

## 富民枳叶绿体基因组密码子偏好性分析

徐 荣<sup>1,2</sup>, 宋文政<sup>1,2</sup>, 乔振升<sup>1,2</sup>, 张晓琳<sup>1,2</sup>, 于国栋<sup>3</sup>, 何承忠<sup>1,2</sup>, 纵 丹<sup>1,2</sup>

1. 西南林业大学 云南省高校林木遗传改良与繁育重点实验室, 云南 昆明 650224

2. 西南林业大学 西南山地森林资源保育与利用教育部重点实验室, 云南 昆明 650224

3. 昆明市林业和草原科技推广总站, 云南 昆明 650223

**摘要:** 目的 分析药用植物富民枳 *Poncirus polyandra* 叶绿体基因组蛋白编码基因密码子使用模式及影响因素。方法 以富民枳叶绿体基因组 53 个蛋白编码基因为材料, 利用 CodonW 软件、CUSP 在线程序和 R 语言等整理并分析基因密码子使用偏性参数和核苷酸组成。结果 富民枳叶绿体基因组密码子各位置 GC 含量不同, 偏好以 A/U 结尾; 有效密码子数 (effective number of codon, ENC) 介于 39.79~59.37, 密码子偏好性较弱; 密码子第 3 位碱基 GC<sub>3</sub> 和 ENC 之间呈显著正相关; 影响因素绘图分析结果表明自然选择对密码子使用模式有重要影响, 共确定了 12 个最优密码子, 且均偏向于 A/U 结尾。结论 富民枳叶绿体基因组密码子第 3 位碱基偏向使用 A 或 U 结尾, 密码子偏性受自然选择影响最大, 同时还受到突变等因素的共同影响。

**关键词:** 富民枳; 叶绿体基因组; 密码子偏好性; 最优密码子; 有效密码子数

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2024)20-7093-08

DOI: 10.7501/j.issn.0253-2670.2024.20.024

## Codon preference analysis of chloroplasts genome in medicinal plants of *Poncirus polyandra*

XU Rong<sup>1,2</sup>, SONG Wenzheng<sup>1,2</sup>, QIAO Zhensheng<sup>1,2</sup>, ZHANG Xiaolin<sup>1,2</sup>, YU Guodong<sup>3</sup>, HE Chengzhong<sup>1,2</sup>, ZONG Dan<sup>1,2</sup>

1. Key Laboratory for Forest Genetic and Tree Improvement & Propagation in Universities of Yunnan Province, Southwest Forestry University, Kunming 650224, China

2. Key Laboratory for Forestry Resources Conservation and Utilization in the Southwest Mountains of China, Ministry of Education, Southwest Forestry University, Kunming 650224, China

3. Kunming Extension Station of Forestry and Grassland Science and Technology, Kunming 650223, China

**Abstract: Objective** To investigate codon usage patterns and influencing factors in the chloroplast genome of *Poncirus polyandra*.

**Methods** The 53 protein-coding genes of the *P. polyandra* chloroplast genome were used as materials to organize and analyze the codon usage patterns and nucleotide composition using CodonW software, CUSP online program and R language. **Results** The results showed that the GC content of the codons in the chloroplast genome of *P. polyandra* varied in each position, and the preference ended with A/U; The ENC value ranged from 39.79 to 59.37, and the codon usage bias was weak; There was a significant positive correlation between the third codon base GC<sub>3</sub> and ENC, and the codon usage bias was correlated with the third base. The results of neutral mapping analysis, parity preference analysis and effective codon mapping analysis showed that natural selection had an important influence on codon usage patterns, and the 12 optimal codons further proved that codon usage bias ended in A/U.

**Conclusion** The third codon base of *P. polyandra* chloroplast genome codon is preferred to end in A/U. Codon usage bias is most affected by natural selection and is also affected by a combination of mutations and other factors.

**Key words:** *Poncirus polyandra* S. Q. Ding; chloroplast genome; codon usage bias; optimal codons; effective number of codon

mRNA 上 3 个连续的碱基被称作为密码子, 决定着氨基酸和蛋白质的生成, 在自然界的生物遗传和变异中起着至关重要的作用<sup>[1]</sup>。由于遗传密码的简并性, 2 个及以上的密码子可翻译成同一种氨基

收稿日期: 2024-02-02

基金项目: 昆明市第七批博士后工作扶持站 (992223004)

作者简介: 徐 荣 (1999—), 硕士研究生, 研究方向为生物化学与分子生物学。E-mail: x13529154284@163.com

\*通信作者: 纵 丹 (1988—), 副教授, 硕士生导师, 从事林木遗传育种研究。E-mail: skyzd907@swfu.edu.cn

酸, 此类密码子称为同义密码子。植物的基因在不受到任何的外界条件的干扰下, 同义密码子应该拥有相同的使用频率<sup>[2]</sup>, 但研究发现, 不同的生物对编码同一个氨基酸的多个同义密码子是具有选择性的, 并且在同一物种内不同基因的密码子的使用频率也是存在差异的, 这种现象被称为密码子的使用偏好性, 这类被偏好使用的密码子称为最优密码子<sup>[3]</sup>。密码子的使用偏好性会影响蛋白质翻译的速度和准确度、mRNA 的转录调控和外源基因的表达<sup>[4-5]</sup>, 有研究表明, 使用特定密码子可提高基因表达近 1 000 倍<sup>[6]</sup>, 可见密码子在基因调控、分子育种方面有着重要的作用。

叶绿体是植物光合作用发生的细胞器, 承担着能量转换的重要作用, 其基因组结构较为简单, 但却包含了完整的母系遗传信息, 因此广泛被应用于植物多样性研究、系统发育、DNA 条形码、植物适应性调查和基因工程研究等诸多方面<sup>[7-8]</sup>。叶绿体基因组作为基因工程受体进行植物性状改良可有效的避免位置效应和基因沉默效应<sup>[9]</sup>, 而选用最优密码子设计的叶绿体基因表达载体进行转化, 可提高叶绿体基因组中基因表达量<sup>[10-11]</sup>。因此, 叶绿体基因组密码子使用模式的研究, 有助于了解物种进化历程, 提高叶绿体外源基因的表达量。

近年来, 随着测序技术的发展, 大量植物的叶绿体基因组被公布, 目前 NCBI 上已经发表了 2 000 多个植物叶绿体全基因组, 为叶绿体基因组的密码子偏好性分析提供了基础数据。富民枳 *Poncirus polyandra* S. Q. Ding 作为云南省富民县特有的芸香科 (Rutaceae) 枳属 *Poncirus* L. 的一个种, 是数量极为稀少的云南省柑桔矮化种, 已被中国珍稀濒危植物信息系统 (ISCRPE) 列为国家二级保护野生植物<sup>[10]</sup>, 作为珍贵的柑橘育种种质资源, 其果实可入药, 当地人称“止咳树”, 具有抑制新生血管形成, 降低肿瘤细胞表达, 其果实可做枳壳的代用品入药治疗咳嗽、根部能够治疗风湿<sup>[11-12]</sup>。目前, 对于富民枳的研究主要集中在生物学特性<sup>[13]</sup>、核型分析<sup>[14]</sup>、分类地位及遗传多样性方面<sup>[15-17]</sup>, 但尚未见关于富民枳叶绿体基因组密码子使用模式的研究报道。基于此, 本研究从富民枳种质资源的利用和保护出发, 利用 CodonW 软件、CUSP 在线程序和 R 语言等程序分析富民枳的密码子使用模式, 旨在将密码子使用模式用于基因表达领域, 为后续的富民枳上的性状改良和遗传育种等方面的研究上提供理论依据<sup>[18-19]</sup>。

## 1 材料与仪器

### 1.1 材料

本研究所用富民枳叶片采自昆明市濒危动植物收容拯救中心 (25°3'N, 102°78'E), 经西南林业大学何承忠教授鉴定为富民枳 *P. polyandra* S. Q. Ding, 采集长势良好、无病虫害的新鲜叶片经液氮冷冻后保存在西南山地森林资源保育与利用教育部重点实验室-80 °C 冰箱。

### 1.2 仪器

HFLTP 86 (728E) 型立式低温保存箱 (Heal Force 公司)、UNIVERSAL 320(R)型离心机 (Hettich 公司, 德国)。

## 2 方法

### 2.1 样品处理

通过 Illumina 高通量测序 (北京百迈客生物科技有限公司) 获得高质量数据, 经拼接、组装与注释完成后将富民枳叶绿体基因组序列提交到 NCBI 数据库 (GenBank 登录号 MK250977), 基因组序列全长为 160 211 bp, 根据基因注释获得 91 条可以编码蛋白的基因序列 (coding DNA sequence, CDS), 为保证结果的准确性与可比性, 选取以 ATG 为起始密码子且长度大于 300 bp 的 CDS 序列, 去除重复序列后确定 53 条蛋白编码 CDS 用于后续分析的样本。

### 2.2 密码子相关参数计算

利用 CUSP 在线程序 (<http://www.Bioinformatics.nl/emboss-explorer/>) 计算各基因编码区总 GC 含量 (GC<sub>all</sub>) 以及密码子第 1 位 (GC<sub>1</sub>)、第 2 位 (GC<sub>2</sub>)、第 3 位 (GC<sub>3</sub>) 的 GC 含量。采用 Codon W1.4.2 软件计算相对同义密码子使用度 (RSCU)、密码子适应指数 (CAI) 和有效密码子数 (ENC)。RSCU 是用来表示密码子在编码某一氨基酸的相对使用频率, RSCU 值大于 1 则偏好性高, RSCU 值等于 1 时没有偏好性, RSCU 值小于 1 时偏好性弱<sup>[20-21]</sup>, 通过 Microsoft Excel 和 IBM SPSS21.0 软件对分析数据进行整理分析。

### 2.3 ENC-plot 绘图分析

有效密码子数 (effective number of codon, ENC) 是分析基因整体密码子偏好性水平的重要指标, 高表达基因密码子偏性强, ENC 值较小; 低表达基因使用了较多种稀有的密码子, ENC 值较大。以实际 ENC 值和 GC<sub>3</sub> 值作为 Y 轴和 X 轴绘制二维散点图, 并以预期的 ENC 值绘制标准曲线,  $ENC = 2 + X +$

$29/[X^2+(1-X)^2]$ , 式中  $X$  为  $GC_3$  的值。

基因点越靠近标准曲线, 表明密码子使用偏性受突变作用影响越强; 基因点越远离标准曲线, 表明密码子使用偏性受自然选择的影响程度越强<sup>[22]</sup>。

#### 2.4 PR2-plot 绘图分析

PR2 偏倚分析 (PR2-bias plot analysis) 仅对 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, 避免密码子第 3 位碱基 A/T 和 C/G 之间突变不平衡, 依照 2 个互补的碱基在含量对等时不会发生选择以及突变的偏倚规则<sup>[23]</sup>。以  $A_3/(A_3+T_3)$  作为  $Y$  轴,  $G_3/(G_3+C_3)$  作为  $X$  轴绘制散点图, 中心点 ( $A=T$  且  $C=G$ ) 代表无偏性使用, 各点与中心的位置关系代表基因的偏倚方向及程度<sup>[23]</sup>。

#### 2.5 中性绘图分析

中性绘图分析是衡量密码子偏性影响因素的方法之一, 以  $GC_{12}$  ( $GC_1$  与  $GC_2$  的平均值为  $Y$  轴, 各基因的  $GC_3$  值为  $X$  轴绘制散点图。若相关度高, 则全部的点会沿着对角线分布, 表示各个位置碱基的使用模式相似, 密码子的使用模式主要与突变因素相关; 反之, 随机分布的点表示碱基的使用存在差异, 样本中  $GC$  含量保守性强, 表明自然选择对密码子的使用模式影响更大<sup>[24]</sup>。

#### 2.6 最优密码子确定

根据 ENC 值大小, 将富民枳 53 个候选基因中 ENC 值最高 (10%) 和最低 (10%) 的作为高表达组 (5 个) 和低表达组 (5 个), 分别计算高表达组和低表达组的 RSCU, 并计算高低组的 RSCU 差值 ( $\Delta RSCU$ ), 以  $\Delta RSCU$  大于 0.08 的密码子作为高表达密码子; 将高表达密码子与高频密码子 ( $RSCU > 1$ ) 取交集, 共有的密码子作为最优密码子。

### 3 结果与分析

#### 3.1 富民枳叶绿体基因组密码子特征分析

富民枳叶绿体基因组密码子不同位置  $GC$  含量分析结果显示 (表 1), 各基因  $GC_{all}$  含量在 33.94%~49.35%, 平均值为 39.73%,  $GC_1$ 、 $GC_2$  和  $GC_3$  含量分别在 34.28%~58.19%, 29.36%~60.43%, 21.82%~51.94%, 平均  $GC$  含量呈现为  $GC_1$  (47.41%)  $> GC_2$  (40.18%)  $> GC_3$  (31.59%), 且均小于 50%, 表明密码子不同位置碱基  $GC$  含量存在差异, 第 3 位碱基的组成与前 2 位碱基差异大, 密码子相对保守, 且偏好以 A/U 结尾。

富民枳叶绿体基因组 ENC 介于 39.79(*rps18*)~

59.37 (*clpP*), 均值为 49.83, 53 个基因中仅 10 个基因 ENC 值小于 45, 43 个基因 ENC 值大于 45, 说明富民枳叶绿体基因组密码子偏好性较弱。富民枳叶绿体基因组蛋白编码基因序列 CAI 取值范围在 0.10~0.32, 均值为 0.17, 表明其基因表达水平不高。

富民枳叶绿体基因密码子各位置  $GC$  含量以及有效密码子等参数的相关性分析结果见表 2。  $GC_{all}$  与  $GC_1$ 、 $GC_2$ 、 $GC_3$  在 0.01 水平呈极显著正相关, 而  $GC_1$  与  $GC_2$  呈显著正相关, 但是  $GC_3$  与  $GC_1$ 、 $GC_2$  之间相关性不显著, ENC 与  $GC_3$  呈显著正相关, 但是与  $GC_1$  和  $GC_2$  相关性不显著, 表明富民枳叶绿体基因组密码子的使用偏好性主要受到  $GC_3$  的影响而受  $GC_1$  和  $GC_2$  的影响不是很大。

#### 3.2 中性绘图分析

富民枳叶绿体基因组中性绘图结果显示 (图 1), 大部分基因分布于对角线上方, 仅 2 条基因位于对角线下方,  $GC_3$  分布范围在 26.47~31.59,  $GC_{12}$  分布范围在 40.69~43.80,  $GC_3$  与  $GC_{12}$  的相关系数为 0.124 0, 回归系数为 0.121 0, 表明  $GC_{12}$  和  $GC_3$  相关性较弱, 内部突变贡献率占 12.10%, 自然选择贡献率为 87.90%, 表明自然选择对富民枳叶绿体基因组密码子偏好性的作用强度大于内部突变压力。

#### 3.3 ENC-plot 分析

ENC-plot 分析结果见图 2。只有 6 个基因在预期 ENC 曲线附近 (11.3%), 偏性较弱, 剩余 47 个基因与理论值有所偏移 (88.7%), 大多数基因 ENC 值低于预期值, 位于曲线下方, 偏性较强, 表明自然选择对富民枳叶绿体密码子的偏好性的影响大于内部突变压力。

#### 3.4 PR2-plot 分析

奇偶偏好性分析结果见图 3。各编码基因不均匀地分布于平面图的 4 个区域内, 有的点距离轴较远, 平面图右下方基因分布较多,  $G_3/(G_3+C_3)$  大于 0.5 的点有 30 个,  $A_3/(A_3+T_3)$  小于 0.5 的点有 37 个, 同时满足的有 21 个点, 说明富民枳叶绿体基因组密码子第 3 位碱基使用频率  $U > A$ 、 $G > C$ , 证明富民枳密码子使用模式受到内部突变压力影响之外, 还受到自然选择的干预。

#### 3.5 最优密码子的确定

去掉 3 个不编码氨基酸的终止密码子以及有唯一的密码子的色氨酸 (Trp)、蛋氨酸 (Met), 对高、低偏性库各个密码子的 RSCU 和  $\Delta RSCU$  计算

表1 富民枳叶绿体基因组密码子各位置 GC 含量

Table 1 GC content of different positions of codon in chloroplast genome of *P. polyandra*

基因	GC/%					ENC	CAI
	GC <sub>1</sub>	GC <sub>2</sub>	GC <sub>3</sub>	GC <sub>12</sub>	GC <sub>all</sub>		
<i>rps18</i>	38.24	43.14	26.47	40.69	35.95	39.79	0.10
<i>ndhI</i>	44.64	36.90	22.02	40.77	34.52	40.77	0.21
<i>psbA</i>	55.91	39.37	29.92	47.64	41.73	40.88	0.32
<i>atpF</i>	45.95	32.43	31.35	39.19	36.58	41.26	0.13
<i>rps14</i>	34.65	47.52	43.56	41.09	41.91	41.43	0.14
<i>petD</i>	50.93	39.75	27.95	45.34	39.54	42.74	0.16
<i>rps7</i>	53.21	44.87	23.72	49.04	40.60	44.23	0.19
<i>ccsA</i>	34.28	39.31	28.93	36.80	34.17	44.30	0.13
<i>rpl16</i>	48.18	51.09	31.39	49.64	43.55	44.31	0.13
<i>rps12</i>	51.61	48.39	30.65	50.00	43.55	44.88	0.15
<i>rps8</i>	42.34	40.15	33.58	41.25	38.69	45.28	0.12
<i>accD</i>	39.92	37.50	29.03	38.71	35.48	45.93	0.20
<i>psbC</i>	53.80	45.99	32.28	49.90	44.02	46.57	0.21
<i>ndhA</i>	44.51	39.29	25.00	41.90	36.26	46.73	0.13
<i>ndhF</i>	38.71	36.29	27.55	37.50	34.18	47.48	0.14
<i>rpl20</i>	38.98	43.22	27.12	41.10	36.44	48.13	0.10
<i>ndhC</i>	47.11	32.23	29.75	39.67	36.36	48.60	0.20
<i>psbD</i>	52.54	43.22	35.03	47.88	43.60	48.75	0.24
<i>atpB</i>	56.71	41.68	32.87	49.20	43.75	48.87	0.19
<i>psbB</i>	54.81	46.17	31.43	50.49	44.14	49.14	0.19
<i>atpA</i>	55.91	39.37	29.92	47.64	41.73	49.28	0.20
<i>matK</i>	42.94	31.76	32.16	37.35	35.62	49.39	0.16
<i>ndhE</i>	39.22	35.29	32.35	37.26	35.62	49.72	0.15
<i>petB</i>	47.23	40.85	36.17	44.04	41.42	49.82	0.21
<i>rpl14</i>	43.56	47.52	34.65	45.54	41.91	49.85	0.16
<i>ndhB</i>	42.07	38.55	32.29	40.31	37.64	50.13	0.17
<i>rps3</i>	46.82	33.18	21.82	40.00	33.94	50.24	0.14
<i>atpI</i>	49.60	37.50	30.24	43.55	39.11	50.38	0.16
<i>ndhG</i>	45.25	34.64	27.93	39.95	35.94	50.43	0.15
<i>psaB</i>	48.98	43.13	34.29	46.06	42.13	50.44	0.18
<i>rpoC1</i>	51.10	38.33	29.07	44.72	39.50	50.81	0.15
<i>ndhH</i>	51.01	36.62	28.03	43.82	38.55	50.82	0.18
<i>ndhD</i>	42.40	37.60	33.20	40.00	37.73	50.85	0.13
<i>rpoB</i>	50.79	37.91	30.63	44.35	39.78	50.92	0.16
<i>psaA</i>	53.00	43.28	34.89	48.14	43.72	50.96	0.19
<i>rps4</i>	51.49	37.62	26.24	44.56	38.45	51.52	0.15
<i>rbcL</i>	58.19	43.91	33.40	51.05	45.17	51.88	0.25
<i>ycf68</i>	48.06	48.06	51.94	48.06	49.35	52.10	0.14
<i>petA</i>	52.02	37.07	33.02	44.55	40.71	52.23	0.18
<i>rpoC2</i>	47.65	37.89	32.12	42.77	39.22	52.45	0.15
<i>ycf2</i>	42.11	34.95	37.31	38.53	38.12	53.58	0.16
<i>ndhK</i>	42.54	45.61	29.39	44.08	39.18	53.59	0.17
<i>atpE</i>	48.51	39.55	30.60	44.03	39.55	53.76	0.16
<i>rpoA</i>	45.59	33.43	32.22	39.51	37.08	53.92	0.17
<i>rps2</i>	47.26	43.46	31.22	45.36	40.65	54.29	0.16
<i>rpl2</i>	51.64	49.09	33.09	50.37	44.61	55.29	0.14
<i>rps11</i>	56.83	60.43	30.22	58.63	49.16	55.42	0.15
<i>ycf4</i>	43.78	40.00	35.68	41.89	39.82	55.62	0.16
<i>ndhJ</i>	52.83	37.11	33.33	44.97	41.09	57.36	0.16
<i>ycf3</i>	49.11	37.87	31.36	43.49	39.45	58.10	0.17
<i>cemA</i>	41.70	29.36	32.34	35.53	34.47	58.15	0.20
<i>ycf1</i>	40.93	31.04	35.44	35.99	35.81	58.34	0.18
<i>clpP</i>	55.84	39.09	38.07	47.47	44.33	59.37	0.16
平均值	47.41	40.18	31.59	43.80	39.73	49.83	0.17

表 2 富民枳叶绿体基因组密码子各参数相关性分析

Table 2 Correlation analysis of chloroplast genomic codons in *P. polyandra*

因素	相关系数			
	GC <sub>1</sub>	GC <sub>2</sub>	GC <sub>3</sub>	GC <sub>all</sub>
GC <sub>2</sub>	0.304*			
GC <sub>3</sub>	-0.110	0.215		
GC <sub>all</sub>	0.686**	0.773**	0.539**	
ENC	0.204	-1.270	0.302*	0.174

\*表示相关性显著 ( $P < 0.05$ ), \*\*表示相关性极显著 ( $P < 0.01$ ).  
\*indicates significant correlation ( $P < 0.05$ ) and \*\*indicates highly significant correlation ( $P < 0.01$ ).

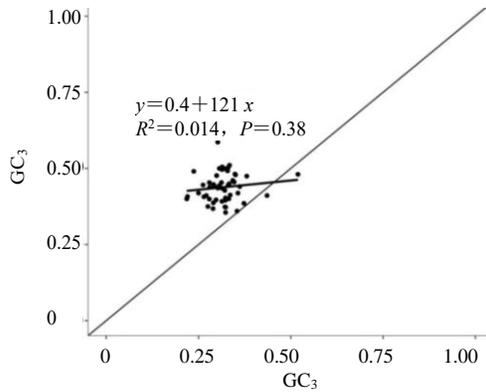


图 1 富民枳叶绿体基因组中性图分析

Fig. 1 Neutral map analysis of the chloroplast genome of *P. polyandra*

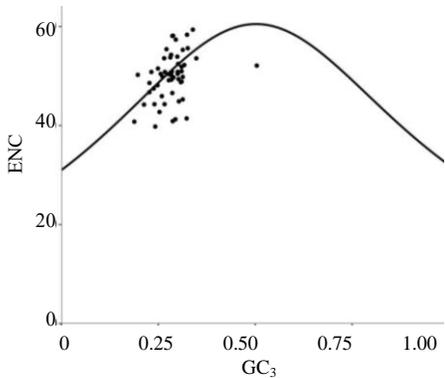


图 2 富民枳 ENC-plot 分析

Fig. 2 ENC-plot analysis of *P. polyandra*

结果见表 3。由表 3 可知, 有 26 个  $\Delta RSCU \geq 0.08$  的高表达优越密码子,  $RSCU > 1$  的高频密码子有 23 个。将二者结合, 同时满足高表达优越密码子和高频密码子条件的为最优密码子, 12 个, 分别为 UUA、AUU、GUU、GUA、UCU、AGU、CCU、ACU、GCU、CAA、AAA、UGU。其中 12 个最优

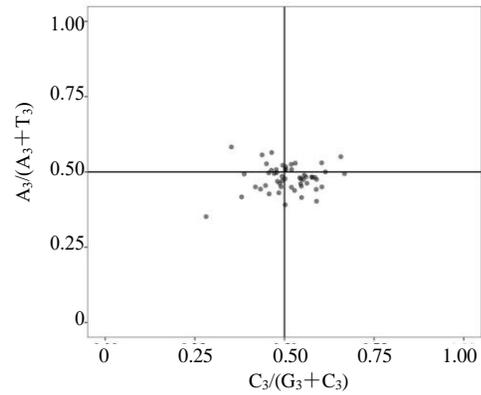


图 3 富民枳 PR2-plot 分析

Fig. 3 PR2-plot analysis of *P. polyandra*

密码子中 8 个以 U 结尾, 4 个以 A 结尾, 进一步说明了富民枳叶绿体基因组密码子偏好 A/U 结尾。

#### 4 讨论

植物叶绿体基因组高度保守, 对其叶绿体基因组密码子偏好及影响因素研究有助于理解物种起源及进化。富民枳作为典型的极小种群野生植物, 近年来其药用价值不断得到研究重视, 本研究对富民枳叶绿体基因组偏好性及影响因素进行了相关的分析。43 条 CDS (81%) 均显示  $ENC > 45$ , 说明富民枳叶绿体基因组密码子的偏好性较弱, 密码子第 3 位碱基偏好以 A/U 结尾。

密码子相关参数的分析中发现  $GC_{all}$  的平均值是 39.73%; GC 在密码子 3 个位置的含量分别是:  $GC_1$  (47.41%)  $> GC_2$  (40.18%)  $> GC_3$  (31.59%), 表明密码子位置分布不均且密码子偏好以 A/U 结尾, 这与双子叶植物偏好使用 A/U 结尾, 以及在叶绿体基因组密码子普遍偏好以 A/T 结尾<sup>[25-27]</sup>的结果是一致的。

如果密码子第 3 位碱基受到的自然选择压力较小, 不会影响与氨基酸的对应关系, 因此对研究密码子偏好性具有重要意义<sup>[23]</sup>。从  $GC_1$ 、 $GC_2$ 、 $GC_3$ 、与 ENC 的相关性这个层面上来看,  $GC_{all}$  与  $GC_1$ 、 $GC_2$ 、 $GC_3$  相关性呈极显著, 并且  $GC_1$  与  $GC_2$  的相关性呈显著, 但是  $GC_3$  与  $GC_1$ 、 $GC_2$  之间的相关性呈不显著, ENC 与  $GC_3$  相关性呈显著, 但是与  $GC_1$ 、 $GC_2$  相关性呈不显著, 这表明了富民枳叶绿体基因组密码子中第 1、2 位碱基的差异并不是很大, 但是前 2 位与第 3 位的差异是有的, 结合与 ENC 的相关性分析来看, ENC 值与  $GC_3$  显著相关, 但与  $GC_1$ 、 $GC_2$  相关性不显著, 这表明了富民枳叶绿体基因组

表 3 富民枳叶绿体基因组最优密码子

Table 3 Optimal codons for chloroplast genome of *P. polyandra*

氨基酸	密码子	高表达基因		低表达基因		ΔRSCU
		数目	RSCU	数目	RSCU	
苯丙氨酸	UUU	21	0.93	48	1.35	-0.42
	UUC	24	1.07	23	0.65	0.42
亮氨酸	UUA	27	2.00	32	1.63	0.37
	UUG	21	1.56	15	0.76	0.80
	CUU	16	1.19	27	1.37	-0.18
	CUC	2	0.15	13	0.66	-0.51
	CUA	10	0.74	17	0.86	-0.12
	CUG	5	0.37	14	0.71	-0.34
异亮氨酸	AUU	45	1.80	39	1.29	0.51
	AUC	18	0.72	22	0.73	-0.01
	AUA	12	0.48	30	0.99	-0.51
甲硫氨酸	AUG	25	1.00	34	1.00	0
缬氨酸	GUU	16	1.52	24	1.41	0.11
	GUC	2	0.19	9	0.53	-0.34
	GUA	22	2.10	27	1.59	0.51
	GUG	2	0.19	8	0.47	-0.28
丝氨酸	UCU	17	1.70	21	1.62	0.08
	UCC	8	0.80	12	0.92	-0.12
	UCA	9	0.90	13	1.00	-0.10
	UCG	4	0.40	12	0.92	-0.52
	AGC	5	0.50	5	0.38	0.12
	AGU	17	1.70	15	1.15	0.55
	CCU	21	2.21	15	1.30	0.91
脯氨酸	CCC	3	0.32	14	1.22	-0.90
	CCA	9	0.95	9	0.78	0.17
	CCG	5	0.53	8	0.70	-0.17
	ACU	20	1.78	17	1.31	0.47
苏氨酸	ACC	13	1.16	10	0.77	0.39
	ACA	11	0.98	18	1.38	-0.40
	ACG	1	0.09	7	0.54	-0.45
	GCU	38	2.34	21	1.56	0.78
丙氨酸	GCC	4	0.25	14	1.04	-0.79
	GCA	15	0.92	11	0.81	0.11
	GCG	8	0.49	8	0.59	-0.10
	UAU	15	1.25	47	1.68	-0.43
酪氨酸	UAC	9	0.75	9	0.32	0.43
	UAA	4	2.40	3	1.80	0.60
终止密码子	UAG	1	0.60	0	0	0.60
	CAU	10	1.05	20	1.43	-0.38
组氨酸	CAC	9	0.95	8	0.57	0.38
	CAA	25	1.79	29	1.49	0.30
谷氨酰胺	CAG	3	0.21	10	0.51	-0.30
	AAU	34	1.39	41	1.49	-0.10
天冬酰胺	AAC	15	0.61	14	0.51	0.10
	AAA	23	1.53	33	1.18	0.35
赖氨酸	AAG	7	0.47	23	0.82	-0.35
	GAU	19	1.46	32	1.64	-0.18
天冬氨酸	GAC	7	0.54	7	0.36	0.18
	GAA	36	1.50	65	1.55	-0.05
谷氨酸	GAG	12	0.50	19	0.45	0.05
	UGU	12	1.71	6	1.00	0.71
半胱氨酸	UGC	2	0.29	6	1.00	-0.71
	UGA	0	0	2	1.20	-1.20
终止密码子	UGG	15	1.00	24	1.00	0
	CGU	21	1.80	9	1.02	0.78
色氨酸	CGC	5	0.43	3	0.34	0.09
	CGA	18	1.54	16	1.81	-0.27
	CGG	2	0.17	6	0.68	-0.51
	AGA	15	1.29	14	1.58	-0.29
	AGG	9	0.77	5	0.57	0.20
	GGU	33	2.40	14	0.92	1.48
甘氨酸	GGC	6	0.44	10	0.66	-0.22
	GGA	11	0.80	23	1.51	-0.71
	GGG	5	0.36	14	0.92	-0.56

下划线为最优密码子。

Glid line is the optimal codon.

密码子的偏好性主要是与密码子的第3位碱基的变化有关,这与柔毛郁金香 *Tulipa buhseana* Pall.<sup>[28]</sup>和梔子 *Gardenia jasminoides* Ellis.<sup>[29]</sup>的研究结果一致。

PR2-plot 分析的结果来看富民枳叶绿体基因组密码子第3位碱基使用频率呈现 U>A、G>C 的分布,证明富民枳密码子使用模式受到内部突变压力影响之外,还受到自然选择的干预。ENc-plot 分析结果表明 47 个基因与理论值有所偏移 (88.7%),表明自然选择对富民枳叶绿体密码子的偏好性的影响大于内部突变压力。中性绘图分析显示大部分点落在对角线上方,仅 2 个点沿对角线分布,且 GC<sub>3</sub> 与 GC<sub>12</sub> 的相关系数为 0.124 0,再次证明自然选择是叶绿体基因组进化的主要驱动力。在其它研究中也得到了相似的结果,如五节芒 *Miscanthus floridulus* (Lab.) Warb. ex Schum. et Laut.<sup>[30]</sup>翠雀 *Delphinium grandiflorum* L.<sup>[31]</sup>和刺榆 *Hemiptelea davidii* (Hance) Planch.<sup>[32]</sup>。

进行基因表达载体的设计,提高目的基因的表达到量,离不开最优密码子的确定<sup>[33]</sup>。本研究从富民枳叶绿体基因组中筛选了 53 条 CDS 序列,以 ENC 值从小到大选择前 10%和后 10%建立高偏性库和低偏性库,筛选出来了 12 个最优密码子,其中 NNA 型有 4 个、NNU 型有 8 个,最优密码子的结尾全是 U/A,再次印证了富民枳偏好选择 A/U 结尾的密码子的结论。

本研究首次基于叶绿体基因组密码子偏好性对富民枳进行了分析,系统性的阐明了影响密码子偏好性的因素,揭示了影响其密码子偏好性的因素与自然选择压力相关,筛选出了最优密码子,为后续的富民枳基因工程研究、资源保护利用及药用价值的开发等提供了科学依据和理论支撑。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] 夏晞,彭劲谕,王大玮,等. 3 种榕属叶绿体基因组密码子偏好性分析 [J]. 西北林学院学报, 2022, 37(5): 88-94.
- [2] 王超,赵懿琛,赵德刚. 朝仓花椒转录组密码子偏好性分析 [J]. 分子植物育种, 2020, 18(21): 6989-6996.
- [3] Grantham R, Gautier C, Gouy M. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type [J]. *Nucleic Acids Res*, 1980, 8(9): 1893-1912.
- [4] Wang Z J, Xu B B, Li B, et al. Comparative analysis of codon usage patterns in chloroplast genomes of six *Euphorbiaceae* species [J]. *PeerJ*, 2020, 8: e8251.
- [5] 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]. *Mol Biol Evol*, 2017, 34(11): 2944-2958.
- [6] Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression [J]. *Trends Biotechnol*, 2004, 22(7): 346-353.
- [7] Asaf S, Waqas M, Khan A L, et al. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species [J]. *Front Plant Sci*, 2017, 8: 304.
- [8] Abdullah, Mehmood F, Shahzadi I, et al. Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): Comparative analyses and identification of mutational hotspots [J]. *Genomics*, 2020, 112(1): 581-591.
- [9] 赵森,邓力华,陈芬. 秋茄叶绿体基因组密码子使用偏好性分析 [J]. 森林与环境学报, 2020, 40(5): 534-541.
- [10] 李革,欧阳志勤,祁云,等. 昆明地区珍稀濒危植物及其保护对策 [J]. 环境科学导刊, 2010, 29(1): 27-29.
- [11] 高雪松,张玉萍,唐金武,等. 浅谈富民枳生存现状与保护发展 [J]. 防护林科技, 2017(12): 83-85.
- [12] 黎发宏,高雪松,董锦润,等. 苗药富民枳不同组织挥发性化学成份 GC-MS 分析 [J]. 天然产物研究与开发, 2021, 33(增刊): 43-58.
- [13] 钟瑞芳,王仕玉,梁明清,等. 富民枳的生物学特性及利用研究初报 [J]. 云南农业大学学报, 2001, 16(3): 244-246.
- [14] 刘利勤,杨静,顾志建. 富民枳的核型与分类位置探讨 [J]. 云南植物研究, 2007, 29(2): 198-200.
- [15] 龚桂芝,洪棋斌,彭祝春. 枳属种质遗传性及其与近缘属植物亲缘关系的 SSR 和 cpSSR 分析 [J]. 园艺学报, 2008, 35(12): 1742-1750.
- [16] 张珊珊,杨文忠,张玉萍,等. 基于 SSR 分析的富民枳样本量对其遗传多样性指标的影响 [J]. 东北林业大学学报, 2017, 45(9): 35-39.
- [17] Ji Y M, Zhang W J, Li D, et al. The complete chloroplast genome sequence of *Annamocarya sinensis* (Juglandaceae), an endangered species endemic to Yunnan Province, China [J]. *Mitochondrial DNA B Resour*, 2020, 5(3): 2021-2023.
- [18] 冯茜莉,王慧慧,汪梦竹,等. 同义密码子通过精微翻译选择机制实现对基因的表达调控 [J]. 微生物学报, 2022, 62(10): 3681-3695.
- [19] 高灿,樊智丰,马长乐,等. 菠萝蜜叶绿体基因组密码子偏好性分析 [J]. 福建农林大学学报: 自然科学版, 2023, 52(6): 776-784.

- [20] 张东旭, 孔丽丽. 连香树叶绿体基因组测序与特征分析 [J]. 分子植物育种, 2023, 21(17): 5654-5662.
- [21] 韩慧杰, 邱锐, 刘叶飞, 等. 5种冰草叶绿体基因组密码子偏好性分析 [J]. 分子植物育种, 2023, 21(18): 6010-6018.
- [22] 李路丽, 张萍, 曾强, 等. 芝麻叶绿体基因组的密码子使用偏性及其影响因素 [J]. 分子植物育种, 2023, 21(4): 4535-4544.
- [23] 王晓双, 杨芳, 罗茜, 等. 石斛属叶绿体基因组密码子使用偏性及系统发育分析 [J]. 分子植物育种, 2022, 20(23): 7759-7769.
- [24] 李江飞, 李熙颜, 王瑜, 等. 滇楸叶绿体基因组密码子偏好性分析 [J]. 基因组学与应用生物学, 2022, 41(4): 843-853.
- [25] 简雪艳, 杨菊. 木通叶绿体基因组密码子偏好性分析 [J]. 分子植物育种, 2023, 9: 1-12.
- [26] 原晓龙, 陈伟, 王毅, 等. 思茅松叶绿体基因组密码子偏好性分析 [J]. 基因组学与应用生物学, 2021, 40(3): 1278-1283.
- [27] 王占军, 丁亮, 蔡倩文, 等. 3种木薯全基因组的密码子偏好性模式与变异来源比较 [J]. 应用与环境生物学报, 2021, 27(4): 1013-1021.
- [28] 秦斗文, 徐庭亮, 闫京艳, 等. 柔毛郁金香叶绿体基因组密码子偏好性分析 [J]. 江苏农业科学, 2023, 51(22): 41-47.
- [29] 覃兴化, 张麒功, 汪其双, 等. 梔子叶绿体基因组密码子偏好性分析 [J]. 福建农林大学学报: 自然科学版, 2023, 52(5): 632-639.
- [30] Sheng J J, She X, Liu X Y, *et al.* Comparative analysis of codon usage patterns in chloroplast genomes of five *Miscanthus* species and related species [J]. *Peer J*, 2021, 9: e12173.
- [31] Duan H R, Zhang Q, Wang C M, *et al.* Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints [J]. *Peer J*, 2021, 9: e10787.
- [32] Liu H B, Lu Y Z, Lan B L, *et al.* Codon usage by chloroplast gene is bias in *Hemiptelea davidii* [J]. *J Genet*, 2020, 99: 8.
- [33] 吴宪明, 吴松锋, 任大明, 等. 密码子偏性的分析方法及相关研究进展 [J]. 遗传, 2007, 29(4): 420-426.

[责任编辑 时圣明]