

紫叶苔(*Pleurozia purpurea*)线粒体基因组密码子 偏好性分析

杜梦雪¹, 张鑫¹, 高洁¹, 毕胜¹, 郝杰威^{1,2}, 张莉娜^{1,2}

(1. 海南大学热带特色林木花卉遗传与种质创新教育部重点实验室, 海南海口 570228;
2. 海南大学海南省环南海陆域生物多样性国际联合研究中心, 海南海口 570228)

摘要: 为了丰富对紫叶苔属遗传基础的认识, 探究了紫叶苔(*Pleurozia purpurea*)线粒体基因组密码子使用的偏好性及其影响因素。首先, 基于完整紫叶苔线粒体基因组筛选蛋白编码序列; 然后, 利用 CodonW、EMBOSS explorer 计算各基因的 GC 含量、有效密码子数(ENC)、密码子适应指数(CAI)和同义密码子相对使用度(RSCU)等指标; 最后, 采用 R 软件进行参数间的相关性分析并绘图。结果表明: 1) 从紫叶苔线粒体基因组中筛选出 31 条蛋白编码序列, 平均 GC 含量和密码子第 3 位碱基的 GC 含量分别为 38.17%和 32.28%, ENC 和 CAI 平均值为 50.29、0.17, 表明密码子偏好使用以 A/U 结尾的碱基, 且密码子偏好性较弱; 2) ENC 分别与 GCall、GC3 以及基因编码氨基酸的长度呈极显著正相关, 表明 GC 含量和基因的长度对密码子的偏好有显著影响; 3) 中性绘图、ENC-plot 和 PR2-plot 分析表明, 突变压力和自然选择等因素共同塑造了密码子的使用偏好性, 其中突变是最主要的影响因素; 4) 筛选出 23 个最优密码子, 多数以 A/U 结尾。因此, 紫叶苔线粒体基因组密码子偏好性较弱, 除了主要受突变压力影响外, 还受到自然选择、碱基组成和基因长度等因素的影响, 研究结果可为苔藓植物线粒体基因组的进化研究提供参考。

关键词: 分子遗传学; 紫叶苔属; 密码子偏好性; 线粒体基因组; 最优密码子

中图分类号: Q949.35; Q755

文献标识码: A

Analysis of codon usage bias in the mitochondrial genome of *Pleurozia purpurea*

DU Mengxue¹, ZHANG Xin¹, GAO Jie¹, BI Sheng¹, HAO Jiewei^{1,2}, ZHANG Lina^{1,2}

(1. Ministry of Education Key Laboratory for Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Hainan University, Haikou, Hainan 570228, China;

2. International Joint Center for Terrestrial Biodiversity around South China Sea of Hainan Province, Hainan University, Haikou, Hainan 570228, China)

Abstract: To enrich our understanding of the genetic basis of the genus *Pleurozia*, the codon usage bias (CUB) in the mitochondrial genome of *Pleurozia purpurea* and its main influencing factors were investigated. First, protein-coding sequences were screened from the complete mitochondrial genome of *P. purpurea*. Subsequently, CodonW and EMBOSS explorer were utilized to calculate various metrics for each gene, including GC content, the effective number of codons (ENC), codon adaptation index (CAI), and the relative synonymous codon usage (RSCU). Finally, correlation analysis between the parameters and plotting were performed using R software. The results showed: 1) A total of 31 protein-coding sequences were identified from the mitochondrial genome of *P. purpurea*. The average GC content and the GC content at the third codon position were 38.17% and 32.28%, respectively. The average values for ENC and CAI were 50.29 and 0.17, indicating a preference for codons ending with A/U bases and a relatively weak CUB. 2) ENC exhibited highly significant positive correlations with GCall, GC3, and the length of amino acid encoding genes respectively, suggesting significant influences of GC content and gene length on CUB. 3) Analyses including neutrality plot, ENC-plot and PR2-plot indicated that mutation pressure and natural selection jointly shaped the CUB, with mutation

收稿日期: 2024- ; 修回日期: 2024-08-21; 责任编辑: 张士莹

基金项目: 国家自然科学基金(32160315, 31760054); 海南大学科研启动项目 (KYQD(ZR)-22105)

第一作者简介: 杜梦雪(1994—), 女, 河南平顶山人, 硕士研究生, 主要从事苔藓植物基因组和系统进化方面的研究。

通信作者: 张莉娜, 副教授。E-mail: zhanglina@hainanu.edu.cn

being the most important influencing factor. 4) A total of 23 optimal codons were selected, most of which end with A/U. In conclusion, the CUB in the mitochondrial genome of *P. purpurea* is relatively weak, influenced primarily by mutational pressure, but also by factors such as natural selection, base composition, and gene length. These findings can provide a reference for evolutionary studies of mitochondrial genomes in bryophytes.

Keywords: molecular genetics; *Pleurozia*; codon usage bias; mitochondrial genomes; optimal codon

植物线粒体是关键细胞器，负责代谢、呼吸和能量转换等重要细胞过程^[1]。和叶绿体一样，线粒体是半自主性细胞器，有自己的基因组，具有独立复制表达及母系遗传等特点^[2]。线粒体基因组编码与呼吸作用等代谢过程有关的蛋白质中含有丰富的遗传信息。在遗传信息的解读过程中，密码子扮演着至关重要的角色，是基因与蛋白质之间联系的基础。已知构成天然蛋白质的 20 种基本氨基酸中，除了色氨酸和甲硫氨酸各由一个密码子编码外，其他 18 种氨基酸由 2—6 个不同的密码子进行翻译，这种现象被称为密码子简并性^[3]，而编码同一种氨基酸的不同密码子被定义为同义密码子。同义密码子在蛋白质合成中的使用并不均匀，一些同义密码子被优先使用，这种对特定密码子的偏倚，即为密码子使用偏好(codon usage bias, CUB)^[4]。CUB 信息的获得可以为基因表达水平的预测和优化、基因功能预测、遗传分歧评估和生物进化探索等提供重要的见解^[5-7]，并有助于解释物种对环境适应的分子机制^[8]。

苔藓植物是由水生到陆生过渡的高等植物类群，大多生活在潮湿的陆地环境中，包含地钱门(Marchantiophyta)、真藓门(Bryophyta)和角苔门(Anthocerotophyta)3 个门类。紫叶苔属(*Pleurozia* Dumort.)隶属于地钱门，在最新的世界苔类名录中共记录 12 种^[9]，中国记录 5 种。紫叶苔属是唯一具有两面体顶细胞结构的茎叶体苔类^[10]，被视为苔类植物中由简单叶状体向茎叶体过渡的关键类群^[11-12]。同时，紫叶苔属具有形态独特且变异较大的叶型和生殖结构，如囊状叶、不育蒴萼等^[13]，使得这一类群在苔类系统发育与进化研究中具有非常重要的科学意义。此外，紫叶苔属多为树附生类型，主要分布于热带和亚热带山地高海拔多云雾的雨林中^[13]，对于气候等环境因子的变化十分敏感。因此，紫叶苔属不仅是高海拔雨林或云雾林生态系统动态变化潜在的指示植物，而且极易成为受威胁物种。在《中国生物多样性红色名录——高等植物卷(2020)》中，紫叶苔(*Pleurozia purpurea* Lindb.)和狭尖紫叶苔(*P. subinflata* (Austin) Austin)被列为近危物种。

地钱门是现存陆生植物最早分化的系群之一^[14-15]，开展紫叶苔属基因组的研究有助于揭示陆生植物起源和早期进化过程中基因组的演化。已有研究表明，紫叶苔线粒体基因组在苔藓植物中具有较高的进化保守性和代表性^[16]。然而，尽管测序技术不断进步，对维管植物细胞器基因组的研究日益增多，但关于苔藓植物线粒体基因组的研究仍然较少^[14,17-18]，特别是关于密码子使用偏好性较为全面的分析更是寥寥无几^[19-20]。本研究深入分析紫叶苔线粒体基因组的密码子使用模式，探讨其偏好性形成的潜在机制，并筛选出最优密码子。通过研究，期望能够丰富对紫叶苔属这一进化和生态关键类群遗传基础的认识，为理解该属的生物学特性及其在苔藓植物系统发育、进化和物种保护中的关系提供新的视角和科学依据。

1 材料和方法

1.1 紫叶苔线粒体基因组序列的获取

紫叶苔线粒体基因组序列从 GenBank 数据库下载，登录号为 NC_013444.1。该基因组总长为 168 526 个碱基，含有 41 条蛋白编码序列(coding sequence, CDS)。为了提高后续密码子偏好分析的准确性，本文依照长度大于 300 bp，起始密码子为 ATG，终止密码子为 TGA、TAA 和 TAG 的标准，最终选取了 31 条 CDS 序列用于后续数据分析。

1.2 密码子使用特征参数分析

运用 CodonW (版本号 v1.4.2) 对紫叶苔线粒体基因组 CDS 序列的密码子使用指标进行分析，包括 ENC、CAI、RSCU、GC3s、A3s、T3s、C3s 和 G3s。其中 A3s、T3s、C3s 和 G3s 代表了同义密码子第 3 碱基位置分别使用 A/T/C/G 的频率；GC3s 指同义密码子第 3 位置的平均 GC 含量；ENC (effective number of codon)代表有效密码子使用个数；CAI (codon adaptation index)为密码子适应指数；RSCU (relative synonymous codon usage) 代表相对同义密码子使用度。EMBOSS explorer (<http://www.bioinformatics.nl/emboss-explorer/>) 被用来计算密码子的平均 GC 含量(GCall)以及密码子的第 1、

2 和 3 位置的 GC 含量(分别表示为 GC1、GC2 和 GC3), GC12 为 GC1 和 GC2 的平均值。

1.3 中性绘图分析

中性绘图分析是一种用于研究突变压力和自然选择对密码子使用偏好影响的方法^[21]。利用 R 语言中的 ggplot2 包, 以 GC3 作为横坐标, GC12 作为纵坐标, 构建一个二维散点分布图以表征之前筛选出来的 31 条序列。此外, 使用 SPSS 18.0, 基于皮尔逊相关性分析, 绘制 GC12 和 GC3 的线性关系回归线。该回归曲线的斜率大小表示密码子使用偏好的主导原因。如果回归曲线接近对角线(斜率=1.0), GC12 和 GC3 之间存在显著相关性, 表示突变压力是密码子偏好的主要驱动因素, 自然选择压力无或较弱; 相反, 如果回归曲线偏离对角线, 斜率趋近于 0, 且 GC12 和 GC3 之间没有显著相关性, 则表明自然选择是塑造密码子使用的主要力量^[22]。

1.4 ENC-plot 分析

有效密码子数(effective number of codon, ENC)是评估同义密码子使用不均衡度的重要指标^[23], 其范围通常在 20~61 之间, 与密码子的偏好性呈负相关。ENC 值为 20 表示极端偏倚, 每个氨基酸仅使用一个密码子; 而 ENC 值接近 61 时, 可以认为基本没有偏好性。较低的 ENC 值(≤ 35)表示基因具有较强的密码子偏好, 即倾向于使用特定的密码子来编码每个氨基酸; 较高的 ENC 值则表明密码子使用偏好较弱^[24]。ENC-plot 常用于探究影响特定基因密码子偏好性的决定性因素, 将各基因没有密码子使用偏好的期望 ENC 值^[23]作为纵坐标, 其 GC3s 值作为横坐标来绘制期望曲线。如果基因分布在预期曲线上或接近曲线, 可以认为突变压力对密码子使用偏好产生了显著影响; 反之, 如果这些点落在标准曲线下方且显著偏离预期曲线, 则可能是自然选择等因素在密码子偏好的形成中发挥了主要作用^[22]。为了更准确地评估期望 ENC 值与实际 ENC 值之间的差异, 可以使用公式: ENC 比率 = (ENC 期望-ENC 实际) / ENC 期望来计算, 并通过直方图将这些 ENC 比率的分布情况呈现出来。

1.5 PR2-plot 分析

Parity rule 2 (PR2)-plot 分析应用于评估具有 4 个同义密码子氨基酸的密码子偏好, 并进一步确定突变压力和自然选择对密码子使用的影响^[25]。因此, 本研究选择了具有 4 个同义密码子的氨基酸(如缬氨酸、脯氨酸、苏氨酸、丙氨酸、甘氨酸)来计算每个基因第 3 个碱基位置的组成频率。散点图的构建以 $A3s/(A3s + T3s)$ 作为纵坐标, 以 $G3s/(G3s + C3s)$ 作为横坐标。图的中心点代表链内核苷酸组成遵循 PR2 ($A = T$ 和 $G = C$), 表示 DNA 两条互补链间的突变和选择压力没有偏差, 距中心的距离和方向表示偏离 PR2 的程度和方向。如果在密码子的第 3 位置上, AT 的频率等于 GC 的频率, 图中的点将沿坐标轴成比例分布, 则表明基因的密码子使用偏好仅受到突变压力的影响; 否则, 说明自然选择或两者共同作用影响了密码子的不均等使用^[26]。

1.6 RSCU 分析和最优密码子筛选

密码子相对使用频率(relative synonymous codon usage, RSCU)是指观察到的密码子使用频率与该密码子的预期使用频率的比值(假设特定氨基酸的同义密码子不存在密码子使用偏好)^[27]。当 RSCU 等于 1 时, 表示密码子使用是无偏的, 即在 RNA 转录中密码子被随机或平等地选择; 当 RSCU 值大于 1 时, 表示该密码子的使用频率高于预期, 为高频密码子; RSCU 值小于 1, 则表示该密码子的使用频率较低^[28]。

已知较低的 ENC 值代表较高的密码子使用偏倚, 而高表达基因通常具有较大程度的密码子偏好, 因此具有较小的 ENC 值。最优密码子的确定是将基因按其 ENC 值降序排列, 选择排在 ENC 值高低两端基因中的 10% 分别建立低表达和高表达基因库。根据 CodonW 软件计算 2 个基因库中密码子的 RSCU 值以及二者的差异值($\Delta RSCU$)。满足 $RSCU > 1$ 且 $\Delta RSCU > 0.08$ 的密码子被定义为基因的最优密码子^[29]。

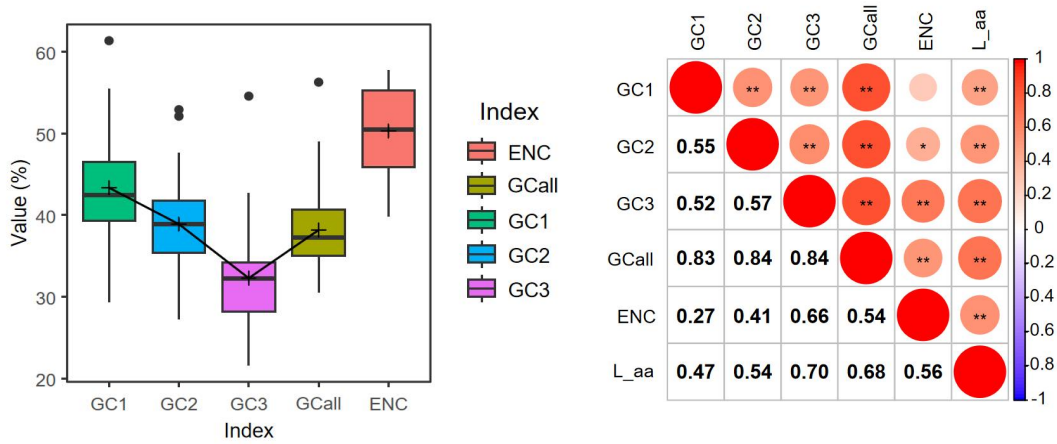
2 结果与分析

2.1 密码子偏好相关指标分析

紫叶苔线粒体基因组 CDS 序列的碱基组成分析显示, A3s、T3s、C3s 和 G3s 分别为 40.93%、45.44%、18.51% 和 19.80%, A3s 和 T3s 远高于 C3s 和 G3s, 表明紫叶苔线粒体基因组密码子的第 3 位置碱基以 A/U 结尾为主。在密码子不同位置的碱基 GC 含量分布如图 1 a) 所示, GC1、GC2、GC3 和 GCall 的分布区间分别为 29.29%~61.36%、27.18%~52.91%、21.57%~54.57% 和 30.46%~56.28%。不同位置碱基 GC 含量的平均值大小如下: $GC1(43.35\%) > GC2(38.90\%) > GCall(38.17\%) > GC3(32.28\%)$ 。结果

表明, 紫叶苔线粒体基因组密码子不同位置的 GC 含量分布不均衡, 且偏好使用 A/U 碱基, 在密码子的第三位上表现尤为明显。不同编码基因的 ENC 值分布在 39.80~57.79 区间, 平均值为 50.29。此外, 密码子适应指数(CAI)集中在 0.13~0.21 之间, 平均值为 0.17, 远小于 1。上述结果表明, 紫叶苔线粒体基因组存在密码子使用偏好性, 但整体偏倚程度较弱, 基因表达水平较低。

紫叶苔线粒体密码子偏好性主要指标的相关性分析如图 1 b) 所示, 密码子的 GC1 和其他不同位置的 GC 含量及 ENC 均呈极显著正相关, GC3 与 ENC 呈极显著正相关($r=0.66, p<0.01$), 且二者的相关系数在不同位置碱基的 GC 含量与 ENC 的相关系数中最大。该结果表明 GC 含量, 尤其是密码子第 3 位置的 GC 含量对密码子的偏好使用存在显著影响。此外, ENC 与基因编码氨基酸的长度 L_{aa} 有极显著正相关($r=0.56, p<0.01$), 表明在紫叶苔线粒体中, 基因的长度影响密码子偏好性。



a) GC1、GC2、GC3、GC11 和 ENC 的分布, 平均值的位置用“+”表示

b) 紫叶苔线粒体基因组密码子偏好主要指标的相关性分析

图 1 紫叶苔线粒体基因组密码子相关指标的分布和相关性分析

Fig.1 Distribution and correlation analysis of codon-related indicators in the mitochondrial genomes of *P. purpurea*

紫叶苔线粒体基因组蛋白编码区密码子的 RSCU 值分布如图 2 和表 1 所示, 不同密码子的 RSCU 值差异较大, 其中编码亮氨酸的密码子 UUA 的 RSCU 值最高, 为 1.91, 而另一个同样编码亮氨酸的密码子 CUC 的 RSCU 值仅为 0.41。此外, 共鉴定出 27 个高频密码子, 其 RSCU 值大于 1。在这 27 个密码子中, 以 U、A、G 和 C 结尾的密码子数分别为 16、10、1 和 0, 这表明紫叶苔线粒体基因组的高频密码子在使用偏好上多以 A 或 U 结尾。

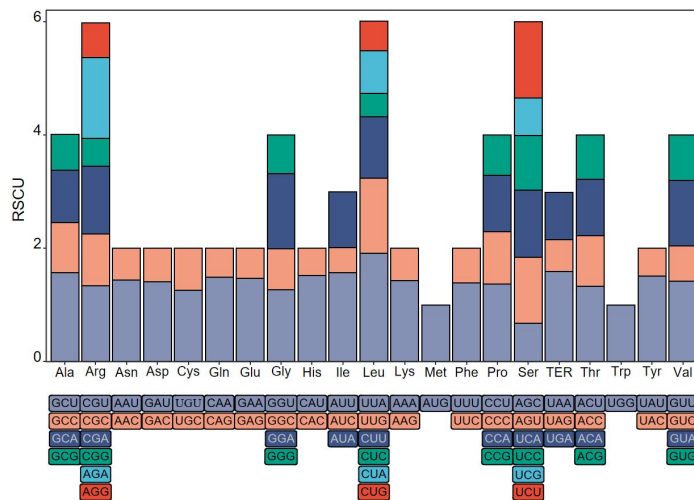


图 2 紫叶苔线粒体基因组蛋白编码区密码子的 RSCU 分析

Fig.2 Analysis of RSCU in the protein-coding regions of the mitochondrial genome of *P. purpurea*

2.2 中性绘图分析

GC3 和 GC12 之间的关系用中性图(neutrality plot)分析(见图 3)。GC3 的范围为 0.2157~0.5457, GC12 的范围为 0.3306~0.5714, 平均值分别为 0.322 8 和 0.411 2。GC12 和 GC3 之间存在显著的相关性($r=0.6203$, $p<0.01$), 说明基因组中大多数编码基因的前 2 个密码子碱基与第 3 个碱基在组成上不存在显著差异, 定向突变压力作用于所有密码子位置。回归直线的斜率(回归系数)为 0.534 0, 表明突变压力对密码子偏好效应的贡献率为 53.40%, 自然选择等因素的贡献率为 46.60%。根据中性绘图可知, 突变压力和自然选择共同塑造了紫叶苔线粒体基因组的密码子偏好, 其中突变压力占主要作用。

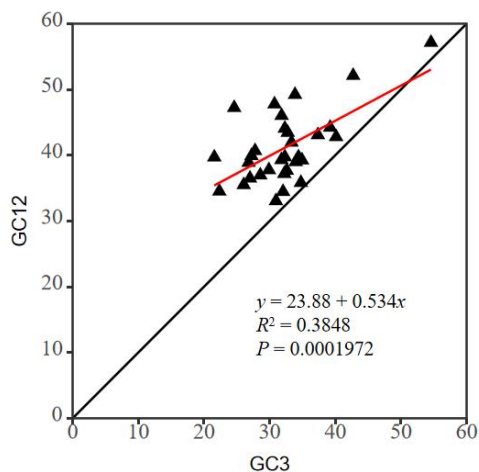
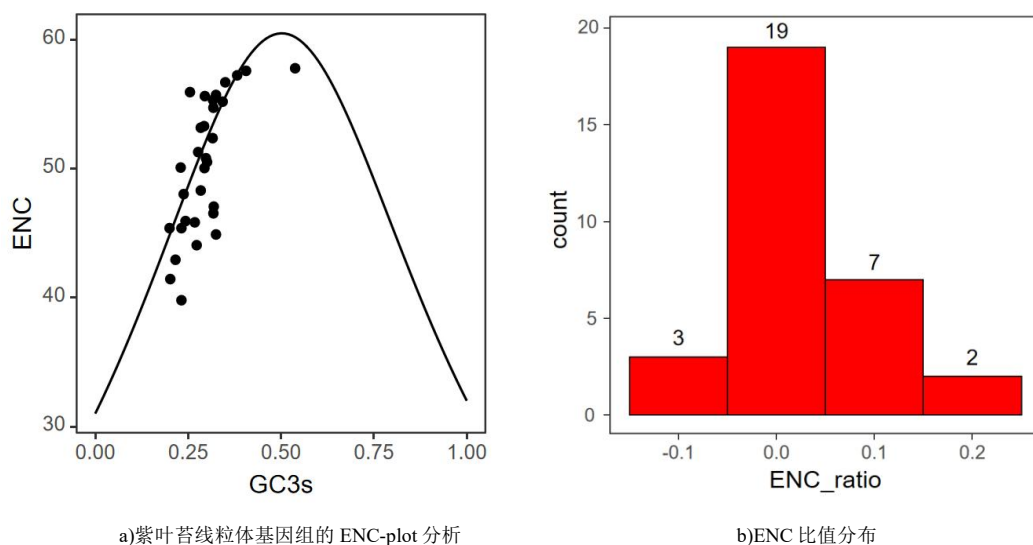


图 3 紫叶苔线粒体基因组的中性绘图分析

Fig.3 Neutral plot analysis of the mitochondrial genomes of *P. purpurea*

2.3 ENC 曲线分析

图 4 中 ENC-plot 结果表明, 在紫叶苔线粒体基因组蛋白编码序列中, 基因的 ENC 值分布较聚集, 大量的基因落在或靠近 ENC 标准曲线, 而较少数的基因离标准曲线较远。从 ENC 比值频率统计分布图中也可以看出: ENC 比率从 -0.15 到 0.25 不等; 在 31 个蛋白编码基因中, 靠近 ENC 标准曲线的区间(-0.05~0.05)内共分布 19 个基因(61.29%), 其实际和预期的 ENC 值差异较小; 有 9 个基因(29.03%)的 ENC 比率大于 0.05, 反映了这些基因不仅分布在标准曲线下方且离标准曲线较远。以上结果进一步表明, 紫叶苔线粒体基因组的密码子偏好性更多地受到突变压力的作用, 自然选择的影响次之。



a)紫叶苔线粒体基因组的 ENC-plot 分析

b)ENC 比值分布

图 4 紫叶苔线粒体基因组的 ENC-plot 分析和 ENC 比值分布

Fig.4 ENC-plot analysis and ENC ratio distribution of the mitochondrial genomes of *P. purpurea*

2.4 PR2-plot 分析

为进一步评估紫叶苔线粒体基因 4 种碱基的偏倚关系，并了解突变和选择压力对密码子使用的影响，对同义密码子的第 3 位碱基进行了 PR2-plot 分析。如图 5 所示，PR2 水平 4 个区域的基因分布不均匀。在垂直方向上，大部分基因集中在平面的下半区域，表明在密码子的同义位置上 T 的使用频率高于 A；在水平方向上，更多的基因分布在平面的右侧，说明 G 的使用频率高于 C。因此，紫叶苔线粒体基因组同义密码子的第 3 位置碱基在选择上具有偏好性。根据碱基的不均衡使用，以及少数基因显示出接近中心的位置，可以推测自然选择与突变压力可能共同影响了紫叶苔线粒体基因的密码子使用偏好。

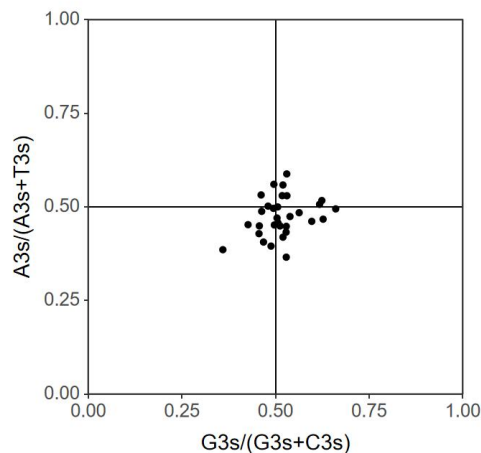


图 5 紫叶苔线粒体基因组的 PR2-plot 分析

Fig.5 PR2-plot analysis of the mitochondrial genomes of *P. purpurea*

2.5 最优密码子筛选

按照 ENC 值的大小顺序，分别选取 3 个基因(*rps14*、*rpl6*、*cox2*)作为高表达基因库，3 个基因(*ccmFC*、*rpl2*、*rtl*)作为低表达基因库。按照最优密码子的标准(RSCU > 1 且 Δ RSCU > 0.08)，共筛选出 23 个最优密码子(见表 1)，其中以 U 结尾的密码子有 14 个，以 A 结尾的密码子有 8 个，以 G 结尾的密码子有 1 个，而以 C 结尾的密码子 0 个。该结果进一步表明了紫叶苔线粒体基因的最优密码子偏好以 A 或 U 结尾。

表 1 紫叶苔线粒体基因组最优密码子筛选

Tab.1 Optimal codon selection in the mitochondrial genome of *P. purpurea*

氨基酸	密码子	基因组 RSCU	高表达基因 RSCU	低表达基因 RSCU	Δ RSCU	氨基酸	密码子	基因组 RSCU	高表达基因 RSCU	低表达基因 RSCU	Δ RSCU
TER	<u>UAA</u>	1.59	1.00	2.00	-1.00	Arg	<u>AGA</u>	1.42	1.11	1.45	-0.34
	UAG	0.56	2.00	1.00	1.00		AGG	0.62	0.44	0.96	-0.52
	UGA	0.84	0.00	0.00	0.00		<u>CGA</u>	1.20	1.56	1.05	0.51
Ala	GCC	0.88	1.39	1.38	0.01		CGC	0.91	0.89	1.25	-0.36
	GCG	0.63	0.35	0.76	-0.41		CGG	0.49	0.22	0.73	-0.51
	GCA	0.93	0.35	0.79	-0.44		<u>CGU</u>	1.34	1.78	0.56	1.22
	<u>GCU</u>	1.57	1.91	1.07	0.84	Asp	GAC	0.59	0.80	1.02	-0.22
Cys	UGC	0.74	0.33	1.08	-0.75		<u>GAU</u>	1.41	1.20	0.98	0.22
	<u>UGU</u>	1.26	1.67	0.92	0.75	Asn	AAC	0.56	0.62	0.75	-0.13
Gly	<u>GGA</u>	1.33	1.45	1.06	0.39		<u>AAU</u>	1.44	1.38	1.25	0.13
	GGC	0.72	0.36	1.14	-0.78	Glu	<u>GAA</u>	1.47	1.47	1.37	0.10
	GGG	0.68	0.73	0.89	-0.16		GAG	0.53	0.53	0.63	-0.10
	<u>GGU</u>	1.27	1.45	0.91	0.54	Gln	<u>CAA</u>	1.49	1.62	1.24	0.38
Ile	AUA	0.99	0.82	1.22	0.40		CAG	0.51	0.38	0.76	-0.38
	AUC	0.44	0.20	0.49	-0.29	Lys	<u>AAA</u>	1.43	1.67	1.14	0.53
	<u>AUU</u>	1.57	1.98	1.29	0.69		AAG	0.57	0.33	0.86	-0.53
Leu	CUA	0.75	0.30	1.11	-0.81	Met	AUG	1.00	1.00	1.00	0.00
	CUC	0.41	0.00	0.72	-0.72	Pro	CCA	1.00	1.52	0.61	0.91
	CUG	0.53	0.15	0.75	-0.60		CCC	0.92	0.19	1.19	-1.00
	<u>CUU</u>	1.08	1.35	1.33	0.02		CCG	0.71	0.00	0.94	-0.94
	<u>UUA</u>	1.91	2.25	0.90	1.35		<u>CCU</u>	1.37	2.29	1.26	1.03
	<u>UUG</u>	1.33	1.95	1.19	0.76	Thr	ACA	1.00	0.71	0.84	-0.13
His	CAC	0.48	0.31	0.81	-0.50		ACC	0.89	0.24	0.89	-0.65

Ser	<u>CAU</u>	1.52	1.69	1.19	0.50	ACG	0.78	0.00	1.38	-1.38	
	AGC	0.68	0.00	0.84	-0.84	<u>ACU</u>	1.33	3.06	0.89	2.17	
	<u>AGU</u>	1.16	1.74	0.89	0.85	Tyr	UAC	0.49	0.10	1.09	-0.99
	<u>UCA</u>	1.19	1.16	0.62	0.54	<u>UAU</u>	1.51	1.90	0.91	0.99	
	UCC	0.96	1.35	1.11	0.24	Trp	UGG	1.00	1.00	1.00	0.00
	UCG	0.66	0.39	1.11	-0.72	Val	<u>GUA</u>	1.16	1.60	0.96	0.64
Phe	<u>UCU</u>	1.35	1.35	1.42	-0.07	GUC	0.62	0.40	0.58	-0.18	
	UUC	0.61	0.40	0.80	-0.40	GUG	0.8	0.27	1.06	-0.79	
	<u>UUU</u>	1.39	1.60	1.20	0.40	<u>GUU</u>	1.42	1.73	1.40	0.33	

注: 高频密码子(RSCU > 1) 带下划线; 加粗的密码子表示最优密码子。

3 讨 论

在基因编码中, 生物体倾向于选择某些特定的密码子, 即密码子使用偏好(CUB), 对 CUB 的解析有助于理解基因调控和进化过程。密码子的使用偏好不同基因组、基因之间以及同一基因内部都存在差异^[30], 同一物种的核基因组和细胞器基因组的 CUB 往往也有较大差别。植物具有较高的 GC 含量, 其中单子叶植物的核基因编码区富含以 C/G 结尾的密码子, 双子叶植物则有较高比例的以 A/U 结尾的密码子^[24], 叶绿体和线粒体基因组在密码子使用上更倾向于使用 A/U 结尾的密码子。在对多种植物的叶绿体基因组进行的研究中, 如水稻属^[26]、胡颓子属^[31]、柏科^[32]、枸杞^[33]等, 都发现了这种密码子偏好性。目前, 对植物线粒体基因组 CUB 模式的研究还相对有限, 但已有研究^[19,34-35]也揭示了相似的规律, 为理解植物基因组的进化提供了宝贵信息。

本文对紫叶苔线粒体基因组蛋白编码序列进行了全面的密码子使用偏好性(CUB)分析。密码子组成计算结果显示, 紫叶苔线粒体基因组的平均 GC 含量和密码子不同位置的 GC 含量均低于 50%, 表明紫叶苔线粒体基因组倾向于富含 A/T 的核苷酸, 偏好以 A/U 结尾的密码子, 且基因组中的高频密码子和最优密码子也主要以 A/U 结尾。通过对维管植物、苔藓植物和藻类线粒体基因组的密码子偏好进行初步比较^[19-20,34], 发现绿色植物各类群在密码子偏好上既表现出共性特征, 也存在动态变化。其中大多数物种偏好以 A/U 结尾的密码子, 但这种偏好的程度在不同类群之间以及类群内部又表现出一定的变异性。总体来看, 苔藓植物对 A/U 结尾密码子的偏好低于水生轮藻类, 而高于在陆地繁盛的维管植物, 这似乎也暗示了苔藓植物在植物界中从水生到陆生过渡的进化地位。地钱门、真藓门和角苔门在线粒体基因组的密码子偏好上也存在一定差异, 其中地钱门在密码子末端位点对 A/U 的偏好最弱, 而真藓门的偏好性较强^[34]。本文的结果同样显示, 紫叶苔线粒体基因组密码子对 A/U 具有较弱的偏好性, 为陆生植物早期分化类群在基因组组成上的共性特征提供了又一证据。植物在线粒体基因组密码子偏好上的相似性可能与植物线粒体的共同起源有关, 而偏好性强度的变化可能与不同植物类群对陆生环境变化的适应性密切相关。例如, 在陆生植物谱系的进化过程中, 线粒体基因组中的密码子使用逐渐从强烈的 AT 偏好转变为越来越多的 GC 积累, 这可能有助于减少 DNA 在强紫外线辐射下的损伤^[20], 从而更好地适应干燥的环境。

在基因组学研究领域, 对密码子使用偏好(CUB)的形成机制进行了广泛探讨, 其中自然选择理论与中性理论得到了学术界的普遍认可^[36]。在研究密码子使用偏好的成因时, 中性图是一种重要的分析工具, 它能够揭示自然选择和突变压力对 CUB 的相对影响^[37]。紫叶苔线粒体基因组密码子中性绘图结果显示, GC3 与 GC12 显著相关, 表明定向突变压力在整体密码子组成上发挥了重要作用; 同时回归直线的斜率(回归系数)为 0.534 0, 进一步表明, 相较于自然选择等因素, 突变压力对于密码子的组成成分发挥了更为重要的作用。为了进一步说明同义密码子使用偏好性, 基于同义密码子第 3 位置的 GC 含量(GC3s)绘制的 ENC-plot 图, 结果显示大部分基因接近预期曲线, 表明这些基因的密码子偏好性主要受突变压力的制约, 自然选择的作用次之。在 PR2-plot 分析中, 只有少量基因位于或接近中心坐标(0.5, 0.5), 大部分基因散落在 $A3/(A3+T3)<0.5$ 和 $G3/(G3+C3)>0.5$ 的区域范围内, 显示在同义位点上不同碱基的使用频率存在差异。这也表明, 在形成紫叶苔线粒体基因组的 CUB 时, 除了受到中性突变压力的影响外, 还受到自然选择及其他因素的影响。已知植物线粒体基因组密码子的偏好性可能以突变或自然选择作为主要的驱动力, 或同时受到其他多种因素的共同作用^[30]。然而目前针对苔藓植物的相关研究极少, 有待加大对该植物类群线粒体基因组的探讨, 寻找更广泛的密码子偏好规律, 为揭示植物类群的系统发育关系, 以及理解其进化历程等提供更全面深入的见解。

此外, 本研究从紫叶苔线粒体基因组中共筛选出 23 个最优密码子。这些密码子的大多数在第 3 位置碱基上或以 U 结尾, 或以 A 结尾, 更进一步表明紫叶苔线粒体基因组对于 A/U 结尾密码子的偏好, 而这种偏好 NNA 或 NNU 密码子的使用模式也与之之前其他植物细胞器基因组最优密码子的筛选结果一致^[38-39]。基于紫叶苔线粒体基因组密码子使用模式进行外源序列优化, 可以在后续以基因工程为手段的紫叶苔属遗传发育、物种保护等研究工作中, 用于改善基因表达以及提高翻译效率。

4 结语

本研究基于紫叶苔线粒体基因组 CDS 序列, 对紫叶苔线粒体基因组的密码子组成、中性绘图、ENC-plot 绘图、PR2-plot 绘图等进行了分析, 主要结论如下: 紫叶苔线粒体基因组 CDS 序列倾向于富含 A/T 的核苷酸和以 A/U 结尾的密码子, 但偏倚程度较弱; 其同义密码子的使用偏好性除主要受到突变压力的影响外, 还受到自然选择、碱基组成和基因长度等因素的影响; 筛选出了 23 个最优密码子, 可用于紫叶苔分子发育和遗传保护等方面的研究。

本文仅针对紫叶苔线粒体基因组的蛋白编码序列进行了密码子偏好分析, 后续将进一步研究非编码序列如内含子数量等因素对密码子偏好性的影响, 还将增加苔藓及其他不同植物类群的物种作为线粒体基因组密码子偏好性的研究对象, 结合系统发育树深入探讨密码子偏好与进化的关系。

参考文献/References:

- [1] CHEN X J, BUTOW R A. The organization and inheritance of the mitochondrial genome[J]. *Nature Reviews Genetics*, 2005, 6(11): 815-825.
- [2] TAANMAN J W. The mitochondrial genome: structure, transcription, translation and replication[J]. *Biochimica et Biophysica Acta*, 1999, 1410(2): 103-123.
- [3] MCCLELLAN D A. The codon-degeneracy model of molecular evolution[J]. *Journal of Molecular Evolution*, 2000, 50: 131-140.
- [4] BIRO J C. Does codon bias have an evolutionary origin ?[J]. *Theoretical Biology and Medical Modelling*, 2008, 5: 16. doi: 10.1186/1742-4682-5-16.
- [5] SHARP P M, Bailes E, Grocock R J, et al. Variation in the strength of selected codon usage bias among bacteria[J]. *Nucleic Acids Research*, 2005, 33(4): 1141-1153.
- [6] LIN K, KUANG Y Y, JOSEPH J S, et al. Conserved codon composition of ribosomal protein coding genes in *Escherichia coli*, *Mycobacterium tuberculosis* and *Saccharomyces cerevisiae*: lessons from supervised machine learning in functional genomics[J]. *Nucleic Acids Research*, 2002, 30(11): 2599-2607.
- [7] NOVEMBRE J A. Accounting for background nucleotide composition when measuring codon usage bias[J]. *Molecular Biology and Evolution*, 2002, 19(8): 1390-1394.
- [8] BARBHUIYA P A, UDDIN A, CHAKRABORTY S. Genome-wide comparison of codon usage dynamics in mitochondrial genes across different species of amphibian genus *Bombina*[J]. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 2019, 332(3-4): 99-112.
- [9] SÖDERSTRÖM L, HAGBORG A, VON KONRAT M, et al. World checklist of hornworts and liverworts[J]. *PhytoKeys*, 2016, 59: 1-828.
- [10] CRANDALL-STOTLER B. The apical cell and early development of *Pleurozia purpurea* Lindb.[J]. *Lindbergia*, 1976, 3: 197-208.
- [11] FORREST L L, DAVIS E C, LONG D G, et al. Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses[J]. *The Bryologist*, 2006, 109(3): 303-334.
- [12] LI Y F, LUO L, LIU Y, et al. The bryophyte phylogeny group: a revised familial classification system based on plastid phylogenomic data[J]. *Journal of Systematics and Evolution*, 2024. doi: 10.1111/jse.13063.
- [13] THIERS B M. A monograph of *Pleurozia* (Hepaticae; Pleuroziaceae)[J]. *The Bryologist*, 1993, 96(4): 517-554.
- [14] DONG S S, LIU Y. The mitochondrial genomes of bryophytes[J]. *Bryophyte Diversity and Evolution*, 2021, 43(1): 112-126.
- [15] PUTTICK M N, MORRIS J L, WILLIAMS T A, et al. The interrelationships of land plants and the nature of the ancestral embryophyte[J]. *Current Biology*, 2018, 28: 733-745.
- [16] WANG B, XUE J Y, LI L B, et al. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts[J]. *Current Genetics*, 2009, 55: 601-609.
- [17] LIU Y, MEDINA R, GOFFINET B. 350 my of mitochondrial genome stasis in mosses, an early land plant lineage[J]. *Molecular Biology and Evolution*, 2014, 31(10): 2586-2591.
- [18] DONG S S, ZHAO C X, ZHANG S Z, et al. Mitochondrial genomes of the early land plant lineage liverworts (Marchantiophyta): conserved genome

- structure, and ongoing low frequency recombination[J]. *BMC Genomics*, 2019, 20: 953. doi.org/10.1186/s12864-019-6365-y.
- [19] ZHOU M, LI X. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes[J]. *Molecular Biology Reports*, 2009, 36: 2039-2046.
- [20] WANG B, YUAN J, LIU J, et al. Codon usage bias and determining forces in green plant mitochondrial genomes[J]. *Journal of Integrative Plant Biology*, 2011, 53(4): 324-334.
- [21] 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, 2653-2657.
- [22] ZHANG K, WANG Y H, ZHANG Y, et al. Codon usage characterization and phylogenetic analysis of the mitochondrial genome in *Hemerocallis citrina*[J]. *BMC Genomic Data*, 2024, 25: 6. doi.org/10.1186/s12863-024-01191-4.
- [23] WRIGHT F. The 'effective number of codons' used in a gene[J]. *Gene*, 1990, 87: 23-29.
- [24] PARVATHY S T, UDAYASURIYAN V, BHADANA V. Codon usage bias[J]. *Molecular Biology Reports*, 2022, 49: 539-565.
- [25] SUEOKA N. Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G+C content of third codon position[J]. *Gene*, 1999, 238: 53-58.
- [26] CHAKRABORTY S, YENKHO S, UDDIN A. Analysis of codon usage bias of chloroplast genes in *Oryza* species[J]. *Planta*, 2020, 252(4): 67. doi: 10.1007/s00425-020-03470-7.
- [27] SHIELDS D C, SHARP P M, HIGGINS D G, et al. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons[J]. *Molecular Biology and Evolution*, 1988, 5(6): 704-716.
- [28] WANG H J, MENG T, WEI W Q. Analysis of synonymous codon usage bias in *helicase* gene from *Autographa californica* multiple nucleopolyhedrovirus[J]. *Genes and Genomics*, 2018, 40(7): 767-780.
- [29] WU P, XIAO W Q, LUO Y Y, et al. Comprehensive analysis of codon bias in 13 *Ganoderma* mitochondrial genomes[J]. *Frontiers in Microbiology*, 2023, 14: 1170790. doi: 10.3389/fmicb.2023.1170790.
- [30] SALIM H M W, CAVALCANTI A R O. Factors influencing codon usage bias in genomes[J]. *Journal of the Brazilian Chemical Society*, 2008, 19(2): 257-262.
- [31] LI C L, ZHOU L, NIE J B, et al. Codon usage bias and genetic diversity in chloroplast genomes of *Elaeagnus* species (Myrtiliflorae: Elaeagnaceae)[J]. *Physiology and Molecular Biology of Plants*, 2023, 29(2): 239-251.
- [32] 黄思琦, 张麒功, 叶泽霖, 等. 5种柏科植物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)*, 2024, 53(2): 214-220.
- HUANG S Q, ZHANG Q G, YE Z L, et al. Codon bias analysis of chloroplast genomes of 5 Cupressaceae plants[J]. *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, 2024, 53(2): 214-220.
- [33] 杜雨, 李效雄, 贾西贝, 等. 枸杞叶绿体基因组密码子偏好性分析[J]. *中草药*, 2024, 55(4): 1316-1325.
- DU Y, LI X X, JIA X B, et al. Codon preference analysis of chloroplast genome of Chinese wolfberry[J]. *Chinese Traditional and Herbal Drugs*, 2024, 55(4): 1316-1325.
- [34] XU W J, XING T, ZHAO M M, et al. Synonymous codon usage bias in plant mitochondrial genes is associated with intron number and mirrors species evolution[J]. *PLoS ONE*, 2015, 10(6): e0131508. doi:10.1371/journal.pone.0131508.
- [35] LIU Q P, FENG Y, XUE Q Z. Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa*[J]. *Mitochondrion*, 2004, 4: 313-320.
- [36] 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.
- [37] LI T T, MA Z, DING T M, et al. Codon usage bias and phylogenetic analysis of chloroplast genome in 36 Gracilariaceae species[J]. *Functional and Integrative Genomics*, 2024, 24(2): 45. doi: 10.1007/s10142-024-01316-z.
- [38] 耿晓姗, 贾魏, 陈佳宁, 等. 金花茶叶叶绿体基因组密码子偏好性分析[J]. *分子植物育种*, 2022, 20(7): 2196-2203.
- GENG X S, JIA W, CHEN J N, et al. Codon usage bias analysis of chloroplast genome in *Camellia nitidissima*[J]. *Molecular Plant Breeding*, 2022, 20(7): 2196-2203.
- [39] 罗永坚, 王茹, 赵仁菲, 等. 珙桐叶绿体基因组同义密码子使用偏好性分析[J]. *北京林业大学学报*, 2024, 46(3): 8-16.
- LUO Y J, WANG R, ZHAO R F, et al. Analysis of synonymous codon usage bias in the chloroplast genome of *Davidia involucrata*[J]. *Journal of Beijing Forestry University*, 2024, 46(3): 8-16.