

侯 哲, 姜晓鸣, 李 昂, 等. 11 种唐松草属(*Thalictrum*)*rbcL* 基因的密码子偏好性研究[J]. 江苏农业科学, 2023, 51(3): 46–53.
doi:10.15889/j.issn.1002-1302.2023.03.007

11 种唐松草属(*Thalictrum*)*rbcL* 基因的密码子偏好性研究

侯 哲, 姜晓鸣, 李 昂, 黄长兵

(苏州农业职业技术学院, 江苏苏州 215008)

摘要:唐松草属(*Thalictrum*)是毛茛科(Ranunculaceae)的重要植物类群,为了明确唐松草属*rbcL*基因的密码子使用偏好性,从NCBI下载了11条唐松草属植物的*rbcL*序列,并对其密码子偏好性进行了分析。研究表明,*rbcL*基因的密码子中,GC₁的平均值为0.583,GC₂与GC₃分别为0.435、0.310,其中,GC₁含量最大,GC₃含量最小,表现出显著的差异;ENC值与GC₁和GC₂的相关性不显著,而与GC₃的相关系数为0.335($P < 0.05$),说明在*rbcL*基因中,密码子第3位的碱基组成与密码子的偏好性显著相关;中性绘图分析表明,GC₃与GC₁₂的相关性不显著,且选择对密码子的偏好性有着决定性的作用;ENC-plot及PR-plot结果表明,*rbcL*基因的ENC值低于预期值,且相比于嘌呤C/T,嘧啶A/G的使用频率要更低;RSCU分析表明,RSCU值大于1的密码子大多数以A/U结尾,密码子使用度最高的3个密码子为CGU、UCU、ACU,用RSCU值做聚类分析,在一定程度上也可以反映唐松草属植物的亲缘关系。对于唐松草属植物的*rbcL*基因来说,酵母可以作为其表达受体,拟南芥可作为其转化受体。本研究为唐松草属叶绿体基因工程的开展奠定了基础,为其后续的遗传多样性、遗传结构及基因功能提供了借鉴信息。

关键词:唐松草属;*rbcL*基因;密码子;密码子偏好性

中图分类号:S184 **文献标志码:**A **文章编号:**1002-1302(2023)03-0046-08

植物的生长发育离不开光合作用,而光合作用最重要的场地在叶绿体中。*rbcL*基因在光合作用的整个过程中起到至关重要的作用,如*rbcL*基因可以编码Rubisco酶中的大亚基,从而催化光合作用的核心过程:羧化反应与加氧反应。由此可以看出,*rbcL*基因对植物的净光合率有着十分重要的影响^[1]。*rbcL*基因在不同植物、不同自然条件下会产生不同的耐受性,因此基因的表达也会各不相同,这对植物的光合作用也会产生一定的影响^[2]。所以,了解其密码子偏好性可以更好地理解不同植物在不同自然条件下是如何适应与进化的。因此,对于*rbcL*基因密码子偏好性的研究是非常有意义的^[3]。

充分了解植物的密码子偏好性为理解物种进化提供了重要的信息^[3]。密码子是指在mRNA上每3个相邻核苷酸排列成的三联体,一种密码子决

定一种氨基酸,利用A、U、G、C,可以组成64种密码子,除UAA、UGA和UAG外的61种密码子用来编码常见的20种氨基酸^[4]。而一种氨基酸可以由一至多个密码子编码,由于基因突变和自然选择的存在,在不同物种翻译过程中,氨基酸的编码中出现偏爱某些密码子的现象,称为密码子的偏好性^[5]。漂变、突变、基因表达水平、自然选择及tRNA丰度等都会对物种的密码子使用情况产生影响^[6-9]。

唐松草属植物在全世界大约有150种,在中国各地有76种,其中约30种由于富含生物碱、三萜、酚类及黄酮等具有药理活性的化合物,因而具有非常高的药用价值,根和茎常被作为传统的民间药物,用于治疗流感、癌症、高血压、细菌感染及结膜炎等症状^[10]。目前关于唐松草属的研究内容较少,且多是对其化学成分的研究^[10]。迄今为止,有关唐松草属叶绿体基因组方面的研究还比较少,对其*rbcL*基因密码子偏好性的研究也未见报道。因此,笔者从NCBI上下载唐松草属目前已公布的所有叶绿体基因组,并从中获取*rbcL*的基因序列,深度分析其序列特征,并对其密码子使用情况进行分析,优选出使用频率最高的密码子,为其后续的叶绿体基因组应用、种质资源保护及基因功能等研究提供理论支撑。

收稿日期:2022-05-23

基金项目:江苏省特色花卉工程中心项目(编号:苏发改高技发[2020]1460号);江苏省省级作物种质资源库(球宿根花卉)项目(编号:JS-ZW-K05);四川省自然科学基金(编号:2022NSFSC1141)。

作者简介:侯 哲(1991—),男,山西运城人,博士,讲师,从事园林植物抗逆性分子机制研究。E-mail:houshedota@163.com。

通信作者:黄长兵,硕士,副研究员,从事观赏植物新品种选育与抗性研究。E-mail:cbhuang@szai.edu.cn。

1 材料与方 法

1.1 唐松草属植物 11 条 *rbcL* 基因序列获取

于 2022 年 5 月 20 日,在苏州农业职业技术学院植物生理实验室开展本研究,从 NCBI 公共数据库上下载唐松草属目前已知的所有叶绿体基因组序列,并从蛋白编码序列文件中获取 *rbcL* 基因序列,用于后续密码子偏好性分析。

1.2 密码子组成分析

利用 CodonW 软件(<http://codonw.sourceforge.net>)分析 11 条序列的密码子偏好参数:(1)密码子的出现次数(N);(2)有效密码子数(ENC);(3)计算出 RSCU 值,并将 RSCU > 1 的密码子过滤到高频密码子中,并分别用 TTools^[11]及 R 语言绘图,对 RSCU 的聚类分析在 SPSS 在线软件(<https://spssau.com/help.html>)中进行。使用 EMBOSS 在线网站(<https://www.bioinformatics.nl/emboss-explorer/>)分析星毛唐松草叶绿体基因组中各个密码子的 GC₁、GC₂、GC₃ 和 GC_{all},最后用 SPSS 在线软件(<https://www.spssau.com/102000000>)对各个参数进行了相关性分析。

1.3 中性绘图分析

通过中性绘图分析,可以对密码子的使用偏好性的影响因素进行分析^[12]。以 GC₃ 为横坐标,GC₁ 与 GC₂ 的平均值为纵坐标,然后绘制二维散点图并对二者进行相关性分析。如果图中的分散点沿对角线分布集中,则回归系数接近 1,表明密码子偏好性主要受突变的影响。反之,如果图中的分散点不沿对角线分布,说明选择对该基因的密码子偏好性产生了重要影响^[12]。

1.4 ENC - plot 绘图分析

ENC 既可直观反映基因密码子使用模式,也可以用于判断突变或选择对基因密码子偏好性的影响,ENC 值一般在 20 ~ 61 区间。当 ENC 接近 20 时,表明突变是密码子偏好的主要影响因素,而当 ENC 接近 61 时,说明选择起到了决定性作用^[13]。以 11 条 *rbcL* 基因序列中的 ENC 值为纵坐标,GC₃ 作为横坐标,然后用 R 语言作图,构建 ENC 值的期望曲线,并结合频率对差异进行一定量化分析。

1.5 PR2 - plot 绘图

PR2 - plot 分析,根据每个密码子的第 3 位数计算 A、T、C 和 G 的含量。首先横坐标选取 A₃/(A₃ + T₃) 值,而纵坐标为 G₃/(G₃ + C₃) 值,然后绘制

PR2 - plot 散点图。

1.6 聚类分析

11 条基因通过 MAFFT 软件^[14] 比对后,用 MEGA 7.0^[15] 软件的邻接法(NJ)生成系统进化树。

1.7 密码子使用频率比较

通过计算唐松草属植物与模式植物基因组序列的密码子使用频率的比值,可以得出它们与模式植物的密码子使用偏好性的不同程度,当比值在 0.5 ~ 2.0 区间内,则差异较小^[16]。在密码子使用数据库中分别下载几种模式植物的密码子使用频率,用在线软件 CUSP(<https://www.bioinformatics.nl/emboss-explorer/>)计算唐松草属植物的密码子使用频率后,计算相对使用频率,再用 Ttools^[11] 绘图。

2 结果与分析

2.1 唐松草属 *rbcL* 基因信息

从公共数据库 NCBI 中获取 11 条唐松草属植物的 *rbcL* 基因信息(表 1)。

表 1 11 种唐松草属植物 *rbcL* 基因信息

| 物种 | 登录号 | CDS 位置 |
|--|-----------|-------------|
| 贝加尔唐松草(<i>Thalictrum baicalense</i>) | MW133265 | 56461—57888 |
| 唐松草叶银莲花(<i>Thalictrum thalictroides</i>) | NC_039433 | 56243—57670 |
| 亚欧唐松草(<i>Thalictrum minus</i>) | NC_041544 | 56723—58150 |
| 丝叶唐松草(<i>Thalictrum foeniculaceum</i>) | NC_053570 | 56470—57897 |
| 粘唐松草(<i>Thalictrum viscosum</i>) | NC_058831 | 56475—57902 |
| 欧洲唐松草(<i>Thalictrum aquilegifolium</i>) | NC_058830 | 56740—58167 |
| 瓣蕊唐松草(<i>Thalictrum petaloideum</i>) | MK253449 | 56421—57848 |
| 细唐松草(<i>Thalictrum tenue</i>) | MK253448 | 56553—57980 |
| 多叶唐松草(<i>Thalictrum foliolosum</i>) | NC_058920 | 56344—57771 |
| 星毛唐松草(<i>Thalictrum cirrhosum</i>) | OL711693 | 56453—57880 |
| 朝鲜唐松草(<i>Thalictrum coreanum</i>) | NC_026103 | 56274—57701 |

2.2 密码子相关偏性指数分析

为了研究唐松草属植物中 *rbcL* 基因密码子使用偏差的程度,计算了 11 个基因的有效密码子数(ENC)的值(表 2)。ENC 值的范围为 50.531 ~ 50.992,平均值为 50.757,其中细唐松草的 *rbcL* 基因的 ENC 值最小,而贝加尔唐松草的 ENC 值最大,虽然不同植物的 ENC 值各不相同,但是差别不是很大(表 2),说明 11 种唐松草植物的 *rbcL* 基因密码子的偏好性不强。此外,笔者发现不同基因的 GC₁、GC₂、GC₃ 含量存在明显的差异,各基因总体的 GC 含量为 0.445,表明 *rbcL* 基因偏好使用 A 与 T 碱基。而 GC₁、GC₂、GC₃ 的平均含量分别为 0.583、0.435、

0.321。其中,第 1 位密码子的 GC 含量最大,而第 3 位密码子的 GC 含量最小,表现出明显的差异。第 3 位密码子的 GC 含量差异最明显,而 GC₃ 也是评估

密码子偏好性的重要指标。第 3 位密码子 A、T、C、G 的含量分别为 0.387、0.477、0.155、0.218,表明 *rbcL* 基因偏好使用 A 与 T 结尾的密码子。

表 2 11 种唐松草属植物 *rbcL* 基因的碱基类别及有效密码子(ENC)

| 植物 | A ₃ | T ₃ | G ₃ | C ₃ | GC | GC ₁ | GC ₂ | GC ₃ | ENC |
|---------|----------------|----------------|----------------|----------------|-------|-----------------|-----------------|-----------------|--------|
| 贝加尔唐松草 | 0.393 | 0.474 | 0.150 | 0.220 | 0.445 | 0.584 | 0.431 | 0.320 | 50.992 |
| 唐松草叶银莲花 | 0.381 | 0.480 | 0.159 | 0.217 | 0.447 | 0.584 | 0.433 | 0.323 | 50.637 |
| 亚欧唐松草 | 0.390 | 0.477 | 0.153 | 0.217 | 0.445 | 0.582 | 0.429 | 0.324 | 50.741 |
| 丝叶唐松草 | 0.391 | 0.477 | 0.154 | 0.217 | 0.444 | 0.584 | 0.423 | 0.319 | 50.778 |
| 粘唐松草 | 0.385 | 0.478 | 0.154 | 0.220 | 0.445 | 0.584 | 0.431 | 0.319 | 50.845 |
| 唐松草 | 0.385 | 0.478 | 0.154 | 0.220 | 0.445 | 0.582 | 0.429 | 0.324 | 50.741 |
| 辨蕊唐松草 | 0.391 | 0.477 | 0.154 | 0.217 | 0.447 | 0.584 | 0.431 | 0.319 | 50.845 |
| 细唐松草 | 0.391 | 0.477 | 0.154 | 0.217 | 0.444 | 0.584 | 0.487 | 0.319 | 50.531 |
| 多叶唐松草 | 0.387 | 0.477 | 0.157 | 0.217 | 0.444 | 0.580 | 0.431 | 0.321 | 50.710 |
| 星毛唐松草 | 0.390 | 0.477 | 0.153 | 0.220 | 0.445 | 0.584 | 0.431 | 0.320 | 50.845 |
| 朝鲜唐松草 | 0.381 | 0.478 | 0.159 | 0.220 | 0.447 | 0.584 | 0.431 | 0.326 | 50.667 |
| 平均值 | 0.387 | 0.477 | 0.155 | 0.218 | 0.445 | 0.583 | 0.435 | 0.321 | 50.757 |

2.3 密码子 RSCU 分析

11 个唐松草属植物的 *rbcL* 基因共包含 5 236 个密码子,其中蛋氨酸和色氨酸仅由 1 个密码子编码,分别为 ATG 和 TGG。其余的氨基酸由 2~6 个密码子编码,并表现出明显的密码子使用偏好性(图 1)。11 个唐松草属植物的 *rbcL* 基因中,最丰富的是亮氨酸(Leu),共包含 461(8.8%)个。同时,

在编码亮氨酸的 6 个密码子中,CUA 的 RSCU 值最大,为 1.72(表 3),表明 CUA 密码子具有很高的偏好性,是最常用的密码子。

从图 2 可以看出,11 个唐松草属 *rbcL* 基因中,共包含 27 个使用度较高的密码子,即 RSCU 值大于 1,在这 27 个密码子中,有 23 个密码子均以 A 或者 U 结尾,而仅有 4 个密码子 ACC、CGC、AUC 和 UCC

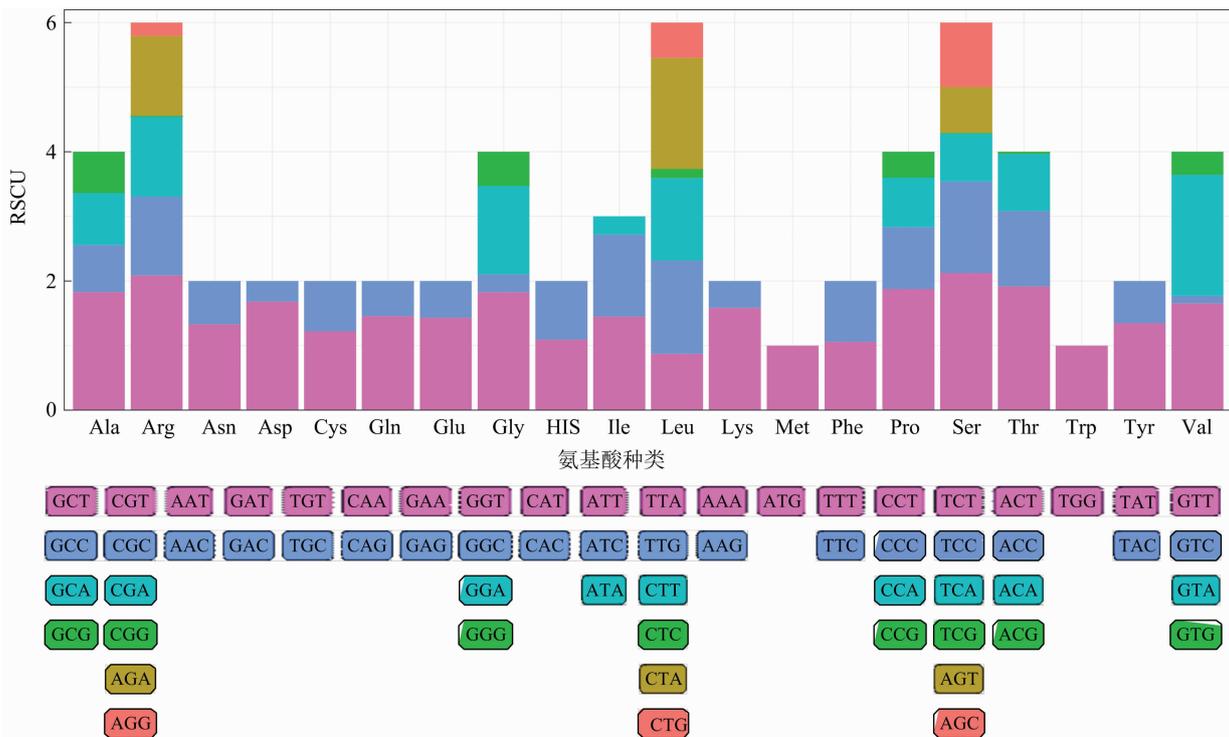


图 1 唐松草属 *rbcL* 基因的 RSCU 分析

表 3 唐松草属 *rbcL* 基因 RSCU 分析

| 氨基酸种类 | 密码子 | 密码子数 | RSCU | 氨基酸种类 | 密码子 | 密码子数 | RSCU |
|-------|-----|------|------|-------|-----|------|------|
| Phe | UUU | 123 | 1.06 | His | CAU | 90 | 1.09 |
| | UUC | 110 | 0.94 | | CAC | 75 | 0.91 |
| Leu | UUA | 67 | 0.87 | Gln | CAA | 88 | 1.45 |
| | UUG | 111 | 1.44 | | CAG | 33 | 0.55 |
| | CUU | 98 | 1.28 | Asn | AAU | 113 | 1.33 |
| | CUC | 11 | 0.14 | | AAC | 57 | 0.67 |
| | CUA | 132 | 1.72 | Lys | AAA | 209 | 1.58 |
| CUG | 42 | 0.55 | AAG | | 55 | 0.42 | |
| Ile | AUU | 113 | 1.45 | Asp | GAU | 231 | 1.68 |
| | AUC | 99 | 1.27 | | GAC | 44 | 0.32 |
| | AUA | 22 | 0.28 | Glu | GAA | 276 | 1.43 |
| Met | AUG | 101 | 1.00 | | GAG | 110 | 0.57 |
| | Val | GUU | 152 | 1.65 | Arg | CGU | 111 |
| GUC | | 11 | 0.12 | CGC | | 65 | 1.22 |
| GUA | | 172 | 1.87 | CGA | | 66 | 1.24 |
| GUG | | 33 | 0.36 | CGG | 0 | 0 | |
| Ser | UCU | 66 | 2.13 | Ser | AGU | 22 | 0.71 |
| | UCC | 44 | 1.42 | | AGC | 31 | 1.00 |
| | UCA | 23 | 0.74 | Arg | AGA | 66 | 1.24 |
| | UCG | 0 | 0 | | AGG | 11 | 0.21 |
| Pro | CCU | 103 | 1.87 | Gly | GGU | 231 | 1.83 |
| | CCC | 53 | 0.96 | | GGC | 35 | 0.28 |
| | CCA | 42 | 0.76 | | GGA | 173 | 1.37 |
| | CCG | 22 | 0.40 | | GGG | 67 | 0.53 |
| Thr | ACU | 161 | 1.92 | His | CAU | 90 | 1.09 |
| | ACC | 98 | 1.17 | | CAC | 75 | 0.91 |
| | ACA | 74 | 0.88 | Gln | CAA | 88 | 1.45 |
| | ACG | 3 | 0.04 | | CAG | 33 | 0.55 |
| Ala | GCU | 223 | 1.83 | Asn | AAU | 113 | 1.33 |
| | GCC | 89 | 0.73 | | AAC | 57 | 0.67 |
| | GCA | 98 | 0.80 | Lys | AAA | 209 | 1.58 |

以 C 结尾,表明 *rbcL* 11 个基因的密码子均偏好以 A 或者 U 结尾。继续分析发现,密码子使用度最高的 3 个密码子(RSCU 值大于 2),CGU、UCU、ACU 均以 U 结尾。表明在 A 与 U 中,*rbcL* 基因更偏好使用 U 结尾的密码子。

2.4 密码子参数相关性分析

对密码子 3 个位置的 GC 含量及 ENC 值做相关性分析(表 4),GC₁ 与 GC₂ 的相关系数为 0.350,显示出显著的相关性;GC_{all} 与 GC₁ 的相关系数为 0.710,极显著相关,与 GC₂ 的相关系数为 0.615,也极显著相关,而与 GC₃ 的相关系数为 0.405,显著相

关。GC₃ 与 GC₁ 的相关系数为 0.060,与 GC₂ 的相关系数为 -0.098,均不具有相关性,表明 *rbcL* 基因中,密码子第 1 位与第 2 位的碱基组成具有高度的相似度,但是第 3 位的碱基组成与第 1 位和第 2 位具有明显的不同。ENC 值与 GC₁ 和 GC₂ 的相关系数分别为 0.255 和 -0.049,相关性不显著,而与 GC₃ 为 0.335,显著相关,说明在 *rbcL* 基因中,密码子第 3 位的碱基组成与密码子的偏好性显著相关,第 3 位的碱基组成对密码子的偏好性有十分重要的影响。GC₃ 与密码子的出现次数(N)也具有显著相关性,表明密码子第 3 位的碱基组成对密码子的出现次数也有一定的影响。ENC 值与 N 不相关,说明基因序列的大小对密码子的偏好性影响较弱。

2.5 密码子中性绘图分析

GC₃ 与 GC₁₂ 的相关性很弱(图 3),相关系数非常低,表明 3 个位置的密码子含量存在明显的不同。在生成的所有基因的中性图中,回归线的斜率接近于 0,而且大多数绘图点不在对角线上或沿对角线分布。这些数据给出的证据表明,密码子的偏好是由自然选择主导的^[12]。

2.6 ENC - plot 结果

ENC - plot 结果表明,所有的基因位点均偏离了预期的曲线,表明所有 *rbcL* 基因的 ENC 值低于预期值,位于曲线下方(图 4)。因此,这些结果证实,11 个唐松草属植物的 *rbcL* 基因密码子偏好性大部分来自于选择的影响^[13]。且 11 个基因位点距离非常接近,表明自然选择对 11 个唐松草属物种的密码子偏好的影响相差不大^[13]。

2.7 PR2 - plot 分析

PR2 - plot 分析结果见图 5,可以看出 11 个唐松草属植物的 *rbcL* 基因全部位于左下方,表明这 11 个基因密码子第 3 位的碱基使用情况为 T 大于 A 且 C 大于 G,也就是说相比嘌呤 C/T,嘧啶 A/G 的使用频率要更低。PR2 - plot 分析结果表明,选择对唐松草属 *rbcL* 基因密码子的使用模式影响较大。

2.8 基于 RSCU 和 CDS 的聚类分析

分别使用 CDS 序列和所有密码子的 RSCU 值对 11 个唐松草属进行聚类分析,结果(图 6)显示,NJ 树与聚类图分别将 11 个唐松草属植物分为 4~5 个分枝,表明不同唐松草属植物 *rbcL* 的氨基酸编码模式具有一定的差异,虽然 NJ 树与聚类图的内部分枝存在明显的不同,但在分枝水平上,也有一致的聚类结果,比如都将星毛唐松草、粘唐松草聚为 1 个



图2 11种唐松草属植物*rbcL*基因RSCU值

表4 *rbcL*基因GC₁、GC₂、GC₃、GC_{all}、ENC及GC₁₂的相关性分析

| 参数 | 相关系数 | | | | | |
|-------------------|-----------------|-----------------|-----------------|-------------------|-------|------------------|
| | GC ₁ | GC ₂ | GC ₃ | GC _{all} | ENC | GC ₁₂ |
| GC ₂ | 0.350 * | | | | | |
| GC ₃ | 0.060 | -0.098 | | | | |
| GC _{all} | 0.710 ** | 0.615 ** | 0.405 * | | | |
| ENC | 0.255 | -0.049 | 0.335 * | 0.285 | | |
| GC ₁₂ | 0.750 ** | 0.801 ** | -0.009 | 0.905 ** | 0.145 | |
| N | -0.066 | -0.321 | 0.388 * | -0.080 | 0.250 | -0.215 |

注：*表示相关性显著(P<0.05)；**表示相关性极显著(P<0.01)。

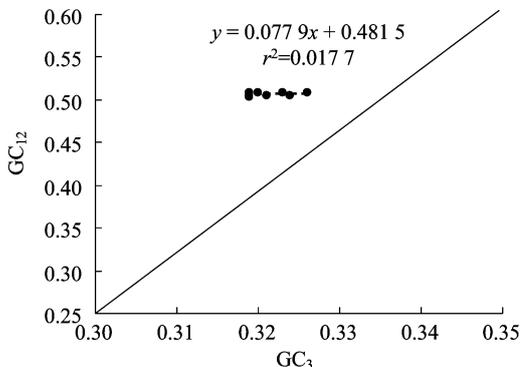


图3 *rbcL*基因中性绘图分析

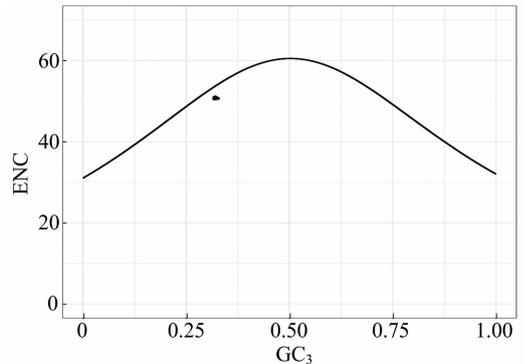
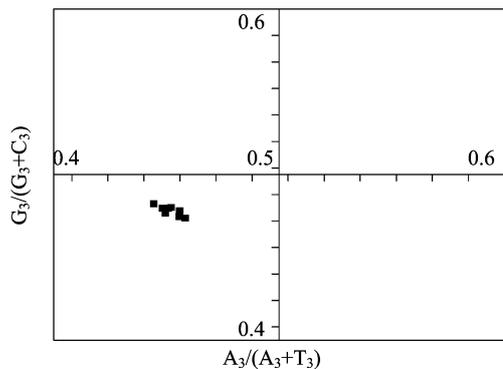
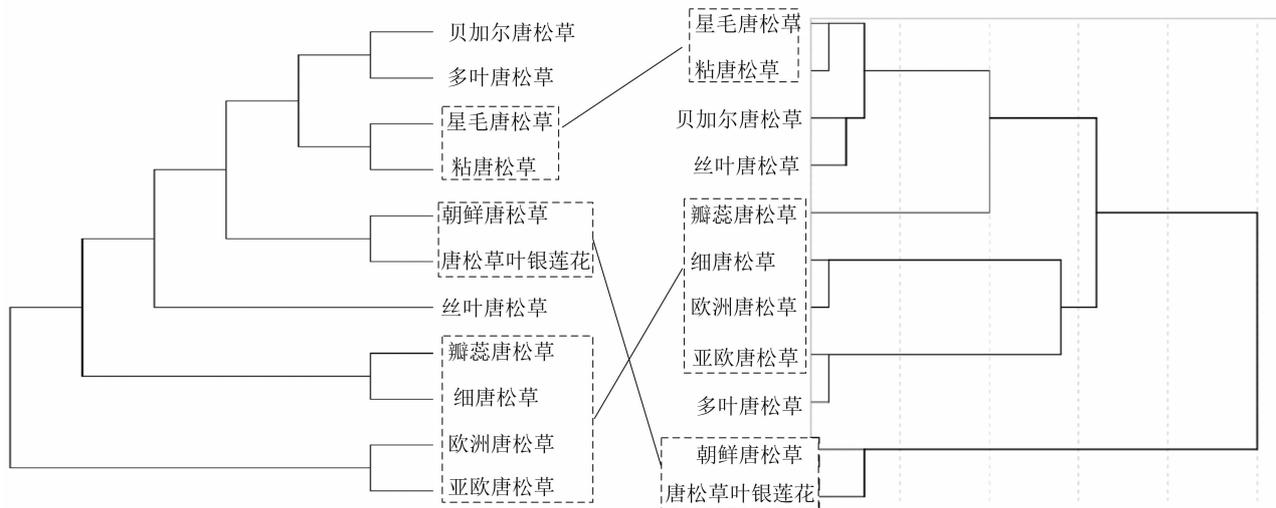


图4 ENC与GC₃关联分析

图5 11个 *rbcL* 基因 PR2-plot 绘图分析图6 11条 *rbcL* 基因构建 NJ 树(左)及用所有密码子的 RSCU 值的聚类结果(右)

(AGA), 而酵母有 30 个密码子的比值大于 1, 最大值为 2.61 (GUU), 表明相比大肠埃希菌来说, 酵母与唐松草属植物的密码子使用偏好性更为接近, 因此可以选用酵母作为其表达受体。烟草有 29 个密码子比值大于 2, 最大值为 2.02 (AAG), 拟南芥有 26 个密码子比值大于 2, 最大值为 2.001 (GCA), 番茄有 29 个密码子的比率大于 2, 最大值为 2.09 (GUG) (图 7), 这些结果充分说明, 与番茄和烟草相比, 拟南芥更适用于唐松草属 *rbcL* 基因的遗传转化受体。

3 讨论与结论

rbcL 是植物重要的编码基因, 主要位于叶绿体基因组, 叶绿体基因组结构很稳定, 因而植物叶绿体基因组常被用于群体进化、种质资源鉴定及群体遗传学研究^[17]。密码子在生物体内信息传递方面发挥着不可或缺的作用, 功能基因的转录、翻译及表达同样受到密码子的影响, 因而密码子的偏好性也会对功能基因的表达产生重要的作用^[18]。密码

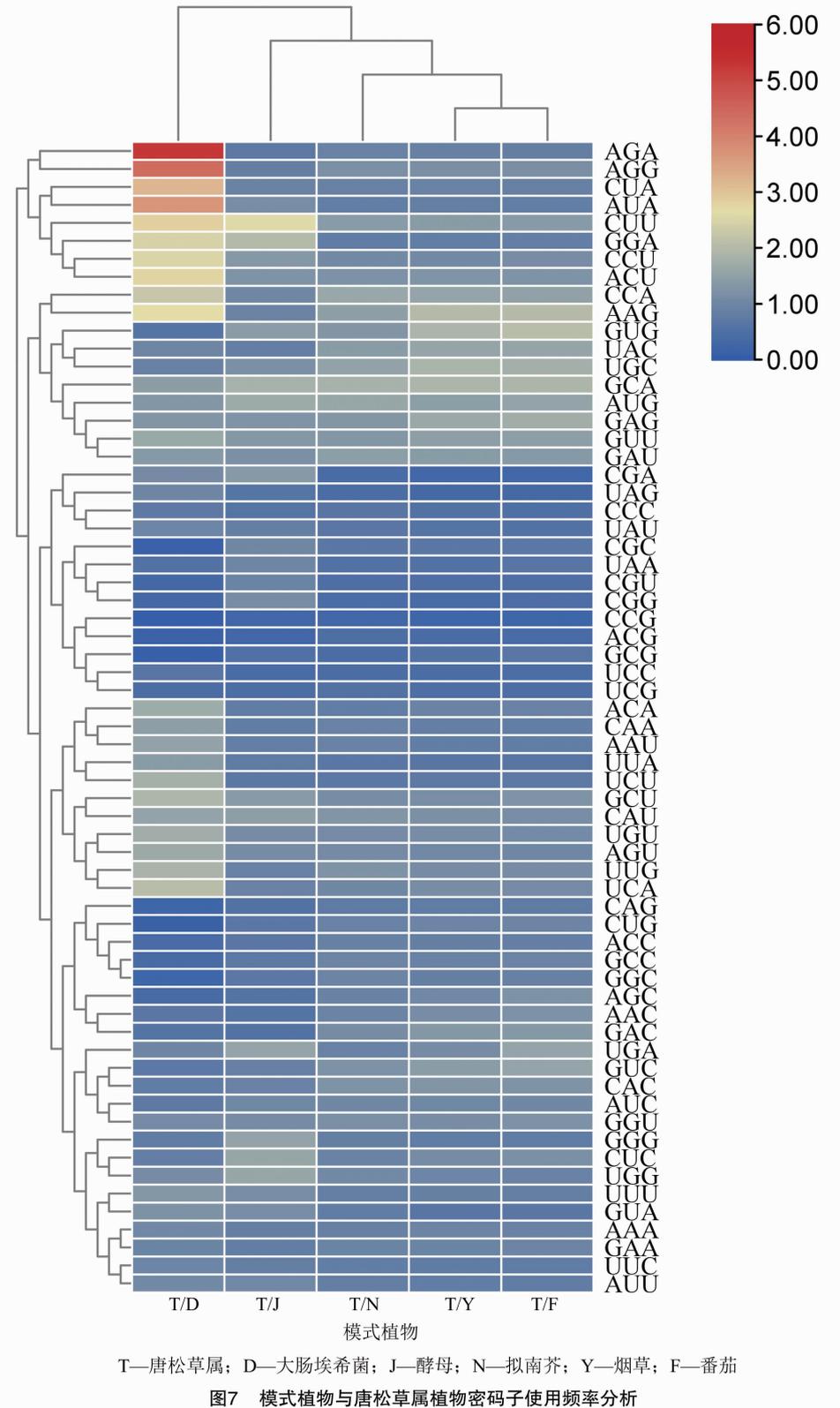
子, 将朝鲜唐松草和唐松草叶银莲花聚在一起, 且都将瓣蕊唐松草、细唐松草、欧洲唐松草及亚欧唐松草聚在 1 个分枝, 表明用 RSCU 值做聚类分析, 在一定程度上也可以反映唐松草属植物的亲缘关系, 即同一属内不同植物的密码子使用偏好性与其亲缘关系也存在着某种程度的关联。

2.9 与模式物种相对密码子使用频率分析

唐松草属植物与模式植物基因组序列的密码子使用频率结果表明, 与唐松草属 *rbcL* 相比, 大肠埃希菌有 36 个密码子的比值大于 2, 最大值为 5.27

子的使用偏好与基因表达密切相关表达, 并影响到基因组中的蛋白质和 mRNA 水平, 密码子独有的使用方法及使用偏好性会在植物长期的进化进程中逐渐累积, 因而在不同的植物中, 甚至同一植物不同基因间, 密码子的偏好性也会各有不同, 对其进行深入研究可以更好地理解功能基因的表达模式及进化方式, 从而进一步了解不同物种间的进化关系^[19]。

11 个唐松草属 *rbcL* 基因中, 最丰富的氨基酸是亮氨酸, 共包含 461 (8.8%) 个, 这与其他被子植物叶绿体基因组所报道^[20]的一致。更有意思的是, 大多数以 A/U 结尾的密码子的 RSCU 值都大于 1, 而以 C/G 结尾的则小于 1, 这一模式与其他植物的叶绿体基因组的使用模式一致^[21]。植物叶绿体基因组 *rbcL* 基因中, 密码子第 3 位通常比较保守, 选择压力对其影响不大, 所以 GC₃ 是密码子偏好性评估的一项重要指标^[21]。本研究发现不同基因的 GC₁、GC₂、GC₃ 含量存在明显的差异, GC₁、GC₂、GC₃ 的平均含量分别为 0.583、0.435、0.321。其中, GC₁ 含量最大, 而 GC₃ 含量最小, 且 GC₁ 与 GC₂ 显著相关, 而



GC₃ 与 GC₁、GC₂、GC₁₂均没有相关性,表明星毛唐松草的密码子 A/U 含量比 C/G 含量要多,中性绘图分析结果表明,密码子的偏好性受到了选择的影响,这与多种植物的研究结果^[20-21]一致。

11 个唐松草属 *rbcL* 基因中,ENC 值均大于 45,表明这些基因的密码子具有较弱的偏好性。ENC - plot 分析结果可以看出,该基因主要受到了选择作用的影响,巨桉的叶绿体基因组密码子偏好研究

中,同样也发现了这一规律^[22]。PR2 - plot 绘图表明,密码子第 3 位的碱基出现的规律为 T、G 大于 A、C,这一结果体现出选择对 11 个唐松草属 *rbcL* 基因中密码子偏好性影响较大,突变及其他因素的作用较小,这与原晓龙等在蒜头果 (*Malania oleifera*) 研究中发现的规律^[12]一致。总结以上结果后发现,11 个唐松草属 *rbcL* 基因中,密码子使用偏好性与多种因素有关,但是选择的作用最为明显。

虽然基于基因序列的 NJ 树与基于 RSCU 值的聚类图的内部分枝存在明显的不同,但在分枝水平上,也有一致的聚类结果,表明用 RSCU 值做聚类分析,在一定程度上也可以反映唐松草属植物的亲缘关系,二者之间存在一定的对应关系。RSCU 聚类树结果的偏差可能是由于仅选取 RSCU 值这一单一的数据导致的,结合密码子偏好性的其他数据,可能会降低误差,在苔藓植物中,也得出了同样的结论^[23]。

模式物种通常被选作目标基因的异源表达受体,来完成转基因试验,而如果二者的基因密码子使用的偏好性具有相似性,这一过程会更加顺利,目标基因可以高效表达^[24]。研究发现,酵母与唐松草属植物的密码子使用偏好性更为接近,可以作为其表达受体,而拟南芥更适用于唐松草属 *rbcL* 基因的遗传转化受体^[25]。

参考文献:

[1] Santos Siqueira A, Lima A R J, Dall'Agnol L T, et al. Comparative modeling and molecular dynamics suggest high carboxylase activity of the *Cyanobium* sp. CACIAM14 *rbcL* protein [J]. *Journal of Molecular Modeling*, 2016, 22(3): 68 - 78.

[2] 李冬林, 金雅琴, 崔梦凡, 等. 夏季遮光对连香树幼苗形态、光合作用及叶肉细胞超微结构的影响 [J]. *浙江农林大学学报*, 2020, 37(3): 496 - 505.

[3] 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.

[4] 钱留华. 密码子 [J]. *生物学通报*, 1997(4): 15 - 17.

[5] 吴学俊, 梁湘兰, 易子群, 等. 湖北山楂叶绿体基因组密码子偏好性分析 [J]. *分子植物育种*, 2021(6): 1 - 16.

[6] 任桂萍, 董璿莹, 党云琨. 密码子中的密码: 密码子偏好性与基因表达的精细调控 [J]. *中国科学(生命科学)*, 2019, 49(7): 839 - 847.

[7] 原晓龙, 康洪梅, 王毅. 云南蓝果树叶绿体基因组密码子偏好性分析 [J]. *西北林学院学报*, 2020, 35(4): 26 - 31, 124.

[8] 冯瑞云, 梅超, 王慧杰, 等. 籽粒苋叶绿体基因组密码子偏好性

分析 [J]. *中国草地学报*, 2019, 41(4): 8 - 15.

[9] 李路丽, 张萍, 曾强, 等. 芝麻叶绿体基因组的密码子使用偏好性及其影响因素 [J]. *分子植物育种*, 2021, 20(3): 1 - 11.

[10] Luo D, Lv N, Zhu L J, et al. Isoquinoline alkaloids from whole plants of *Thalictrum cirrhosum* and their antitrotavirus activity [J]. *Chemistry of Natural Compounds*, 2020, 56(3): 504 - 508.

[11] Chen C J, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. *Molecular Plant*, 2020, 13(8): 1194 - 1202.

[12] 原晓龙, 刘音, 康洪梅, 等. 蒜头果叶绿体基因组密码子偏好性分析 [J]. *西南林业大学学报(自然科学)*, 2021, 41(3): 15 - 22.

[13] Rozas J, Ferrer - Mata A, Sánchez - Delbarrio J C, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets [J]. *Molecular Biology and Evolution*, 2017, 34(12): 3299 - 3302.

[14] Nakamura T, Yamada K D, Tomii K, et al. Parallelization of MAFFT for large - scale multiple sequence alignments [J]. *Bioinformatics*, 2018, 34(14): 2490 - 2492.

[15] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7. 0 for bigger datasets [J]. *Molecular Biology and Evolution*, 2016, 33(7): 1870 - 1874.

[16] 彭丽云, 王云, 孙雪丽, 等. 苋菜 *AmMYB2* 基因密码子偏好性与进化分析 [J]. *应用与环境生物学报*, 2019, 25(3): 679 - 686.

[17] Palmer J D. Comparative organization of chloroplast genomes [J]. *Annual Review of Genetics*, 1985, 19: 325 - 354.

[18] 李真, 李永光, 薄蕾, 等. 大豆和拟南芥 *NFYB* 基因家族同义密码子使用偏好性的比较 [J]. *作物杂志*, 2014(3): 14 - 21.

[19] Lyu X L, Liu Y. Nonoptimal Codon usage is critical for protein structure and function of the master general amino acid control regulator CPC - 1 [J]. *mBio*, 2020, 11(5): e02605 - e02620.

[20] Somaratne Y, Guan D L, Wang W Q, et al. The complete chloroplast genomes of two *Lespedeza* species: insights into codon usage bias, RNA editing sites, and phylogenetic relationships in Desmodieae (Fabaceae: Papilionoideae) [J]. *Plants*, 2019, 9(1): 51 - 61.

[21] Liu X Y, Li Y, Ji K K, et al. Genome - wide codon usage pattern analysis reveals the correlation between codon usage bias and gene expression in *Cuscuta australis* [J]. *Genomics*, 2020, 112(4): 2695 - 2702.

[22] 王鹏良, 吴双成, 杨利平, 等. 巨桉叶绿体基因组密码子偏好性分析 [J]. *广西植物*, 2019, 39(12): 1583 - 1592.

[23] 张家榕, 雷万钧. 18 种苔藓植物 *rbcL* 基因的密码子偏性及聚类分析 [J]. *山西农业大学学报(自然科学版)*, 2020, 40(6): 1 - 12.

[24] Daniell H, Lin C S, Yu M, et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering [J]. *Genome Biology*, 2016, 17(1): 134 - 142.

[25] 郑钢, 顾翠花, 林琳, 等. 20 种千屈菜科植物 *rbcL* 基因密码子使用偏好性分析 [J]. *浙江农林大学学报*, 2021, 38(3): 476 - 484.