

# 大花香水月季叶绿体基因组密码子使用偏性分析

牛元,徐琼\*,王崧德,代立兰,庄健,赵亚兰

(兰州市农业科技研究推广中心,甘肃兰州730010)

**摘要:**以大花香水月季叶绿体基因组为研究对象,利用 codonW 等软件,通过分析大花香水月季叶绿体基因组 54 个基因密码子的使用偏性,探讨影响密码子偏好性形成的主要因素。结果表明:1)中性绘图分析显示  $GC_3$  与  $GC_{12}$  的相关系数为 0.141,不显著相关,说明密码子第 1、2 位与第 3 位碱基组成变异无明显关联。2)ENC 绘图分析显示,期望值曲线下方分布了大部分基因,偏性较强,而仅少量基因沿曲线分布或在曲线附近,偏性较弱。3)基于 RSCU 的对应性分析表明,第 1 轴能够解释基因总变异的 10.35%,第 1 轴与 CAI、CBI 呈负显著相关。4)PR2-plot 分析显示,密码子第 3 位碱基的使用存在偏性,嘧啶比嘌呤使用更为频繁。5)最终确定了 TTT、CTT 和 TCA 等 13 个最优密码子,且均以 A 或 T 结尾。综上所述,大花香水月季叶绿体基因组密码子的使用受突变影响较弱,更多地受到如选择压力等其他因素的影响。

**关键词:**大花香水月季;叶绿体基因组;GC 含量;密码子使用;最优密码子

**中图分类号:**S685.12      **文献标志码:**A      **文章编号:**1001-7461(2018)03-0123-08

## An Analysis on Codon Usage Bias of Chloroplast Genome of *Rosa odorata* var. *gigantea*

NIU Yuan, XU Qiong\*, WANG Yu-de, DAI Li-lan, ZHUANG Jian, ZHAO Ya-lan

(Lanzhou Agro-technical Research and Popularization Center, Lanzhou, Gansu 730010, China)

**Abstract:** Taking chloroplast genome of *Rosa odorata* var. *gigantea* as research object, fifty-four genes were analyzed by CodonW and other software to explore the main factors affecting codon usage bias. The results were as follows. 1) Neutral mapping analysis showed that the correlation coefficient between  $GC_{12}$  and  $GC_3$  was 0.141, and the correlation was not significant, indicating that there was no significant correlation between the first two and the third base composition variation of codon. 2) ENC plot analysis showed that only a few genes in the genome were distributed along the expected value curve or near the expected value curve, which bias were weak, while most genes were distributed below the expected value curve, which bias were strong. 3) RSCU based correspondence analysis showed that the first axis explained 10.35% of the total variation of the gene and was negatively related to CAI and CBI. 4) PR2-plot analysis showed that the third base of codon usage was biased, where pyrimidine was used more frequently than purine. 5) Thirty preferred codon were identified finally, which all ended with A and T. In conclusion, the codon usage bias is formed under less effect of mutation and more effect of other factors such as selection pressure.

**Key words:** *Rosa odorata* var. *gigantea*; chloroplast genome; GC content; codon usage bias; preferred codon

密码子是生命遗传信息传递过程当中的基本单位。20 种自然界中存在的氨基酸中,每种对应至少 1 种,最多 6 种密码子。同义密码子即编码相同氨基酸的密码子<sup>[1]</sup>。如果没有任何选择压力或者基因

收稿日期:2017-08-21 修回日期:2017-12-27

基金项目:甘肃省科技重大专项(1302NKDA028);兰州市科技攻关计划项目(2010-1-239,2016-3-4);兰州市科技局人才项目(2015-RC-87);甘肃省农科院院地科技合作项目(2017GAAS63)。

作者简介:牛元,女,硕士,农艺师,研究方向:花卉种质资源创新技术。E-mail:530144903@qq.com

\*通信作者:徐琼,女,研究员,研究方向:观赏园艺栽培技术。E-mail:2572004227@qq.com

突变偏倚,则同义密码子以相同概率出现。但实际生命体中,同义密码子的使用频率存在差异,形成密码子使用偏性,即某物种或基因偏向使用同一种或几种特定的同义密码子<sup>[2]</sup>。研究表明,密码子使用偏性广泛存在于细菌、真菌、植物、动物及人类中,其影响因素有很多,如基因序列碱基组成的偏好性<sup>[3-4]</sup>、自然选择效应<sup>[5-6]</sup>、tRNA 丰度<sup>[7-10]</sup>、mRNA 二级结构<sup>[11-12]</sup>、氨基酸保守性<sup>[3]</sup>及每个蛋白的疏水性水平等。研究密码子使用偏性不但可以确定最优密码子,从而设计基因表达载体,提高目的基因表达量,还可以利用已知密码子使用模式来判断未知基因的表达,或者根据其与某种功能的关联程度,来预测某些未知功能基因。

蔷薇属(*Rosa*)是现代月季种群的重要亲本,我国拥有诸多蔷薇属植物的基因资源和优良种质<sup>[13]</sup>。大花香水月季(*R. odorata* var. *gigantea*)属蔷薇科蔷薇属的落叶灌木,也称巨花蔷薇,为所有蔷薇属野生种中开花最大者,具有气味芳香、花型大、耐低温等优良园艺性状,为香水月季原始亲本类型之一,是培育蔷薇科大花品种的重要种质资源<sup>[14]</sup>。月季新品种的选育主要依赖于有性杂交和自然突变中的选择,随着技术的发展,目前多利用分子标记等生物学手段达到辅助选育的目的<sup>[13,15]</sup>。大花香水月季主要分布在我国云南及缅甸、泰国及越南等与我国接壤的边境处<sup>[16]</sup>。蹇洪英<sup>[17]</sup>等通过核型分析鉴定其为二倍体;育种方面,王其刚<sup>[18]</sup>等研究了雌蕊群的授粉率和花粉管生长途径,为大花香水月季的资源保护提供了依据;孟令宁<sup>[19]</sup>等、李纯佳<sup>[20]</sup>等建立了以带腋芽的茎段为外植体的再生体系,为遗传改良奠定了基础;天然群体遗传性分析方面,邱显钦<sup>[21]</sup>、邵珠华<sup>[22-23]</sup>等分别通过表型分析、SSR 分子标记等手段进行了研究,J. MENG<sup>[24-25]</sup>等通过开发微卫星分子标记,并从生物地理学和生物多样性保护角度分析了其群体结构;此外,X. L. Guo<sup>[26]</sup>等从花期调控的角度对其与多季开花的品种“月月粉”进行了比较转录组方面的研究。大花香水月季叶绿体基因组的测序现已实现,但未见有文献报道其密码子应用方面的研究。本研究通过对大花香水月季叶绿体基因组密码子使用偏性的分析,讨论影响密码子偏性形成的主要因素,了解密码子用法,为今后对外源基因进行改造,提高以大花香水月季为目的宿主的外源基因的表达效率具有重要意义。

1 材料与方法

1.1 序列获取与整理

从 NCBI 数据库中下载大花香水月季叶绿体基

因组序列(GeneBank 登录号为 KF753637,序列全长 156 634 bp)。根据其注释信息可知该叶绿体基因组共包含 88 条基因编码序列,占全长的50.92%。将下载到的 88 条 CDS(coding DNA sequence),剔除 30 条长度<300 bp 及非 3 整数倍的序列和 4 条重复基因序列,筛选出 54 条序列用于本研究各项分析。

1.2 方法

1.2.1 密码子偏好性分析 运用软件 CodonW 对大花香水月季 54 条基因序列的 ENC(有效密码子数)、GC<sub>1</sub>、GC<sub>2</sub>、GC<sub>3</sub>(密码子第 1、2、3 位的 GC 含量)、GC<sub>all</sub>(密码子总 GC 含量)、RSCU(密码子使用相对概率)、CAI(密码子适应指数)、CBI(密码子偏好指数)及 Fop(最优密码子频率)等参数进行统计,并用 SPSS21.0 做相关性分析。

1.2.2 中性绘图分析 将各基因的 GC<sub>3</sub> 为横坐标,GC<sub>12</sub>(GC<sub>1</sub> 和 GC<sub>2</sub> 的平均值)为纵坐标作图,分析密码子 3 个不同位点碱基组成间的相关性,判断密码子偏性的主要影响因素。若 GC<sub>12</sub> 与 GC<sub>3</sub> 显著性不相关,说明第 1、2 位与第 3 位碱基组成差异较大,密码子偏性更多受选择影响。反之,若两者显著相关,同时回归系数接近 1,说明 3 个位点碱基组成相似,密码子偏性主要受突变影响<sup>[27]</sup>。

1.2.3 ENC 绘图分析 将各基因的 GC<sub>3</sub>、ENC 值分别作为横、纵坐标,分析密码子偏性受碱基组成的影响。理论 ENC 值的计算方法为:ENC=2+GC<sub>3</sub>+29/(GC<sub>3</sub><sup>2</sup>+(1-GC<sub>3</sub>)<sup>2</sup>),理论值代表仅由碱基组成决定其偏性。基因分布在标准曲线之下较远位置,说明该基因密码子偏性受选择因素影响;而在标准曲线上或附近分布的基因,其偏性仅受突变影响<sup>[28]</sup>。

1.2.4 对应性分析 为阐明在密码子使用模式上不同基因的变异趋势,基于 RSCU 值,应用 codonW 软件进行对应性分析。所有基因经软件处理后,分布于一个 58 维(58 个同义密码子)的向量空间中,根据基因间值的变异及向量的情况,可判断密码子偏性的主要影响因素。密码子使用的最大变化差异量体现在第 1 轴,随后副轴依次降低。基因在第 1、2 轴之间的分散程度,代表了密码子的主要变化趋势。以第 1、2 轴分别作横、纵坐标作图,图中每个点(基因)的分布,可作为其密码子使用模式的判断依据<sup>[29]</sup>。

1.2.5 PR2-plot 分析 选择 4 种密码子编码的氨基酸进行密码子第 3 位上 4 种碱基组成情况的分析,以 G<sub>3</sub>/(G<sub>3</sub>+C<sub>3</sub>)|4 作为横坐标,以 A<sub>3</sub>/(A<sub>3</sub>+T<sub>3</sub>)|4 作为纵坐标作图,图的中心点(A=T 且 G=C)发出的矢量表示偏性的方向和程度<sup>[30]</sup>。

1.2.6 最优密码子分析 鉴定大花香水月季叶绿体基因组的最优密码子,须结合高频率密码子和高表达密码子。根据软件计算得到的 *RSCU* 值,将>1 的密码子确定为高频密码子。以 *ENC* 值为偏好性标准,分别选择两级 10% 的基因,建立高低偏性库,将 $\triangle RSCU>0.08$  的密码子视为高表达密码子。最终将同时满足 2 种条件的密码子定义为大花香水月季叶绿体基因组的最优密码子<sup>[31]</sup>。

2 结果与分析

2.1 密码子偏好性分析

大花香水月季叶绿体基因组的 54 条 DNA 编码序列的碱基组成与 *ENC* 值,由 codonW 计算得出(表 1),得到其密码子 3 个碱基平均 GC 含量为 38.54%,第 1 位 GC 含量为 47.09%,第 2 位为 39.55%,第 3 位为 38.54%。可以看出密码子不同位置的 GC 含量并非平均分布,分布趋势为第 1 位

最高,其次为第 2 位,第 3 位最低。因此,在大花香水月季中,叶绿体基因组的密码子使用频率以 A/T 结尾的高于以 G/C 结尾的,这与叶绿体基因组富含 AT 的特征相符合。

*ENC* 的理论值范围是 21~61,其值表示密码子偏性的强弱<sup>[28]</sup>。值为 20 时,偏性最强,即同义密码子完全偏倚,值为 61 时,代表完全没有偏倚。从表 1 中可以看到,54 个基因的 *ENC* 值取值范围 35.94~57.63,平均为 48.40,说明该基因组密码子偏性较弱。

各位置的 GC 含量与 *ENC* 值的相关性分析见表 2,GC<sub>1</sub> 与 GC<sub>2</sub> 极显著相关,但 GC<sub>1</sub>、GC<sub>2</sub> 与 GC<sub>3</sub> 均不显著相关,说明叶绿体基因组密码子碱基中,第 1 位与第 2 位相似度极高,与第 3 位差别较大。*ENC* 值与 GC<sub>1</sub> 相关性不显著,与 GC<sub>2</sub> 显著相关,与 GC<sub>3</sub> 相关性极显著,说明第 2、3 位碱基对密码子偏好性有影响,但第 3 位影响较大。

表 1 依据功能分类的各基因密码子不同位置的 GC 含量  
Table 1 GC content of different positions of each gene classified according to function

类别	基因	GC 含量/%				ENC	类别	基因	GC 含量/%				ENC
		GC <sub>1</sub>	GC <sub>2</sub>	GC <sub>3</sub>	GC <sub>all</sub>				GC <sub>1</sub>	GC <sub>2</sub>	GC <sub>3</sub>	GC <sub>all</sub>	
1	<i>rps14</i>	43.56	48.51	33.66	41.91	44.00	2	<i>petD</i>	50.93	39.13	23.60	37.89	39.69
	<i>rps18</i>	36.27	43.14	25.49	34.97	38.27		<i>rbcL</i>	57.35	43.49	29.41	43.42	46.35
	<i>rpl20</i>	37.29	39.83	23.73	33.62	46.02		<i>petB</i>	47.96	40.72	28.51	39.06	41.60
	<i>rpl14</i>	55.28	38.21	25.20	39.57	48.81		<i>psbC</i>	54.22	46.20	32.70	44.37	46.82
	<i>rps8</i>	39.26	40.00	27.41	35.56	35.94		<i>atpB</i>	57.40	43.00	26.77	42.39	44.67
	<i>rps11</i>	51.80	56.12	28.78	45.56	51.48		<i>ndhD</i>	39.92	37.92	29.54	35.80	47.56
	<i>rpl22</i>	45.71	39.29	28.57	37.86	57.45		<i>atpA</i>	56.10	39.37	29.72	41.73	49.13
	<i>rps7</i>	52.56	44.87	23.08	40.17	44.08		<i>psbB</i>	54.62	46.17	30.26	43.68	47.49
	<i>rps4</i>	50.99	37.13	29.70	39.27	51.91		<i>psaB</i>	48.71	42.99	31.84	41.18	48.43
	<i>rps3</i>	48.40	31.51	26.48	35.46	52.92		<i>ndhB</i>	41.49	39.14	31.51	37.38	48.79
	<i>rps2</i>	43.46	42.62	29.96	38.68	52.44		<i>ndhF</i>	37.83	36.10	24.33	32.75	44.55
	<i>rpl16</i>	49.64	53.28	27.74	43.55	40.38		<i>psaA</i>	53.00	43.28	32.89	43.05	50.75
	<i>rpl2</i>	51.09	48.19	31.16	43.48	52.37		<i>ndhA</i>	43.41	38.46	23.90	35.26	45.10
	<i>rps12</i>	51.61	47.58	27.42	42.20	42.53	3	<i>ycf4</i>	43.78	41.08	34.59	39.82	57.63
2	<i>ndhE</i>	39.22	34.31	33.33	35.62	52.99		<i>ycf3</i>	47.93	37.87	30.18	38.66	55.27
	<i>ndhC</i>	47.93	33.88	27.27	36.36	43.27		<i>ycf1</i>	37.04	28.92	26.14	30.70	49.07
	<i>atpE</i>	50.00	39.55	32.09	40.55	55.43	4	<i>ycf2</i>	41.45	33.99	36.80	37.41	53.28
	<i>ndhJ</i>	50.94	37.74	30.19	39.62	50.57		<i>rpoA</i>	46.08	32.23	27.11	35.14	50.12
	<i>ndhI</i>	41.67	36.31	20.83	32.94	44.49		<i>rpoC1</i>	50.66	38.25	29.34	39.42	50.44
	<i>ndhG</i>	42.94	33.33	27.68	34.65	53.74	5	<i>rpoB</i>	50.61	37.91	28.76	39.09	49.51
	<i>atpF</i>	45.95	32.97	34.59	37.84	52.91		<i>rpoC2</i>	45.85	37.62	29.53	37.67	50.61
	<i>ndhK</i>	45.37	45.81	28.63	39.94	52.08		<i>orf188</i>	48.15	40.21	20.11	36.16	44.68
	<i>atpI</i>	49.19	36.29	26.21	37.23	47.17		<i>cemA</i>	38.26	27.39	30.43	32.03	48.88
	<i>petA</i>	54.52	36.14	33.33	41.33	52.19		<i>ccsA</i>	34.16	37.27	26.40	32.61	46.31
	<i>psbA</i>	49.72	42.94	31.64	41.43	40.31		<i>accD</i>	39.55	37.70	30.53	35.93	47.55
	<i>psbD</i>	52.26	43.22	34.46	43.31	45.74		<i>matK</i>	39.29	31.75	28.37	33.13	51.17
	<i>ndhH</i>	52.54	36.04	29.70	39.42	51.58		<i>clpP</i>	58.16	38.78	32.65	43.20	57.25

注:*ENC* 即有效密码子数;GC<sub>1</sub> 即密码子第 1 位的 GC 含量;GC<sub>2</sub> 即密码子第 2 位的 GC 含量;GC<sub>3</sub> 即密码子第 3 位的 GC 含量;GC<sub>all</sub> 即密码子总 GC 含量;类别 1 表示核糖体蛋白编码基因;类别 2 表示光合作用相关蛋白编码基因;类别 3 表示保守基因 *ycf*;类别 4 表示遗传结构相关的基因;类别 5 表示其他蛋白表达基因;下同。

表 2 各基因相关参数的相关性分析

基因	GC <sub>1</sub>	GC <sub>2</sub>	GC <sub>3</sub>	GC <sub>all</sub>
GC <sub>2</sub>	0.395 **			
GC <sub>3</sub>	0.165		0.066	
GC <sub>all</sub>	0.823 **	0.757 **	0.452 **	
ENC	0.117	-0.271 *	0.472 **	0.081

注：\* 代表  $P<0.05$ ；\* \* 代表  $P<0.01$ 。表 4 同。

2.2 中性绘图分析

大花香水月季叶绿体基因组各基因中性绘图分析见图 1。GC<sub>12</sub> 最小值为 0.328,最大值0.540;GC<sub>3</sub> 最小值 0.201,最大值 0.368。其中仅有 *ycf2* 基因的 GC<sub>12</sub> 与 GC<sub>3</sub> 的比值为 1.03,在图中接近对角线。此外,图中其余各点均分布在对角线上方。Pearson 相关性统计检验表明 GC<sub>3</sub> 与 GC<sub>12</sub> 相关性不显著,且回归系数(斜率)为 0.196 1,表示在碱基组成上的变异无明显关联。由此推断,密码子的用法受突变影响较弱,可能是受到翻译选择压力的限制<sup>[27]</sup>。

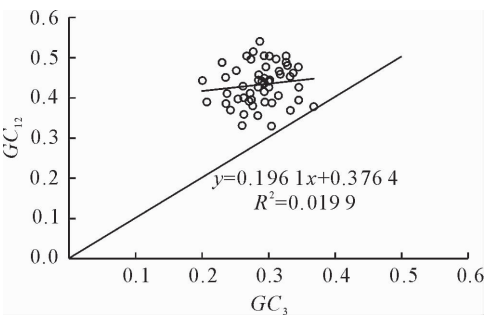


图 1 中性绘图分析

Fig. 1 Neutrality plot analysis

2.3 ENC 绘图分析

从 ENC 与 GC<sub>3</sub> 的相关性分析可以看出(图 2),大花香水月季叶绿体基因大部分都偏离了基因预期值的标准曲线,表明除由碱基突变引发的差异影响该基因组密码子偏性外,影响更多来自选择效应。ENC 值相对较小的基因位于标准曲线下方,说明这种类型偏好使用特定密码子,其密码子偏性较强;而位于曲线上方的基因的 ENC 值相对较大,说明此类基因使用密码子具有随机性,具有相对较弱的密码子偏性。此外,ENC 比值频数的分析结果表明(表 3),多数分布在 -0.05~0.15,此范围内 ENC 的实际值与期望值相差较大。

2.4 对应性分析

以第 1 轴和第 2 轴分别为横、纵坐标,将各基因在平面中分布。其中,分布相对集中的是光合系统基因,其他基因的分布较为分散。由表 4 可见,第一轴与 CAI、CBI 呈负显著相关,相关系数分别为 -0.509和-0.35,与 ENC、GC 和 GC<sub>3</sub> 相关系数分

别为 0.069、-0.221 和 0.18,但均未达显著水平。说明大花香水月季叶绿体密码子偏性的影响因素中,CAI 与 CBI 影响较大,其他因素也有一定影响。

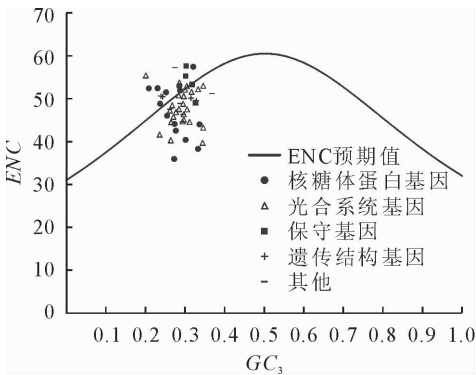


图 2 GC<sub>3</sub> 与 ENC 关联分析

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

表 3 ENC 比值频数分布

Table 3 Frequency distribution of ENC ratio

组限(区间)	组中值	组数	组频
-0.2~-0.1	-0.15	3	0.056
-0.1~0.0	-0.05	19	0.352
0.0~0.1	0.05	23	0.426
0.1~0.2	0.15	7	0.130
0.2~0.3	0.25	2	0.037
总数		54	1

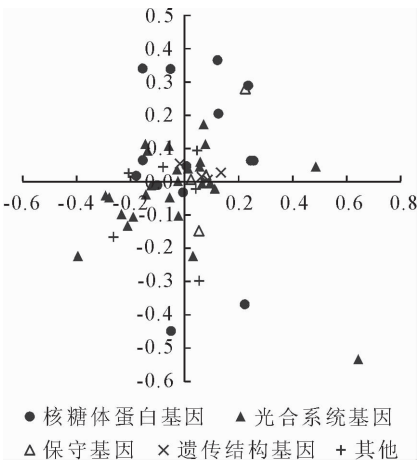


图 3 对应性分析(RSCU)

Fig. 3 Corresponding analysis (RSCU)

2.5 PR2 分析

通过 PR2-plot 分析,可以说明大花香水月季叶绿体基因组中 4 个密码子编码的氨基酸家族中第 3 位碱基的类别的使用情况(图 4),基因在平面图的 4 个区域中不均匀分布,基因较多分布在左下方图中,说明密码子第 3 位碱基的使用存在偏好性,嘧啶的使用比嘌呤更频繁,即从使用频率上看,T 高于 A, C 高于 G。而当密码子的使用仅由突变影响时,这 4 种碱基从使用频率上看是一致的。由此推断本研究中大花香水月季叶绿体基因组密码子使用模式,

受突变影响的同时,还受到其他因素的影响,如选择压力等。

表 4 各度量指标与第一轴的相关性

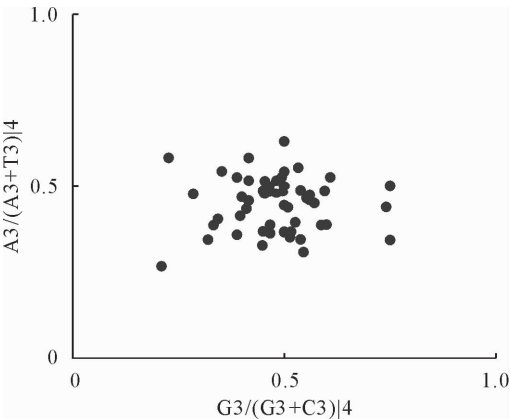
Table 4 Correlation between the measurement index and the first axis						
	Axis1	CAI	CBI	Fop	GC3	ENC
CAI	-0.509**					
CBI	-0.350**	0.704**				
Fop	-0.362**	0.739**	0.966**			
GC3	0.018	0.230	0.216	0.296*		
ENC	0.069	-0.093	-0.292*	-0.205	0.472**	
GC	-0.221	0.320*	0.483**	0.497**	0.443**	0.080

2.6 最优密码子分析

将 54 条 CDS 作为一个整体,运行 codonW 计算 RSCU 值,选出 RSCU>1 的密码子为高频密码子,即 TTT、TTA 等 30 个密码子为大花香水月季叶绿体的高频密码子(表 5 中以下划线显示)。将 codonW 计算得到的 ENC 值大小排序,前 10%(前 5 条基因)构成高表达基因库,后 10%构成低表达基因库。分别计算密码子在 2 个基因库中的 RSCU 值,从中筛选出 $\Delta RSCU \geq 0.08$  的密码子,即 TTT、CTT 等 32 个密码子为高表达优越密码子(表 5 中以 \* 号表示),其中有 12 个以 G 结尾,8 个以 T 结

尾,7 个以 A 结尾,5 个以 C 结尾。

结合高频密码子和高表达优越密码子,将同时满足 2 种条件的密码子确定为大花香水月季叶绿体基因组的最优密码子。即最终确 TTT、CTT、TCA、ACA、TAT、TAA、CAT、AAT、GAT、TGT、CGA、AGT 和 GGA 等 13 个密码子,且均以 A 或 T 结尾,其中 8 个以 T 结尾,5 个以 A 结尾。



注:图中各点代表了所研究的各个基因。G<sub>3</sub>/(G<sub>3</sub>+C<sub>3</sub>)/4 置信区间(95%)为 0.480±0.029,A<sub>3</sub>/(A<sub>3</sub>+T<sub>3</sub>)/4 置信区间(95%)为 0.445±0.021。

图 4 PR2 bias-plot 分析图

Fig. 4 Analysis of PR2 bias plot

表 5 大花香水月季叶绿体基因组最优密码子确定

Table 5 Determination of preferred codons in chloroplast genome of *Rosa odorata* var. *gigantea*

氨基酸	密码子		高表达基因组		低表达基因组		$\Delta RSCU$	基因组	
			数目	RSCU	数目	RSCU		数目	RSCU
Phe	<u>TTT</u>	*	27	1.50	25	1.06	0.44	820	1.36
	TTC		9	0.50	22	0.94	-0.44	386	0.64
LeT	<u>TTA</u>		20	1.69	30	2.17	-0.48	760	2.02
	<u>TTG</u>		14	1.18	19	1.37	-0.19	450	1.20
	<u>CTT</u>	*	19	1.61	21	1.52	0.09	479	1.27
	CTC	*	4	0.34	0	0.00	0.34	139	0.37
	CTA		5	0.42	12	0.87	-0.45	281	0.75
	CTG	*	9	0.76	1	0.07	0.69	148	0.39
Ile	<u>ATT</u>		37	1.42	42	1.58	-0.16	948	1.52
	ATC		14	0.54	18	0.68	-0.14	345	0.55
	ATA	*	27	1.04	20	0.75	0.29	576	0.92
Met	ATG		23	1.00	30	1.00	0.00	492	1.00
Val	<u>GTT</u>		13	1.16	23	1.70	-0.54	435	1.45
	GTC	*	4	0.36	0	0.00	0.36	129	0.43
	<u>GTA</u>		20	1.78	24	1.78	0.00	464	1.55
	GTG	*	8	0.71	7	0.52	0.19	168	0.56
Ser	<u>TCT</u>		13	1.66	25	2.54	-0.88	445	1.68
	TCC	*	7	0.89	5	0.51	0.38	259	0.98
	<u>TCA</u>	*	7	0.89	7	0.71	0.18	286	1.08
	TCG	*	6	0.77	3	0.31	0.46	165	0.62
Pro	<u>CCT</u>		12	1.66	20	1.70	-0.04	338	1.48
	CCC		4	0.55	10	0.85	-0.30	186	0.81

续表 5

氨基酸	密码子				高表达基因组		低表达基因组		$\Delta RSCU$	基因组	
					数目	<i>RSCU</i>	数目	<i>RSCU</i>		数目	<i>RSCU</i>
Thr	<u>CCA</u>				7	0.97	13	1.11	−0.14	256	1.12
	CCG	*	*		6	0.83	4	0.34	0.49	134	0.59
	<u>ACT</u>				11	1.07	23	1.92	−0.85	429	1.56
	ACC				10	0.98	12	1.00	−0.02	211	0.77
	<b><u>ACA</u></b>	*	*	*	16	1.56	12	1.00	0.56	335	1.22
Ala	ACG	*	*		4	0.39	1	0.08	0.31	122	0.44
	<u>GCT</u>				29	1.78	39	2.44	−0.66	543	1.81
	GCC	*	*	*	12	0.74	3	0.19	0.55	192	0.64
	<u>GCA</u>				14	0.86	16	1.00	−0.14	316	1.06
	GCG	*			10	0.62	6	0.38	0.24	147	0.49
Tyr	<b><u>TAT</u></b>	*			27	1.74	19	1.46	0.28	660	1.62
	TAC				4	0.26	7	0.54	−0.28	153	0.38
TER	<u>TAA</u>	*	*	*	4	2.40	3	1.80	0.60	30	1.67
	TAG				0	0.00	2	1.20	−1.20	13	0.72
His	<b><u>CAT</u></b>	*			7	1.27	8	1.14	0.13	381	1.49
	CAC				4	0.73	6	0.86	−0.13	130	0.51
Gln	<u>CAA</u>				24	1.45	19	1.90	−0.45	607	1.54
	CAG	*	*		9	0.55	1	0.10	0.45	179	0.46
Asn	<b><u>AAT</u></b>	*			30	1.46	31	1.32	0.14	794	1.53
	AAC				11	0.54	16	0.68	−0.14	241	0.47
Lys	<u>AAA</u>				27	1.64	28	1.75	−0.11	858	1.53
	AAG	*			6	0.36	4	0.25	0.11	260	0.47
Asp	<b><u>GAT</u></b>	*	*		26	1.68	12	1.20	0.48	713	1.62
	GAC				5	0.32	8	0.80	−0.48	166	0.38
GLT	GAA				44	1.40	37	1.64	−0.24	860	1.49
	GAG	*			19	0.60	8	0.36	0.24	295	0.51
Cys	<b><u>TGT</u></b>	*			6	1.50	3	1.20	0.30	177	1.51
	TGC				2	0.50	2	0.80	−0.30	58	0.49
TER	TGA	*	*	*	1	0.60	0	0.00	0.60	11	0.61
Trp	TGG				14	1.00	17	1.00	0.00	379	1.00
Arg	<u>CGT</u>				8	0.76	18	1.46	−0.70	291	1.38
	CGC				5	0.48	5	0.41	0.07	88	0.42
	<b><u>CGA</u></b>	*			18	1.71	19	1.54	0.17	286	1.36
	CGG	*			6	0.57	4	0.32	0.25	98	0.46
Ser	<b><u>AGT</u></b>	*	*		11	1.40	9	0.92	0.48	327	1.24
	AGC				3	0.38	10	1.02	−0.64	105	0.40
Arg	<u>AGA</u>				16	1.52	23	1.86	−0.34	374	1.77
	AGG	*	*	*	10	0.95	5	0.41	0.54	129	0.61
Gly	<u>GGT</u>				18	1.29	36	2.00	−0.71	493	1.31
	GGC	*			10	0.71	8	0.44	0.27	170	0.45
	<b><u>GGA</u></b>	*			18	1.29	21	1.17	0.12	562	1.50
	GGG	*	*		10	0.71	7	0.39	0.32	278	0.74

注：\* 代表 $\Delta RSCU>0.08$ ，\* \* 代表 $\Delta RSCU>0.3$ ，\* \* \* 代表 $\Delta RSCU>0.5$ 。带下划线显示的密码子表示其基因组  $RSCU>1$ ，加粗显示的密码子表示其为最优密码子。

### 3 结论与讨论

大花香水月季叶绿体基因组密码子的使用受突变影响较弱,更多地受到如选择压力等其他因素的影响。最终结合高频密码子和高表达优越密码子,确定了 TTT、CTT、TCA、ACA、TAT、TAA、CAT、

AAT、GAT、TGT、CGA、AGT 和 GGA 等 13 个密码子为大花香水月季叶绿体基因组的最优密码子。

密码子使用偏性是物种对基因组环境和自然进化压力的适应性选择,影响密码子使用偏性的因素有很多,如基因序列碱基组成的偏好性、自然选择效应、mRNA 二级结构、基因长度、tRNA 丰度及每个

蛋白的疏水性水平及氨基酸保守性等<sup>[3-12]</sup>。其中,最普遍的影响因素即密码子的碱基组成,密码子第 3 位上碱基突变受到的选择压力较小,因此,GC<sub>3</sub> 通常被作为分析密码子使用的重要依据。本研究中中性绘图分析显示密码子第 1、2 位与第 3 位碱基组成变异无明显关联。而 PR2-plot 分析显示,密码子第 3 位碱基的使用存在偏性,嘧啶比嘌呤使用更为频繁。这与前人对陆地棉(*Gossypium hisutum*)<sup>[32]</sup> 和小麦(*Triticum aestivum*)<sup>[33]</sup> 叶绿体基因组的研究结果相一致。而杨国锋<sup>[34]</sup> 等对蒺藜苜蓿(*Medicago truncatula*)叶绿体基因组的研究发现,密码子第 3 位的碱基使用频率上,T 高于 A,G 高于 C。张月荣<sup>[35]</sup> 研究禾本科植物发现,第 3 位碱基 T 的使用频率高于 A,G 和 C 的使用频率相当。说明不同植物的叶绿体基因组的碱基组成存在一定差异,密码子偏好的影响因素也不同。

ENC 值指基因序列中,同义密码子平均使用频率与密码子使用频率,两者偏差的量化值,取值范围 20~61。参照 F. Wright<sup>[28]</sup> 等的研究,以 35 作为偏性强弱的区分标准,据此判断大花香水月季叶绿体基因组密码子偏性较弱。本研究中,确定高表达优越密码子时,基于 ENC 值建立了高低表达基因库。依据△RSCU 值划分等级,△RSCU>0.08、0.3 和 0.5 的分别有 32、17 个和 6 个。其中,△RSCU>0.5 的 6 个密码子全是以 A 或 T 结尾的密码子。结合高频密码子和高表达优越密码子得到的 13 个最有密码子均以 A 或 T 结尾,这与蝴蝶兰(*Phalaenopsis aphrodite* subsp. *formosana*)<sup>[36]</sup>、刺五加(*Eleutherococcus senticosus*)<sup>[37]</sup>、陆地棉<sup>[32]</sup> 和苦荞(*Fagopyrum tataricum*)<sup>[38]</sup> 等植物叶绿体基因的密码子偏性分析结果一致,说明大花香水月季与其他双子叶植物一样偏爱使用以 AT 结尾的密码子。

参考文献:

[1] RÉDEI G P. Encyclopedia of genetics,genomics,proteomics and informatics[M]. Berlin:Springer,2008:383.

[2] DURET L. Evolution of synonymous codon usage in metazoan [J]. Current Opinion in Genetics & Development,2002,12 (6):640-649.

[3] ROMERO H,ZAVALA A,MUSTO H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces[J]. Nucleic Acids Research,2000, 28(10):2084.

[4] KARLIN S,MRÁZEK J. What drives codon choices in human genes[J]. Journal of Molecular Biology,1996,262(4):459-472.

[5] SHARP P M,LI W H. An evolutionary perspective on synonymous codon usage in unicellular organisms[J]. Journal of Mo-

lecular Evolution,1986,24(1/2):28-38.

[6] SHARP P M,COWE E. Synonymous codon usage in *Saccharomyces cerevisiae*[J]. Yeast,1991,7(7):657-678.

[7] MORIYAMA E N,POWELL J R. Codon usage bias and tRNA abundance in *Drosophila*[J]. Journal of Molecular Evolution, 1997,45(5):514.

[8] IKEMURA T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes;a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system [J]. Journal of Molecular Biology,1981,146(1):1-21.

[9] PERCUDANI R,PAVESI A,OTTONELLO S. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*[J]. Journal of Molecular Biology,1997,268(2):322-330.

[10] DURET L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes [J]. Trends in Genetics,2000,16(7):287.

[11] GU W,ZHOU T,MA J,*et al.* Folding type specific secondary structure propensities of synonymous codons [J]. IEEE Transactions on Nanobioscience,2003,2(3):150-157.

[12] GU W,ZHOU T,MA J,*et al.* The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*[J]. Biosystems,2004,73(2):89-97.

[13] 白锦荣,张启翔,潘会堂,等. 蔷薇属分子生物技术研究进展 [J]. 西北林学院学报,2009,24(6):43-49.

BAI J R,ZHANG Q X,PAN H T,*et al.* Recent advances in molecular biotechnology of the genus *Rosa* L[J]. Journal of Northwest Forestry University,2009,24(6):43-49. (in Chinese)

[14] WYLIE A P. The history of garden roses [M]. London: Journal of the Royal Horticultural Society,1954:555-571.

[15] 张非亚,罗玉兰,有祥亮,等. 基于转录组数据的月季 SSR 标记开发及花色遗传性分析[J]. 西北林学院学报,2017,32(4): 81-86.

ZHANG F Y,LUO Y L,YOU X L,*et al.* Development of SSR markers based on the transcriptome data of rose and analysis of flower color inheritance[J]. Journal of Northwest Forestry University,2017,32(4):81-86. (in Chinese)

[16] GU C Z,ROBERTSON K P. Rosa (Rosaceae): Vol. 9. [M]. Beijing & St. Louis:Science Press & Missouri Botanical Garden Press,2003:339-381.

[17] 蹇洪英,张颢,张婷,等. 香水月季(*Rosa odorata Sweet*) 不同变种的染色体及核型分析[J]. 植物遗传资源学报,2010,11 (4):457-461.

JIAN H Y,ZHANG H,ZHANG T,*et al.* Karyotype analysis of different varieties on *Rosa odorata Sweet*[J]. Journal of Plant Genetic Resources,2010,11(4):457-461. (in Chinese)

[18] 王其刚,刘红明,晏慧君,等. 大花香水月季的授粉率及花粉管生长途径[J]. 云南农业大学学报,2012,27(3):391-395.

WANG Q G,LIU H M,YAN H J,*et al.* Pollination rate and pollen tube growth in the gynoecium of *Rosa odorata Sweet* var. *gigantea*[J]. Journal of Yunnan Agricultural University, 2012,27(3):391-395. (in Chinese)

[19] 孟令宁,晏慧君,张颢,等. 大花香水月季再生体系的初步建立

[J]. 云南农业大学学报, 2012, 27(6): 870-874.

MENG L N, YAN H J, ZHANG H, *et al.* Preliminary establishment of the regeneration system of *Rosa odorata* var. *gigantea* [J]. Journal of Yunnan Agricultural University, 2012, 27(6): 870-874. (in Chinese)

[20] 李纯佳, 张颢, 周宁宁, 等. 大花香水月季 (*Rosa odorata* var. *gigantea*) 茎段组织培养的抗褐化研究 [J]. 西南农业学报, 2012, 25(3): 1047-1050.

LI C J, ZHANG H, ZHOU N N, *et al.* Study on anti-browning in shoots tissue culture of *Rosa odorata* var. *gigantea* [J]. Southwest China Journal of Agriculture Sciences, 2012, 25(3): 1047-1050. (in Chinese)

[21] 邱显钦, 唐开学, 蹇洪英, 等. 云南大花香水月季居群遗传多样性的 SSR 分析 [J]. 华中农业大学学报, 2011, 30(3): 300-304.

QIU X Q, TANG K X, JIAN H Y, *et al.* SSR markers based genetic diversity of populations of *Rosa odorata* Sweet var. *gigantea* in Yunnan Province [J]. Journal of Huazhong Agricultural University, 2011, 30(3): 300-304. (in Chinese)

[22] 邵珠华, 李名扬, 邱显钦, 等. 大花香水月季天然群体表型多样性研究 [J]. 江苏农业科学, 2010(2): 184-187.

SHAO Z H, LI M Y, QIU X Q, *et al.* Study on phenotypic diversity of *R. Odorata sweet* var. *gigantea* under natural population [J]. Jiangsu Agricultural Sciences, 2010(2): 184-187. (in Chinese)

[23] 邵珠华. 大花香水月季与长尖叶蔷薇的遗传多样性研究 [D]. 重庆: 西南大学, 2010.

[24] MENG J, LI D, YI T, *et al.* Development and characterization of microsatellite loci for *Rosa odorata* var. *gigantea* Rehder & E. H. Wilson (Rosaceae) [J]. Conservation Genetics, 2009, 10(6): 1973-1976.

[25] MENG J, HE S L, LI D Z, *et al.* Nuclear genetic variation of *Rosa odorata* var. *gigantea* (Rosaceae): population structure and conservation implications [J]. Tree Genetics & Genomes, 2016, 12(4): 65.

[26] GUO X L, YU C, LUO L, *et al.* Comparative transcriptome analysis of the floral transition in *Rosa chinensis* ‘Old Blush’ and *R. odorata* var. *gigantea* [J]. Scientific Reports, 2017, 7(1): 6068.

[27] SUEOKA N. Directional mutation pressure and neutral molecular evolution [J]. Proceedings of the National Academy of Sciences of the United States of America, 1988, 85(8): 2653.

[28] WRIGHT F. The ‘effective number of codons’ used in a gene [J]. Gene, 1990, 87(1): 23.

[29] MCINERNEY J O. Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi* [J]. Proceedings of the National Academy of Sciences of the United States of America, 1998, 95(18): 10698-10703.

[30] SUEOKA N. Two aspects of DNA base composition; G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C. [J]. Journal of Molecular Evolution, 1999, 49(1): 49.

[31] DURET L, MOUCHIROUD D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis* [J]. Proceedings of the National Academy of Sciences of the United States of America, 1999, 96(8): 4482-4487.

[32] 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.

SHANG M Z, LIU F, HUA J P, *et al.* Analysis on codon usage of chloroplast genome of *Gossypium hirsutum* [J]. Scientia Agricultura Sinica, 2011, 44(2): 245-253. (in Chinese)

[33] ZHANG W J, ZHOU J, LI Z F, *et al.* Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L [J]. 植物学报: 英文版, 2007, 49(2): 246-254.

[34] 杨国锋, 苏昆龙, 赵怡然, 等. 藜蓼苜蓿叶绿体密码子偏好性分析 [J]. 草业学报, 2015, 24(12): 171-179.

YANG G F, SU K L, ZHAO Y R, *et al.* Analysis of codon usage in the chloroplast genome of *Medicago truncatula* [J]. Acta Prataculturae Sinica, 2015, 24(12): 171-179. (in Chinese)

[35] 张月荣. 禾本科叶绿体基因组密码子使用模式及紫茎泽兰叶绿体 RNA 编辑分析 [D]. 陕西杨陵: 西北农林科技大学, 2013.

[36] 续晨, 贲爱玲, 蔡晓宁. 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, 2010, 8(5): 945-950.

XU C, BEN A L, CAI X N. Analysis of synonymous codon usage in chloroplast genome of *Phalaenopsis aphrodite* sub sp. *formosana* [J]. Molecular Plant Breeding, 2010, 8(5): 945-950. (in Chinese)

[37] 邢朝斌, 曹蕾, 周秘, 等. 刺五加叶绿体基因组密码子的用法分析 [J]. 中国中药杂志, 2013, 38(5): 661-665.

XING Z B, CAO L, ZHOU M, *et al.* Analysis on codon usage of chloroplast genome of *Eleutherococcus senticosus*. [J]. China Journal of Chinese Materia Medica, 2013, 38(5): 661-665. (in Chinese)

[38] 胡莎莎, 罗洪, 吴琦, 等. 苦荞叶绿体基因组密码子偏爱性分析 [J]. 分子植物育种, 2016, 14(2): 309-317.

HU S S, LUO H, WU Q, *et al.* Analysis of codon bias of chloroplast genome of Tartary Buckwheat [J]. Molecular Plant Breeding, 2016, 14(2): 309-317. (in Chinese)