心兰叶绿体基因组密码子使用的相关分析[J].广东农业科学,10: 61-65.]
- LIU Q, XUE Q, 2005. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species [J]. J Gene, 84(1): 55-62.
- LIU QP, XUE QZ, 2004. Codon usage in the chloroplast genome of rice (*Oryza sativa* L. ssp. *japonica*) [J]. Acta Agronom Sin, 30(12): 1220-1224. [刘庆坡,薛庆中,2004.水稻叶绿体基因组的密码子用法[J].作物学报,30(12): 1220-1224.]
- LUO C, 2016. The area of *Camellia* achieved 64 million acres [EB/OL]. WWW. cnfood. cn/n/2016/1219/98628.html. [罗晨,2016.全国油茶面积已达6400多万亩[EB/OL]. WWW. cnfood. cn/n/2016/1219/98628.html.]
- LUO H, HU SS, WU Q, et al, 2015. Analysis of buckwheat

法存在问题而导致结果出错,最终确定了 CUU、AUU、GUU、GUA、UAA、CAA、AAA、GAC、GAA、

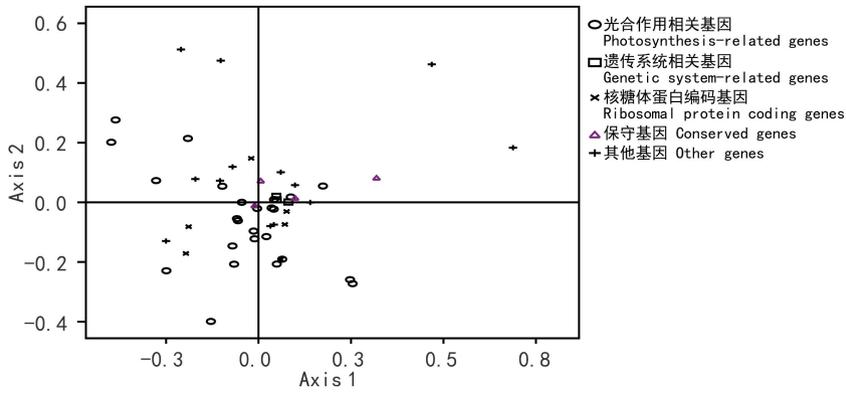


图 3 基于 RSCU 的对应性分析

Fig. 3 Correspondence analysis based on RSCU

表 5 普通油茶叶绿体高/低表达样本组的密码子用法

Table 5 Codon usage of high/low expression sample group in *Camellia oleifera*

氨基酸 Amino acid	密码子 Codon	高表达基因 High expression gene		低表达基因 Low expression gene		$\Delta$ RSCU
		数目 Number	RSCU	数目 Number	RSCU	
Phe	UUU	23	1.05	106	1.09	-0.04
	UUC	21	0.95	89	0.91	0.04
Leu	UUA	16	1.35	77	1.29	0.06
	UUG **	21	1.77	84	1.41	0.36
	CUU **	21	1.77	86	1.45	0.32
	CUC	1	0.08	31	0.52	-0.44
	CUA	11	0.93	52	0.87	0.06
	CUG	1	0.08	27	0.45	-0.37
Ile	AUU *	36	1.52	126	1.29	0.23
	AUC	17	0.72	71	0.73	-0.01
	AUA	18	0.76	95	0.98	-0.22
Met	AUG	26	1.00	73	1.00	0.00
Val	GUU ***	19	2.00	51	1.38	0.62
	GUC	0	0.00	23	0.62	-0.62
	GUA ***	18	1.89	45	1.22	0.67
	GUG	1	0.11	29	0.78	-0.67
Tyr	UAU	16	1.45	118	1.61	-0.16
	UAC *	6	0.55	29	0.39	0.16
TER	UAA ***	3	1.80	2	1.20	0.60
	UAG	2	1.20	2	1.20	0.00
His	CAU	14	1.47	78	1.61	-0.14
	CAC *	5	0.53	19	0.39	0.14
Gln	CAA **	18	1.71	92	1.35	0.36
	CAG	3	0.29	44	0.65	-0.36
Asn	AAU	23	1.12	155	1.51	-0.39

续表5

氨基酸 Amino acid	密码子 Codon	高表达基因 High expression gene		低表达基因 Low expression gene		$\Delta$ RSCU
		数目 Number	RSCU	数目 Number	RSCU	
	AAC **	18	0.88	50	0.49	0.38
Lys	AAA **	28	1.70	143	1.32	0.38
	AAG	5	0.30	74	0.68	-0.38
Asp	GAU	11	1.47	164	1.65	-0.18
	GAC *	4	0.53	35	0.35	0.18
Glu	GAA *	31	1.59	143	1.35	0.24
	GAG	8	0.41	69	0.65	-0.24
Ser	UCU	19	1.70	86	1.62	0.08
	UCC	12	1.07	61	1.15	-0.08
	UCA	8	0.72	60	1.13	-0.41
	UCG	7	0.63	38	0.72	-0.09
Pro	CCU *	13	1.68	54	1.38	0.30
	CCC	4	0.52	33	0.84	-0.32
	CCA	9	1.16	45	1.15	0.01
	CCG	5	0.65	25	0.64	0.01
Thr	ACU ***	17	1.89	48	1.33	0.56
	ACC *	8	0.89	27	0.75	0.14
	ACA	10	1.11	46	1.28	-0.17
	ACG	1	0.11	23	0.64	-0.53
Ala	GCU ***	39	2.36	54	1.73	0.63
	GCC	6	0.36	28	0.90	-0.54
	GCA *	20	1.21	31	0.99	0.22
	GCG	1	0.06	12	0.38	-0.22
Cys	UGU ***	6	2.00	34	1.39	0.61
	UGC	0	0.00	15	0.61	-0.61
TER	UGA	0	0.00	1	0.60	-0.60
Trp	UGG	15	1.00	67	1.00	0.00
Arg	CGU ***	23	1.97	37	0.88	1.09
	CGC	3	0.26	17	0.41	-0.15
	CGA	16	1.37	59	1.41	-0.04
	CGG	2	0.17	24	0.57	-0.40
Ser	AGU **	17	1.52	57	1.08	0.44
	AGC	4	0.36	16	0.30	0.06
Arg	AGA	21	1.80	81	1.94	-0.14
	AGG	5	0.43	33	0.79	-0.36
Gly	GGU ***	28	1.90	45	0.94	0.96
	GGC	7	0.47	19	0.40	0.07
	GGA	18	1.22	80	1.67	-0.45
	GGG	6	0.41	48	1.00	-0.59

注: \* 表示  $\Delta$ RSCU > 0.08; \*\* 表示  $\Delta$ RSCU > 0.3; \*\*\* 表示  $\Delta$ RSCU > 0.5。

Note: \* means  $\Delta$ RSCU > 0.08; \*\* means  $\Delta$ RSCU > 0.3; \*\*\* means  $\Delta$ RSCU > 0.5.

- chloroplast gene codon bias [J]. *Genomics Appl Biol*, 34 (11): 2457-2464. [罗洪,胡莎莎,吴琦,等,2015.甜荞叶绿体基因密码子偏爱性分析[J].基因组学与应用生物学,34(11):2457-2464.]
- MORTON BR, 1998. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages [J]. *J Mol Evol*, 46(4): 449-459.
- MORTON BR, 1999. Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis* [J]. *Proc Nat Acad Sci USA*, 96(9): 5123-5128.
- SHI C, LIU Y, HUANG H, et al, 2013. Contradiction between plastid gene transcription and function due to complex post-transcriptional splicing: An exemplary study of *ycf15* function and evolution in angiosperms [J]. *PLoS ONE*, 8 (3): e59620.
- WANG GL, FANG HY, 2014. Plant genetic engineering [M]. 2nd ed. Beijing: Science Press. [王关林,方宏筠, 2014.植物基因工程[M].第2版.北京:科学出版社.]
- WRIGHT F, 1990. The 'effective number of codons' used in a gene [J]. *Gene*, 87(1): 23-29.
- XU LP, TAN GJ, 2015. Epidemic of *Camellia oleifera* diseases and their relations to the ecological conditions [J]. *J Anhui Agric Univ*, 42(2): 272-275. [徐丽萍,檀根甲, 2015.油茶主要病害流行与生态条件的关系和生态调控技术[J].安徽农业大学学报,42(2):272-275.]
- XU C, BEN AL, CAI XN, 2010. Analysis of synonymous codon usage in chloroplast genome of *Phalaenopsis aphrodite* subsp. *formosana* [J]. *Mol Plant Breed*, 8(5): 945-950. [续晨,贲爱玲,蔡晓宁,2010.蝴蝶兰叶绿体基因组密码子使用的相关分析[J].分子植物育种,8(5):945-950.]
- YANG GF, SU KL, ZHAO YR, et al, 2015. Analysis of codon usage in the chloroplast genome of *Medicago truncatula* [J]. *Acta Pratacul Sin*, 171-179. [杨国锋,苏昆龙,赵怡然,等,2015.蒺藜苜蓿叶绿体密码子偏好性分析[J].草业学报,24(12):171-179.]
- YANG HJ, LIU GS, ZHANG ST, et al, 2012. Codon usage bias studies and cluster analysis on chloroplastic genes in tobacco [J]. *Acta Tabac Sin*, 18(2): 37-43. [杨惠娟,刘国顺,张松涛,等,2012.烟草叶绿体密码子的偏好性及聚类分析[J].中国烟草学报,18(2):37-43.]
- ZHANG EH, WANG XY, QIN ZH, et al, 2016. Genetic diversity analysis of *Camellia oleifera* in Guangxi using SSR markers [J]. *Guihaia*, 36(7): 806-811. [张恩慧,王晓云,覃子海,等,2016.广西普通油茶种质资源遗传多样性的SSR分析[J].广西植物,36(7):806-811.]
- ZHANG WL, 2016. Supply and demand of China's oilseeds and edible vegetable oil and their prospects during the 13th five-year plan period [J]. *Agric Outlook*, (2): 4-8. [张雯丽, 2016.中国油料及食用植物油供需形势及“十三五”展望[J].农业展望,(2):4-8.]
- ZHU SG, XU CF, 2016. Biochemistry [M]. 4th ed. Beijing: Higher Education Press. [朱圣庚,徐长发,2016.生物化学[M].第4版.北京:高等教育出版社.]
- ZHUANG RL, 2007. Chinese oil tea [M]. 2nd ed. Beijing: Chinese Forestry Publishing House. [庄瑞林,2007.中国油茶[M].第2版.北京:中国林业出版社.]
- ZHOU M, LONG W, LI X, 2008. Analysis of synonymous codon usage in chloroplast genome of *Populus alba* [J]. *J For Res*, 19(4): 293-297.
- ZUO JL, WANG YJ, GONG C, et al, 2012. Effects of different management measures on the fruit shape growth of high yield *Camellia oleifera* under during the summer drought period [J]. *J Centr Southern Univ For & Technol*, 32(4): 15-20. [左继林,王玉娟,龚春,等,2012.高产油茶夏旱期不同经营措施对其果形生长的影响[J].中南林业科技大学学报,32(4):15-20.]