

柳燕杰,田旭平,李 倩. 美国红栲叶绿体基因组密码子偏好性分析[J]. 江苏农业科学,2020,48(15):83-88.
doi:10.15889/j.issn.1002-1302.2020.15.014

美国红栲叶绿体基因组密码子偏好性分析

柳燕杰,田旭平,李 倩

(山西农业大学林学院,山西太谷 030801)

摘要:为了提高基因的表达效率,利用叶绿体基因工程提高美国红栲的重要特性,利用 Codon W 1.4.2 和在线软件 CUSP 分析了美国红栲叶绿体基因组中的 52 条基因编码序列密码子偏好性。结果表明,美国红栲叶绿体基因组密码子的 GC 含量依次为 $GC_1(45.23\%) > GC_2(39.23\%) > GC_3(26.19\%)$;有效密码子数(ENC)范围为 37.55~55.28,其中 ENC 值 >45 的有 34 个;RSCU >1 的密码子有 29 个,其中 14 个以 U 结尾、12 个以 A 结尾,表明其偏好以 A、U 结尾,且偏倚很弱。中性点图分析表明, GC_{12} 与 GC_3 的相关系数为 0.321 7,回归系数为 -0.538 5,相关性不显著;美国红栲叶绿体基因组的 GC 含量是高度保守的,密码子偏好主要受环境选择的影响;17 个密码子被确定最优密码子。本研究为美国红栲叶绿体遗传工程和遗传多样性分析提供了科学依据。

关键词:美国红栲;叶绿体基因组;密码子偏好性;选择

中图分类号:S718.43 **文献标志码:**A **文章编号:**1002-1302(2020)15-0083-06

美国红栲 (*Fraxinus pennsylvanica*) 是木犀科 (Oleaceae) 栲属 (*Fraxinus*) 乔木,原产美国,雌雄异株,花先叶开放,喜光、抗寒、抗盐碱、抗水湿,是我国重要的行道树或庭园绿化树种^[1]。

在生物体传递遗传信息的过程中,作为联结核酸和蛋白质的密码子扮演着重要的角色^[2],密码子被称为第二套遗传密码^[3];密码子使用的选择方式不仅影响基因的表达^[4],也影响基因相应的功能^[5]。构成基因组的 4 种核苷酸可形成 64 种密码子,各密码子与氨基酸相对应,除甲硫氨酸和色氨酸外,其余 18 种氨基酸均有 2~6 个密码子,这些编码同一氨基酸的不同密码子被称为同义密码子 (synonymous codon)^[6];在翻译过程中,每个氨基酸相对应同义密码子的使用频率存在差异,即有的同

收稿日期:2020-02-17

基金项目:山西农业大学动植物育种基金(编号:2019yz004)。

作者简介:柳燕杰(1995—),男,山西吕梁人,硕士研究生,主要从事园林植物与观赏园艺。

通信作者:田旭平,副教授,研究生导师,主要从事园林植物与景观设计。E-mail:txp8638@sina.com。

[3] Singh N, Arunkumar A, Peck M, et al. Development of adsorptive hybrid filters to enable two-step purification of biologics [J]. Mabs - Austin, 2017, 9(2): 350-363.

[4] Chiang M J, Pagkaliwangan M, Lute S, et al. Validation and optimization of viral clearance in a downstream continuous chromatography setting [J]. Biotechnology and Bioengineering, 2019, 116(9): 2292-2302.

[5] Angelo J, Chollangi S, Müller - Späth T, et al. Virus clearance validation across continuous capture chromatography [J]. Biotechnology and Bioengineering, 2019, 116(9): 2275-2284.

[6] David L, Bayer M P, Lobedann M, et al. Simulation of continuous low pH viral inactivation inside a coiled flow inverter [J]. Biotechnology and Bioengineering, 2020, 117(4): 1048-1062.

[7] 国家食品药品监督管理局药品审评中心. 生物组织提取制品和真核细胞表达制品的病毒安全性评价技术审评一般原则 [S]. 2005

[8] 韦 薇,徐隆昌,白玉,等. 重组表达生物制品病毒安全性研究与评价的考虑 [J]. 中国新药杂志, 2019, 28(16): 1964-

1968.

[9] 郑丰平,王 炳,郑 琪,等. DV50 纳米过滤去除静注人免疫球蛋白中的病毒时流速的影响因素 [J]. 中国生物制品学杂志, 2012, 25(12): 1712-1713, 1718.

[10] Johnston A, Uren E, Johnston D, et al. Low pH, caprylate incubation as a second viral inactivation step in the manufacture of albumin parametric and validation studies [J]. Biologicals, 2003, 31(3): 213-221.

[11] Buchacher A, Iberer G. Purification of intravenous immunoglobulin G from human plasma - aspects of yield and virus safety [J]. Biotechnology Journal, 2006, 1(2): 148-163.

[12] Lebing W, Remington K M, Schreiner C, et al. Properties of a new intravenous immuneoglobulin (IGIV - C, 10%) produced by virus inactivation with caprylate and column chromatography [J]. Vox Sanguinis, 2003, 84(3): 193-201.

[13] Jeong E K, Sung H M, Kim I S. Inactivation and removal of influenza A virus H1N1 during the manufacture of plasma derivatives [J]. Biologicals, 2010, 38(6): 652-657.

义密码子使用频率高于其他同义密码子,这种现象被称为密码子偏好性(codon usage bias)^[7]。密码子偏好性广泛存在于不同生物中,是因为物种在长期进化过程中受环境选择、碱基突变、基因漂变等因素共同作用,还受到基因组大小、tRNA 丰度和基因表达水平等的影响^[8]。密码子偏好性通过对基因翻译准确性和效率的调节影响基因的表达水平^[9],叶绿体具有基因组小、基因拷贝数多等特点^[10-11];根据最优密码子设计叶绿体基因表达载体,可迅速提高叶绿体基因组中基因表达量,利用目前已知的密码子使用模式推断未知基因的表达,或预测某些未知基因的功能^[12];同时亲缘关系较近的物种具有相近的密码子使用模式^[13],因此,研究叶绿体基因组密码子的使用模式,对于探索物种进化、提高外源基因的表达水平具有重要意义。

与核基因组对比,关于叶绿体基因组的密码子偏性的研究相对滞后,美国红栲的叶绿体基因组测序工作已经完成,但有关密码子偏性的研究迄今尚无报道。研究美国红栲叶绿体密码子的偏性,对于预测基因的表达水平、确定未知基因的位置和改良外源基因有着重要意义。在本研究中,笔者通过对美国红栲叶绿体基因组编码 DNA 序列(CDS)的碱基组成和中性绘图分析,推断了影响美国红栲叶绿体密码子偏好性的主要因素,并确定了美国红栲叶绿体基因组的最优密码子。本研究通过分析美国红栲叶绿体基因组密码子使用模式及影响密码子使用偏好的因素,确定美国红栲叶绿体基因组的密码子偏好性和最优密码子,为美国红栲叶绿体基因组的应用和研究提供科学参考。

1 材料与方法

1.1 材料

美国红栲(*Fraxinus pennsylvanica*)叶绿体基因组序列来自 NCBI 数据库,共下载到 82 条 CDS(coding DNA sequence),剔除 7 条重复基因序列和 23 条长度小于 300 bp 的序列^[14],其余 52 条序列用于本研究的分析。

1.2 方法

在叶绿体通用密码子表中有 64 个密码子,TAA、TGA 和 TAG 是终止密码子,不编码任何氨基酸。ATG 是蛋氨酸的唯一密码子,TGG 是色氨酸的唯一密码子,这些密码子没有偏性,在分析时被剔除。

1.2.1 密码子组成分析 将 50 条符合条件的 CDS 整合为 1 个.fasta 文件中,应用 Codon W 1.4.2 软件分析获得各 CDS 的有效密码子数(effective number of codon,ENC)和同义密码子相对使用度(RSCU),并应用在线软件 CUSP(<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>)分析测定密码子第 1、第 2、第 3 位碱基的 GC 含量(分别为 GC₁、GC₂、GC₃)和 3 位碱基的 GC 平均含量(GC_{all})等参数,结果用 SPSS 和 Excel 进行统计分析。有效密码子数(ENC)是衡量同义密码子使用偏好的重要指标,ENC 的取值范围为 20~61,ENC 值可以反映密码子偏倚的强弱,当 ENC 为 20 时,同义密码子完全处于偏倚状态;当 ENC 值为 61 时,同义密码子没有偏倚;ENC 值从小到大表示偏倚性由强变弱,通常可以 ENC 值 45 作为区分偏倚性强弱的标准^[15]。RSCU 是一个密码子的实际使用频率与无使用偏性时理论频率的比率,无偏倚时,RSCU 为 1;RSCU 小于 1 则代表该密码子的实际使用频率低于其他同义密码子的使用频率,反之实际频率高于其他同义密码子的使用频率^[16]。用 SPSS 软件对不同密码子位置的 GC₁、GC₂、GC₃、GC_{all}进行统计,分析密码子数(N)与 ENC 的相关关系,进而判断各因素对密码子偏倚的影响。

1.2.2 中性绘图分析 简并密码子第 3 碱基通常发生的为同义突变,而简并密码子第 1 位、第 2 位上突变通常会改变基因的功能或活性;也就是说,当没有外压时,密码子 3 个位置的碱基组成应该没有差别;而在存在一定选择压力情况下时,密码子 3 个位置上的碱基组成是存在差异的^[17]。在以 GC₁ 和 GC₂ 的平均值 GC₁₂ 和 GC₃ 分别作为纵坐标和横坐标的中性绘图中,每个散点代表一个基因。如果中性图中的所有基因都沿对角线分布,即 GC₁₂ 和 GC₃ 的变异基本相同,则密码子 3 个位置的碱基组成无显著差异,选择压力弱,但受突变影响较大;回归系数(对角线斜率)是衡量中性程度的指标之一,若回归曲线斜率极小,GC₁₂ 和 GC₃ 的变异的相关性同样很低,说明影响密码子偏好性的主要影响因素为选择效应^[17]。同时,通过分析不同位置密码子碱基组成的相关性,可以分析突变或选择对密码子偏好性的影响,即当 GC₁₂ 与 GC₃ 显著相关时,表明 3 个位置密码子碱基组成无显著差异,且偏好性主要受突变的影响;当 GC₁₂ 与 GC₃ 呈不显著相关时,回归系数趋近于 0,表明密码子的前 2 位碱基和第 3 位碱

基的组成不同,基因组中的 GC 含量比较保守,密码子的偏倚主要受到选择的影响^[13]。

1.2.3 最优密码子的确定 以美国红栲每个叶绿体基因的 ENC 为参考标准,从两端选择 10% 的基因构建高、低偏倚库,将 2 库间 $\Delta\text{RSCU} \geq 0.08$ 的密码子定义为高表达的优越密码子;将 RSCU 值大于 1 的密码子确定为高频密码子^[18]。将既能满足高频率密码子又能满足高表达优越密码子确定为最优密码子。

2 结果与分析

2.1 密码子碱基组成

用 CUSP 软件分析美国红栲 52 条 CDS 的碱基组成,用 Codon W 1.4.2 分析其 ENC 值(表 1),所有 CDS 密码子的平均 GC 含量为 36.93%,第 1 位

GC 含量为 45.23%,第 2 位为 39.23%,第 3 位为 26.19%,GC 含量在密码子不同位置的分布频率不同,由高到低,依次为第 1 位 > 第 2 位 > 第 3 位,第 1、第 2 位的 GC 含量明显高于第 3 位,结果表明,在美国红栲中,叶绿体密码子的最后 1 个碱基主要是 A/U(T),这与植物叶绿体基因中 A/U(T)含量较高的特点是一致的。表示偏好强度的 ENC 值的范围为 37.55 ~ 55.28,平均值为 47.27;52 个 CDS 中有 34 个 ENC 值 > 45,说明美国红栲的大部分编码序列具有弱的密码子偏好性。对不同密码子位置碱基的 GC 含量、密码子数(N)和 ENC 数值的相关分析(表 2)表明,GC_{all}和 GC₁、GC₂、GC₃ 的相关性极显著,GC₁ 和 GC₂ 相关性极显著,相关系数为 0.513,GC₃ 与 GC₁、GC₂ 的相关性不显著,说明密码子的第 1 位和第 2 位碱基的组成与第 3 位碱基组成的相似

表 1 美国红栲叶绿体 CDS 密码子各位置的 GC 含量

| 基因 | GC ₁ (%) | GC ₂ (%) | GC ₃ (%) | GC _{all} (%) | ENC | 基因 | GC ₁ (%) | GC ₂ (%) | GC ₃ (%) | GC _{all} (%) | ENC |
|-------------|---------------------|---------------------|---------------------|-----------------------|-------|--------------|---------------------|---------------------|---------------------|-----------------------|-------|
| <i>accD</i> | 51.61 | 49.19 | 25.81 | 42.20 | 44.14 | <i>psbA</i> | 44.69 | 41.37 | 30.34 | 40.35 | 41.75 |
| <i>atpA</i> | 49.72 | 43.50 | 33.90 | 42.37 | 45.65 | <i>psbB</i> | 55.21 | 46.56 | 30.45 | 44.07 | 47.31 |
| <i>atpB</i> | 41.22 | 32.35 | 26.23 | 33.27 | 46.41 | <i>psbC</i> | 53.38 | 45.78 | 32.28 | 43.81 | 44.39 |
| <i>atpE</i> | 55.12 | 39.96 | 24.41 | 39.83 | 45.05 | <i>psbD</i> | 51.98 | 43.22 | 31.64 | 42.28 | 44.74 |
| <i>atpF</i> | 45.35 | 31.27 | 33.28 | 32.46 | 43.76 | <i>rbcL</i> | 57.14 | 43.49 | 29.83 | 43.49 | 47.34 |
| <i>atpI</i> | 46.97 | 32.34 | 22.30 | 32.02 | 46.31 | <i>rpl14</i> | 50.41 | 37.40 | 28.46 | 38.75 | 52.41 |
| <i>ccsA</i> | 24.43 | 35.47 | 24.68 | 33.45 | 44.47 | <i>rpl16</i> | 50.00 | 52.94 | 29.41 | 44.12 | 43.05 |
| <i>cemA</i> | 38.26 | 29.13 | 31.74 | 33.04 | 53.37 | <i>rpl2</i> | 50.92 | 48.35 | 31.50 | 43.59 | 55.28 |
| <i>clpP</i> | 58.38 | 36.55 | 30.46 | 41.79 | 51.35 | <i>rpl20</i> | 35.00 | 37.50 | 25.00 | 32.50 | 48.31 |
| <i>matK</i> | 36.24 | 29.26 | 28.68 | 31.40 | 46.40 | <i>rpoA</i> | 43.98 | 30.12 | 25.30 | 33.13 | 44.62 |
| <i>ndhA</i> | 40.66 | 37.91 | 21.15 | 33.24 | 48.29 | <i>rpoB</i> | 47.85 | 37.35 | 28.85 | 38.59 | 49.33 |
| <i>ndhB</i> | 42.19 | 39.15 | 31.85 | 37.73 | 49.29 | <i>rpoC1</i> | 48.91 | 37.19 | 25.18 | 37.19 | 49.23 |
| <i>ndhC</i> | 45.45 | 33.88 | 24.79 | 34.71 | 41.32 | <i>rpoC2</i> | 42.32 | 35.19 | 26.16 | 34.27 | 46.64 |
| <i>ndhD</i> | 37.70 | 37.96 | 30.37 | 35.34 | 47.25 | <i>rps11</i> | 52.52 | 54.68 | 24.46 | 43.88 | 51.25 |
| <i>ndhE</i> | 39.22 | 33.33 | 26.47 | 33.01 | 43.95 | <i>rps12</i> | 51.93 | 48.39 | 32.26 | 44.35 | 42.58 |
| <i>ndhF</i> | 35.98 | 33.07 | 23.54 | 30.86 | 42.76 | <i>rps14</i> | 42.34 | 47.52 | 24.75 | 38.61 | 39.34 |
| <i>ndhG</i> | 44.63 | 35.59 | 23.16 | 34.46 | 43.31 | <i>rps18</i> | 36.54 | 40.38 | 25.00 | 33.97 | 37.55 |
| <i>ndhH</i> | 51.27 | 35.53 | 25.63 | 37.48 | 48.43 | <i>rps2</i> | 41.23 | 40.93 | 27.85 | 36.43 | 46.23 |
| <i>ndhI</i> | 41.57 | 36.14 | 23.49 | 33.73 | 46.31 | <i>rps3</i> | 42.92 | 33.79 | 21.00 | 32.57 | 47.94 |
| <i>ndhJ</i> | 49.69 | 37.11 | 27.04 | 37.95 | 47.08 | <i>rps4</i> | 50.99 | 36.14 | 24.75 | 37.29 | 45.30 |
| <i>ndhK</i> | 45.79 | 43.39 | 27.75 | 39.25 | 46.67 | <i>rps7</i> | 52.56 | 42.39 | 23.72 | 40.38 | 46.59 |
| <i>petA</i> | 52.83 | 37.94 | 29.91 | 39.98 | 45.34 | <i>rps8</i> | 37.78 | 40.74 | 26.67 | 35.06 | 41.92 |
| <i>petB</i> | 48.34 | 41.67 | 30.09 | 39.51 | 41.04 | <i>ycf1</i> | 33.56 | 28.21 | 25.09 | 28.95 | 47.37 |
| <i>petD</i> | 50.18 | 38.16 | 22.35 | 35.58 | 38.83 | <i>ycf2</i> | 41.84 | 34.14 | 35.34 | 36.48 | 52.54 |
| <i>psaA</i> | 51.66 | 43.28 | 32.49 | 42.48 | 51.63 | <i>ycf3</i> | 47.93 | 38.46 | 30.18 | 38.86 | 51.50 |
| <i>psaB</i> | 47.39 | 44.93 | 30.32 | 41.34 | 47.83 | <i>ycf4</i> | 42.56 | 38.24 | 31.62 | 39.91 | 51.71 |
| 总平均 | | | | | | | 45.23 | 39.23 | 26.19 | 36.93 | 47.27 |

注:GC_{all}表示密码子各位置的平均数。

表 2 密码子各位置 GC 含量、数量与 ENC 值的相关性分析

| 变量 | 相关系数 | | | | |
|-------------------|-----------------|-----------------|-----------------|-------------------|---------|
| | GC ₁ | GC ₂ | GC ₃ | GC _{all} | ENC |
| GC ₂ | 0.513 ** | | | | |
| GC ₃ | 0.243 | 0.212 | | | |
| GC _{all} | 0.817 ** | 0.832 ** | 0.592 ** | | |
| ENC | 0.249 * | -0.084 | 0.427 ** | 0.231 | |
| 密码子数 | -0.165 | -0.239 * | 0.259 * | -0.147 | 0.264 * |

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

和不同。ENC 与 GC₁ 显著相关,与 GC₂ 相关性不显著,与 GC₃ 极显著相关,说明 ENC 与密码子第 3 位碱基的组成密切相关。ENC 与密码子数(N)显著相关,说明基因编码序列的长度对密码子的偏倚有一定的影响。

编码氨基酸的 RSCU(表 3)显示,RSCU>1 的密码子中以 A 和 U 结尾的频率较高,其中 14 个密码子以 U 结尾、12 个以 A 结尾、1 个以 G 结尾,表明美国红栲叶绿体基因组更倾向于以 A 和 U 结尾;而以 C 和 G 结尾的密码子即为非偏性的密码子。

表 3 美国红栲各氨基酸的 RSCU 分析

| AA | 密码子 | 数目 | RSCU | AA | 密码子 | 数目 | RSCU | AA | 密码子 | 数目 | RSCU |
|-----|------------|-----|------|-----|------------|-----|------|-----|------------|-----|------|
| Phe | <u>UUU</u> | 937 | 1.27 | Pro | <u>CCU</u> | 393 | 1.23 | Lys | <u>AAA</u> | 925 | 1.43 |
| | UUC | 629 | 0.84 | | <u>CCC</u> | 275 | 1.03 | | AAG | 325 | 0.52 |
| | | | | | <u>CCA</u> | 292 | 1.12 | Asp | <u>GAU</u> | 793 | 1.57 |
| Leu | <u>UUA</u> | 782 | 1.86 | Thr | CCG | 117 | 0.54 | | GAC | 192 | 0.41 |
| | <u>UUG</u> | 539 | 1.24 | | <u>ACU</u> | 430 | 1.63 | Glu | <u>GAA</u> | 847 | 1.49 |
| | <u>CUU</u> | 429 | 1.19 | | ACC | 197 | 0.75 | | GAG | 293 | 0.51 |
| | CUC | 142 | 0.35 | | <u>ACA</u> | 331 | 1.25 | Cys | UGU | 176 | 1.46 |
| | CUA | 279 | 0.73 | Ala | ACG | 99 | 0.37 | | UGC | 65 | 0.54 |
| Ile | CUG | 129 | 0.42 | | <u>GCU</u> | 524 | 1.80 | Arg | CGU | 297 | 0.72 |
| | <u>AUU</u> | 927 | 1.39 | | GCC | 192 | 0.66 | | CGC | 80 | 0.38 |
| | AUC | 329 | 0.51 | | <u>GCA</u> | 332 | 1.14 | | <u>CGA</u> | 290 | 1.37 |
| | AUA | 672 | 0.92 | | GCG | 119 | 0.41 | | CGG | 82 | 0.39 |
| Val | <u>GUU</u> | 452 | 1.52 | Tyr | <u>UAU</u> | 656 | 1.65 | | <u>AGA</u> | 391 | 1.85 |
| | GUC | 143 | 0.51 | | UAC | 141 | 0.35 | Ser | AGG | 127 | 0.61 |
| | <u>GUA</u> | 452 | 1.49 | His | <u>CAU</u> | 388 | 1.56 | | <u>UCU</u> | 498 | 1.64 |
| | GUG | 129 | 0.47 | | CAC | 110 | 0.44 | | <u>UCC</u> | 396 | 1.14 |
| Gly | <u>GGU</u> | 592 | 1.33 | Gln | <u>CAA</u> | 613 | 1.59 | | <u>UCA</u> | 372 | 1.22 |
| | GGC | 148 | 0.40 | | CAG | 160 | 0.41 | | UCG | 149 | 0.62 |
| | <u>GGA</u> | 589 | 1.58 | Asn | <u>AAU</u> | 873 | 1.49 | | AGU | 379 | 0.81 |
| | GGG | 273 | 0.72 | | AAC | 225 | 0.49 | | AGC | 173 | 0.42 |

注:下划线表示最优密码子。

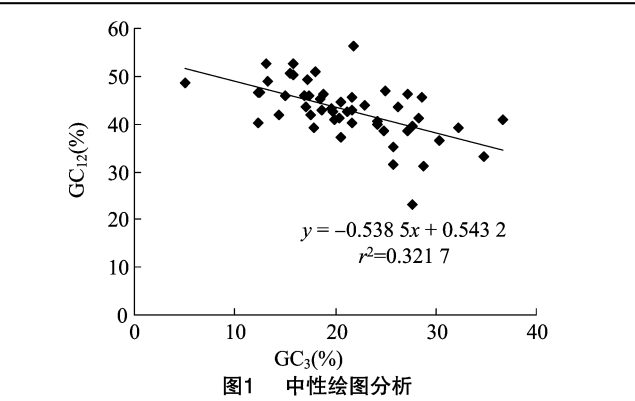
2.2 中性绘图分析

根据美国红栲叶绿体基因组中每个基因的中性绘图分析(图 1),GC₁₂的取值范围略大在 0.314~0.529 之间,GC₃的取值范围很小,为 0.193~0.392 之间,同时大多数基因均位于对角线以上;GC₁₂与 GC₃的相关系数为 0.321 7,相关性不显著,回归系数(趋势线的斜率)为 -0.538 5,表明在美国红栲叶绿体基因组的中性绘图分析中,GC₁₂和 GC₃的相关性很弱,说明密码子第 1 位、第 2 位和第 3 位碱基的组成存在差异,即美国红栲叶绿体基因组的 GC 含量是高度保守的,而密码子第 3 位的 GC 含量相对

较低,其密码子偏好性受选择的影响较大。

2.3 最优密码子确定

将美国红栲叶绿体基因组中的 52 个蛋白编码基因并作一个整体在 Codon W 1.4.2 软件上运行,通过构建高表达基因和低表达基因库,对这 2 个基因数据库的 RSCU 值进行了重新计算,结果(表 4)表明,ΔRSCU≥0.08 的密码子包括 UUU(TTT)、UUA(TTA)等 23 个密码子,都是高表达的优越密码子,其中 11 个以 A 结尾,8 个以 U 结尾,3 个以 C 结尾,1 个为 G 结尾;ΔRSCU≥0.3 有 11 个密码子;ΔRSCU≥0.5 的密码子是 UUU、UUA。以同时满足



高频率密码子和高表达优越密码子作为最优密码子,分别为UUU、UUA、GUA、UCA、CCA、UAU、GCA、CAU、CAA、AAU、AAA、GAA、UGU、CGA、AGA、GGU和GGA等17个,其中11个以A结尾,6个以U结尾。

3 讨论与结论

在生物体内,密码子在核酸和蛋白质的翻译中起着重要作用;植物中不同密码子的使用频率存在差异,这种密码子使用偏好性是物种和基因长期进

表 4 美国红栉叶绿体基因组最优密码子的确定

| AA | 密码子 | 高表达基因 | | 低表达基因 | | ΔRSCU |
|-----|---------|-------|------|-------|------|-------|
| | | 数目 | RSCU | 数目 | RSCU | |
| Phe | UUU *** | 52 | 1.69 | 132 | 1.07 | 0.62 |
| | UUC | 7 | 0.31 | 114 | 0.93 | -0.62 |
| Leu | UUA *** | 54 | 3.23 | 85 | 1.25 | 1.98 |
| | UUG | 17 | 1.15 | 96 | 1.41 | -0.26 |
| | CUU | 23 | 1.21 | 94 | 1.38 | -0.17 |
| | CUC | 0 | 0 | 36 | 0.53 | -0.53 |
| | CUA | 7 | 0.29 | 63 | 0.92 | -0.63 |
| | CUG | 3 | 0.12 | 35 | 0.51 | -0.39 |
| | AUU | 47 | 1.34 | 134 | 1.3 | 0.04 |
| Ile | AUC | 16 | 0.51 | 71 | 0.69 | -0.18 |
| | AUA * | 36 | 1.15 | 105 | 1.02 | 0.13 |
| Val | GUU | 15 | 1.33 | 56 | 1.27 | 0.06 |
| | GUC | 7 | 0.62 | 29 | 0.66 | -0.04 |
| | GUA ** | 20 | 1.78 | 65 | 1.48 | 0.3 |
| | GUG | 3 | 0.27 | 26 | 0.59 | -0.32 |
| Ser | UCU | 21 | 1.62 | 89 | 1.61 | 0.01 |
| | UCC | 6 | 0.46 | 63 | 1.14 | -0.68 |
| | UCA * | 19 | 1.46 | 66 | 1.2 | 0.26 |
| | UCG ** | 14 | 1.08 | 36 | 0.65 | 0.43 |
| | AGU * | 17 | 1.31 | 58 | 1.05 | 0.26 |
| | AGC | 1 | 0.08 | 19 | 0.34 | -0.26 |
| Pro | CCU | 14 | 1.3 | 57 | 1.42 | -0.12 |
| | CCC * | 9 | 0.93 | 33 | 0.82 | 0.11 |
| | CCA * | 19 | 1.4 | 46 | 1.15 | 0.25 |
| | CCG | 3 | 0.37 | 24 | 0.6 | -0.23 |
| Thr | ACU | 18 | 1.33 | 62 | 1.39 | -0.06 |
| | ACC ** | 15 | 1.11 | 35 | 0.79 | 0.32 |
| | ACA | 16 | 1.04 | 61 | 1.37 | -0.33 |
| | ACG | 5 | 0.52 | 20 | 0.45 | 0.07 |
| Tyr | UAU ** | 29 | 1.86 | 100 | 1.54 | 0.32 |
| | UAC | 3 | 0.14 | 30 | 0.46 | -0.32 |

| AA | 密码子 | 高表达基因 | | 低表达基因 | | ΔRSCU |
|-----|--------|-------|------|-------|------|-------|
| | | 数目 | RSCU | 数目 | RSCU | |
| Ala | GCU | 18 | 1.39 | 70 | 1.62 | -0.23 |
| | GCC | 5 | 0.49 | 31 | 0.72 | -0.23 |
| | GCA ** | 24 | 1.71 | 54 | 1.25 | 0.46 |
| | GCG | 4 | 0.41 | 18 | 0.42 | -0.01 |
| His | CAU * | 16 | 1.65 | 97 | 1.56 | 0.09 |
| | CAC | 3 | 0.35 | 27 | 0.44 | -0.09 |
| Gln | CAA ** | 24 | 1.76 | 102 | 1.45 | 0.31 |
| | CAG | 4 | 0.24 | 39 | 0.55 | -0.31 |
| Asn | AAU * | 39 | 1.7 | 155 | 1.49 | 0.21 |
| | AAC | 6 | 0.3 | 53 | 0.51 | -0.21 |
| Lys | AAA ** | 47 | 1.75 | 144 | 1.38 | 0.37 |
| | AAG | 6 | 0.25 | 65 | 0.62 | -0.37 |
| Asp | GAU | 17 | 1.4 | 160 | 1.65 | -0.25 |
| | GAC * | 6 | 0.6 | 34 | 0.35 | 0.25 |
| Glu | GAA * | 31 | 1.52 | 123 | 1.28 | 0.24 |
| | GAG | 10 | 0.48 | 69 | 0.72 | -0.24 |
| Cys | UGU ** | 9 | 1.78 | 33 | 1.43 | 0.35 |
| | UGC | 2 | 0.22 | 13 | 0.57 | -0.35 |
| Arg | CGU * | 14 | 1.06 | 39 | 0.98 | 0.08 |
| | CGC | 2 | 0.23 | 16 | 0.4 | -0.17 |
| | CGA * | 21 | 1.59 | 52 | 1.3 | 0.29 |
| | CGG | 4 | 0.38 | 24 | 0.6 | -0.22 |
| | AGA ** | 29 | 2.2 | 70 | 1.75 | 0.45 |
| Gly | AGG | 7 | 0.53 | 39 | 0.98 | -0.45 |
| | GGU * | 23 | 1.45 | 79 | 1.32 | 0.13 |
| | GGC | 6 | 0.55 | 25 | 1.42 | -0.87 |
| | GGA | 28 | 1.52 | 89 | 1.48 | 0.04 |
| | GGG | 7 | 0.48 | 47 | 0.78 | -0.3 |

注: * 表示 ΔRSCU ≥ 0.08, ** 表示 ΔRSCU ≥ 0.3, *** 表示 ΔRSCU ≥ 0.5。

化和对环境的适应过程中形成的,是多种因子共同作用的结果,其中突变和自然选择是该现象形成的重要影响因素^[7-8]。叶绿体是植物光合作用的细胞器,也包含相对独立的母体遗传基因组信息,因此叶绿体基因组在揭示物种进化、物种亲缘关系、物种鉴定等方面具有重要价值;同时叶绿体基因工程因其可高效表达、安全等特点已成为植物基因工程的研究热点^[10]。因此,对植物叶绿体基因组密码子使用偏好的研究可以揭示物种基因组的进化关系及其主要影响因素。

本研究中美国红栲叶绿体基因组中的 GC₃ 与 GC₁ 和 GC₂ 没有显著相关性,且显著小于 GC₁ 和 GC₂。这说明美国红栲叶绿体基因的密码子偏好以 A 和 U 结尾,RSCU 分析结果从定量分析的角度充分证明了这一观点。这与已报道的黄芩(*Scutellaria baicalensis*)^[19]、普通油茶(*Camellia oleifera*)^[20]、蒺藜苜蓿(*Medicago truncatula*)^[15]的叶绿体基因特征相同;中性绘图分析表明,密码子的第 1 位和第 2 位与第 3 位碱基组成存在着显著差异,由于密码子的 GC 含量高度保守,密码子的偏倚性主要受选择的影响,与蒺藜苜蓿一致^[13];美国红栲叶绿体基因组密码子偏好以 AU 结尾,且其最优密码子为 UUU、UUA,与大多数高等植物的最优密码子 NNA、NNU 的模式一致^[18]。本研究中,美国红栲叶绿体基因组的密码子偏好主要受到选择的影响,同时与其他因素共同作用,确定了 17 个最优密码子,均为 NNA 和 NNU 模式。采取高表达的高频密码子作为最优密码子,在美国红栲叶绿体基因组中选择的 17 个最优密码子分别是 UUU、UUA、GUA、UCA、CCA、UAU、GCA、CAU、CAA、AAU、AAA、GAA、UGU、CGA、AGA、GGU 和 GGA。美国红栲叶绿体基因组最优密码子的确定,为优化目标基因的高效表达密码子,从而通过叶绿体基因工程改良美国红栲的重要性状提供了科学依据。

参考文献:

- [1] 郭学明,高忠明,刘振林,等. 美国红栲雄株和雌株茎导管分子的形态解剖比较[J]. 林业科学,2010,46(8):51-55.
- [2] 梁远楠,陈水莲,张丽君,等. 10 个降香黄檀家系在肇庆地区的早期生长评价[J]. 林业与环境科学,2019,35(2):106-110.
- [3] Nelson D L, Cox M M. Lehninger principles of biochemistr [M]. New York: W. H. Freeman and Company, 2017.
- [4] Zhou Z P, Dang Y K, Zhou M, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription[J]. Proceedings of the National Academy of Sciences, 2016, 113(41):6117-6125.
- [5] Hershberg R, Dmitri A P. Selection on codon bias[J]. Annual Review of Genetics, 2008, 42(1):287-299.
- [6] Duret L. Evolution of synonymous codon usage in metazoans[J]. Curr Opin Genet Dev, 2012, 12(6):640-649.
- [7] Romero 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):2084-2090.
- [8] Chen X, Cai X N, Chen Q Z, et al. Factors affecting synonymous codon usage bias in chloroplast genome of *Oncidium Gower Ramsey* [J]. Evolutionary Bioinformatics, 2011(7):271-278.
- [9] 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.
- [10] Wright F. The 'effective number of codons' used in a gene[J]. Gene, 1990, 87(1):23-29.
- [11] 牛 元,徐 琼,王崙德,等. 大花香水月季叶绿体基因组密码子使用偏性分析[J]. 西北林学院学报,2018,33(3):123-130.
- [12] Wu X M. The analysis method and progress in the study of codon bias[J]. Hereditas, 2007, 29(4):420-426.
- [13] 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, 35(3):362-371.
- [14] Jiang Y, Deng F, Wang H, et al. An extensive analysis on the global codon usage pattern of baculoviruses [J]. Archives of Virology, 2008, 153(12):2273-2282.
- [15] Qin Z, Zheng Y J, Gui L J, et al. Codon usage bias analysis of chloroplast genome of camphora tree (*Cinnamomum camphora*) [J]. Guihaia, 2018, 38(10):1346-1355.
- [16] Chao Y, Chang Y, Wang M F, et al. Codon usage bias and cluster analysis on chloroplastic genes from seven crop species [J]. Acta Agric Bor Sin, 2012, 27(4):60-64.
- [17] Sueoka N. Near homogeneity of PR2 - Bias fingerprints in the human genome and their implications in phylogenetic analyses [J]. Journal of Molecular Evolution, 2001, 53(4/5):469-476.
- [18] Shang M Z, Liu F, Hua J P, et al. Analysis on codon usage of chloroplast genome of *Gossypium hirsutum* [J]. Sci Agric Sin, 2011, 44(2):245-253.
- [19] Wang W B, Yu H, Qiu X P. Analysis of repeat sequence and codon bias of chloroplast genome in *Scutellaria baicalensis* [J]. Molecular Plant Breeding, 2018, 16(8):2445-2452.
- [20] Wang P L, Yang L P, Wu H Y, et al. Codon preference of chloroplast genome in *Camellia oleifera* [J]. Guihaia, 2018(38):135-144.