浙 江 农 林 大 学 学 报，2021，**38**(3)：476–484
Journal of Zhejiang A&F University
doi: 10.11833/j.issn.2095-0756.20200390

20 种千屈菜科植物 *rbcL* 基因密码子使用偏好性分析

郑　钢[1,2,3]，顾翠花[1,2,3]，林　琳[1,2,3]，王　杰[1,2,3]

（1. 浙江农林大学 风景园林与建筑学院，浙江 杭州 311300；2. 浙江农林大学 浙江省园林植物种质创新与利用重点实验室，浙江 杭州 311300；3. 浙江农林大学 南方园林植物种质创新与利用国家林业和草原局重点实验室，浙江 杭州 311300）

摘要：【目的】分析千屈菜科 Lythraceae 植物 *rbcL* 基因密码子使用特性，明确密码子偏好性的影响因素，筛选 *rbcL* 基因异源表达和遗传转化的合适受体。**【方法】**从美国国家生物技术信息中心 (NCBI) 获取 20 种千屈菜科植物的 *rbcL* 基因全长编码区序列 (CDS) 数据，运用 CodonW、EMBOSS 和 DAMBE 软件获取基因碱基组成和密码子使用偏好性的相关参数，分析该类植物叶绿体 *rbcL* 基因使用密码子的偏倚性及其影响因素。**【结果】**千屈菜科植物 *rbcL* 基因 GC 含量 (GC) 为 0.425~0.437，密码子第 3 位碱基 GC 含量 (GC_{3s}) 为 0.275~0.300。GC_{3s}、GC 与有效密码子数 (ENC) 显著相关 ($P<0.01$)；ENC-GC_{3s} 散点图分析、中性绘图分析、奇偶偏差分析均表明：相较于突变压力，自然选择压力对千屈菜科植物 *rbcL* 基因密码子使用偏好性的影响更大。基于同义密码子相对使用度的系统聚类与 CDS 邻接树结果部分一致。与千屈菜科 *rbcL* 基因密码子平均使用频率相比，大肠埃希菌 *Escherichia coli*、酵母 *Saccharomyces cerevisiae*、拟南芥 *Arabidopsis thaliana*、烟草 *Nicotiana tabacum* 和番茄 *Solanum lycopersicum* 分别存在 28、26、20、19 和 17 个使用频率相差较大的密码子。**【结论】**千屈菜科植物 *rbcL* 基因碱基组成上更倾向于选择 A/T 碱基，且偏好使用末端 A/T 碱基的密码子；*rbcL* 基因密码子使用偏好性受多种因素共同作用，但自然选择压力是最主要因素；密码子偏好性的系统聚类可为系统发育研究提供补充；酵母更适合作为千屈菜科植物 *rbcL* 基因异源表达受体，番茄更适合作为 *rbcL* 基因遗传转化和功能研究的受体材料。图 6 表 3 参 32

关键词：碱基组成；选择压力；突变压力；聚类分析；受体

中图分类号：S718.3　　**文献标志码：**A　　**文章编号：**2095-0756(2021)03-0476-09

Codon usage bias analysis of *rbcL* genes of 20 Lythraceae species

ZHENG Gang[1,2,3]，GU Cuihua[1,2,3]，LIN Lin[1,2,3]，WANG Jie[1,2,3]

（1. College of Landscape Architecture, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China; 2. Zhejiang Provincial Key Laboratory of Germplasm Innovation and Utilization for Garden Plants, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China; 3. Key Laboratory of National Forestry and Grassland Administration on Germplasm Innovation and Utilization for Southern Garden Plants, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China）

Abstract: [**Objective**] With an analysis of the codon usage characteristics of the *rbcL* genes in Lythraceae species, this study is aimed to clarify the influencing factors of codon bias, and screen the optimal receptor for heterologous expression and genetic transformation. [**Method**] After *rbcL* gene CDS of 20 Lythraceae species were obtained from NCBI, CodonW, EMBOSS, and DAMBE software were utilized to compute relevant parameters of gene base composition and codon usage bias before an analysis is conducted of the usage bias of

收稿日期：2020-06-19；修回日期：2021-03-08
基金项目：浙江省自然科学基金资助项目 (LY21C160001)
作者简介：郑钢 (ORCID: 0000-0001-5666-4832)，实验师，从事园林植物遗传育种研究。E-mail: 305788868@qq.com。通信作者：王杰 (ORCID: 0000-0003-0038-1045)，从事园林植物遗传育种与种质资源创新研究。E-mail: wangjie@stu.zafu.edu.cn

such genes and its incluencing factors using SPSS and Origin software. [**Result**] The GC content of the *rbcL* gene from Lythraceae species ranged from 0.425 to 0.437, with GC_{3s} being 0.275 to 0.300 and there was a significant correlation between GC_{3s}, GC, and ENC($P<0.01$). As was shown in the analysis of ENC-GC_{3s} plot, the neutral plot and PR2, natural selection pressure affected the codon usage bias of the *rbcL* gene from Lythraceae species more heavily than mutation pressure. The result of clustering analysis based on RSCU is partially consistent with that of the neighbor-joining tree based on CDS. Compared with the average codon usage frequency of the *rbcL* gene from the 20 Lythraceae species, *Escherichia coli*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Nicotiana tabacum*, and *Solanum lycopersicum* possessed 28, 26, 20, 19 and 17 codons, respectively, with significant differences in usage frequency. [**Conclusion**] In terms of the base composition of the *rbcL* gene from 20 Lythraceae species, there was a tendency towards A/T bases and codons with A/T base at ther termonal were generally preferred. Also, of all the factors having an influence on codon bias, natural selection pressure was the most important one. Systematic clustering is a good complement for phylogenetic analysis. *S. cerevisiae* is more suitable as a heterologous expression receptor, while *S. lycopersicum* is more suitable to act as a receptor material for genetic transformation and function research of *rbcL* gene. [Ch, 6 fig. 3 tab. 32 ref.]

Key words: base composition; selection pressure; mutation pressure; clustering analysis; receptor

密码子承担着生物体内遗传信息传递的重要功能，是 DNA 转录与翻译、蛋白质合成与表达过程中的关键单元。在生物体共用的一套密码子中，终止密码子不编码氨基酸，甲硫氨酸 (Met) 和色氨酸 (Trp) 分别由 1 种密码子编码。其余 59 个密码子具有简并性，即 1 种氨基酸可由 2~6 个密码子对应编码，编码相同氨基酸的密码子即为同义密码子[1]。基因并非完全随机地使用同义密码子，而是存在一定的偏好性。特定的密码子偏好性是生物体长期适应性进化的结果，能够反映生物对环境的分子适应机制[2]。分析密码子偏好性及其影响因素，对生物遗传育种、进化基因组学以及系统发育学研究具有深远的意义。1,5-二磷酸核酮糖羧化/加氧酶 (Ribulose-1,5-bisphosphate carboxylase/oxygenase, Rubisco 酶) 是植物叶绿体基质中参与光合作用的关键酶，约占可溶性蛋白质总量的 50%[3]。Rubisco 酶具有催化 1,5-二磷酸核酮糖 (Ribulose-1,5-disphosphate, RuBP) 与二氧化碳 (CO_2) 羧化反应和光呼吸中 RuBP 与氧气 (O_2) 加氧反应的双重活性，对净光合率有决定性影响[4]。Rubisco 酶由 8 个大亚基 (催化亚基) 和 8 个小亚基 (调节亚基) 组成，前者是固定 CO_2 的活性位点和催化位点，由叶绿体基因组大单拷贝区的 *rbcL* 基因编码[5−6]。环境的变化会导致 *rbcL* 基因产生适应性进化，从而影响植物光合效率[7]。因此，研究 *rbcL* 基因的密码子使用模式有利于理解高等植物对环境的适应机制。千屈菜科 Lythraceae 包括许多重要的园林植物，具有重要的观赏价值和经济价值[8]。目前，*rbcL* 基因在千屈菜科中的研究应用仅局限于系统发育[9−10]，对于该科密码子使用偏好性的相关研究尚未见报道。本研究选取了千屈菜科具有代表性的 10 属 20 种植物，分析 *rbcL* 基因的碱基组成、密码子使用偏好性及其影响因素，并与模式物种进行比较，为该科物种 *rbcL* 基因异源高效表达提供理论基础。

1　材料与方法

1.1　基因序列和密码子使用频率数据获取

20 条 *rbcL* 基因全长编码区序列 (CDS) 数据来源于美国国家生物技术信息中心 (NCBI) 的 GenBank 数据库 (https://www.ncbi.nlm.nih.gov/)，详见表 1。

1.2　CDS 碱基组成和密码子使用偏好性参数统计

通过 CodonW 1.4.4 软件和在线工具 EMBOSS explorer(http://emboss.toulouse.inra.fr./) 中的 CUSP 和 CHIPS 程序，统计 *rbcL* 基因密码子末端各类型碱基含量 (A_{3s}、T_{3s}、C_{3s} 和 T_{3s})、GC 总含量 (GC)、密码子各位点 GC 含量 (GC_{1s}、GC_{2s} 和 GC_{3s})、有效密码子数 (ENC) 和密码子适应指数 (CAI)。利用 SPSS 22.0 软件，选用皮尔森相关系数评估碱基组成和密码子偏好性相关显著水平[11]。

表 1 20 种千屈菜科植物 *rbcL* 基因信息

Table 1 Information of *rbcL* genes from 20 Lythraceae species

物种	GenBank登录号	CDS位置	物种	GenBank登录号	CDS位置
萼距花 *Cuphea hyssopifolia*	MN833211	58955~60382	南洋紫薇 *Lagerstroemia siamica*	MK881628	55129~56556
八宝树 *Duabanga grandiflora*	MK881638	56823~58250	绒毛紫薇 *Lagerstroemia tomentosa*	MK881632	54873~56300
黄薇 *Heimia myrtifolia*	MG921615	58612~60039	西双紫薇 *Lagerstroemia venusta*	MK881630	55159~56586
副萼紫薇 *Lagerstroemia calyculata*	MK881636	54873~56300	散沫花 *Lawsonia inermis*	MK881631	58836~60263
川黔紫薇 *Lagerstroemia excelsa*	MK881635	54910~56337	千屈菜 *Lythrum salicaria*	MK881629	59099~60526
屋久岛紫薇 *Lagerstroemia fauriei*	NC_029808	54810~56237	石榴 *Punica granatum*	NC_035240	59017~60444
多花紫薇 *Lagerstroemia floribunda*	NC_031825	54776~56203	圆叶节节菜 *Rotala rotundifolia*	MK881626	58835~60262
桂林紫薇 *Lagerstroemia guilinensis*	NC_029885	54697~56124	细果野菱 *Trapa maximowiczii*	NC_037023	58322~59770
云南紫薇 *Lagerstroemia intermedia*	NC_034662	54948~56375	欧菱 *Trapa natans*	MK881634	58387~59814
福建紫薇 *Lagerstroemia limii*	MK881627	54830~56257	虾子花 *Woodfordia fruticosa*	MK881637	59444~60871

1.3 同义密码子相对使用度统计与分析

同义密码子相对使用度 (RSCU) 是同义密码子的实际使用频次与无使用偏好性时期望频次的比率，去除了碱基成分对密码子使用产生的影响。RSCU＞1，表示该密码子在同义密码子中使用相对较多；RSCU=1，表示该密码子在同义密码子中使用无偏好性；RSCU＜1 表示该密码子在同义密码子中使用相对较少[12]。通过 CodonW 1.4.4 软件计算千屈菜科植物的 RSCU，并利用 TBtools 0.6 软件绘图。

1.4 ENC 绘图分析

以 GC_{3s} 和 ENC 为横、纵坐标，通过 Origin 9.1 绘制 ENC-GC_{3s} 散点图。标准曲线为 ENC 期望值，即 $N_{ENC}=2+M_{GC3s}+29/[M_{GC3s}^2+(1-M_{GC3s})^2]$，其中 N_{ENC} 表示有效密码子数，M_{GC3s} 表示密码子第 3 位碱基平均 GC 含量，该公式的成立表示密码子的偏好性仅受突变压力约束[13]，此条件下，散点应位于标准曲线上部或紧贴标准曲线下部；当散点分布于曲线下方较远距离的区域时，表明除突变压力作用外，选择压力对偏好性产生主要影响。

1.5 中性绘图分析

以 GC_{3s} 为横坐标，密码子第 1、2 位点 GC 含量平均值 (GC_{12}) 为纵坐标，利用 Origin 9.1 绘制散点图并做线性回归分析，分析密码子不同位点碱基组成差异性[14]。当回归曲线斜率趋近 1 时，密码子各位点碱基成分差异不大，偏好性主要受到突变的影响；当斜率趋近 0 时，密码子第 3 位点和第 1、2 位点碱基变异模式差异较大，偏好性主要受到选择压力影响。

1.6 奇偶偏差 (PR2) 分析

奇偶偏差分析可评估密码子第 3 位点嘌呤和嘧啶组成偏差对密码子使用偏好性的影响[15]。以 $G_{3s}/(G_{3s}+C_{3s})$ 和 $A_{3s}/(A_{3s}+T_{3s})$ 为横、纵坐标，利用 Origin 9.1 绘制奇偶偏差图，交点 (0.50, 0.50) 表示无碱基突变和选择压力下，A=T 且 G=C。

1.7 基于 RSCU 和 CDS 的聚类分析

参照巫伟峰等[16] 方法，以 59 个密码子 (去除 AUG、UGG 和 3 个终止密码子 UAA、UAG、UGA) 的 RSCU 为变量，20 条 CDS 为个体，通过 SPSS 进行系统聚类，类间距离为组内联接法，基因间距离为平方欧式距离。分别利用 DAMBE 5.2.73 和 MEGA-X 软件对 CDS 进行碱基替换饱和度检测和总体平均距离 (d) 计算，同时满足替换饱和度指数 (I_{ss}) 小于饱和度标准指数 ($I_{ss.c}$)，即 $I_{ss}<I_{ss.c}$，表明碱基替换未饱和，且 P=0.000 和 $0<d<1$ 后，通过 MEGA-X 软件邻接法 (NJ) 构建系统发生树，重复 1 000 次。

1.8 密码子使用频率比较分析

密码子相对使用频率比值是评估不同生物密码子使用偏好性差异程度的重要参数。当比值为 0.5~2.0 时，认为物种密码子偏好性差异较小[17]。拟南芥 *Arabidopsis thaliana*、烟草 *Nicotiana tabacum*、番茄 *Solanum lycopersicum*、大肠埃希菌 *Escherichia coli* 和酵母 *Saccharomyces cerevisiae* 的基因组密码子使用频率来源于密码子使用数据库 (http://www.kazusa.or.jp/codon/)。千屈菜科物种整体密码子平均使用频率通过 EMBOSS explorer 中 CUSP 计算获得[18]。利用 Origin 9.1 进行绘图。

2　结果与分析

2.1　*rbcL* 基因碱基组成和密码子使用偏好性

从表 2 可见：GC 含量为 0.425~0.437，平均为 0.431。结合密码子各位点 GC 含量 (GC_{1s} 为 0.567~0.582，平均 0.573；GC_{2s} 为 0.429~0.437，平均 0.432；GC_{3s} 为 0.275~0.300，平均 0.288)，表明 *rbcL* 基因 CDS 在组成上更倾向于使用 A/T 碱基。第 3 位点各类型碱基含量从大到小依次为 T_{3s}、A_{3s}、C_{3s}、G_{3s}，表明 *rbcL* 基因更偏向于使用 A/T 碱基结尾的密码子。

表 2　20 种千屈菜科植物 *rbcL* 基因碱基组成和密码子使用特性

Table 2　Base composition and codon usage characteristics of *rbcL* genes from 20 Lythraceae species

物种	A_{3s}	T_{3s}	G_{3s}	C_{3s}	GC	GC_{1s}	GC_{2s}	GC_{3s}	CAI	ENC
萼距花	0.376	0.531	0.157	0.173	0.435	0.582	0.437	0.286	0.276	45.392
八宝树	0.380	0.526	0.152	0.180	0.431	0.571	0.433	0.288	0.278	45.942
黄薇	0.390	0.508	0.145	0.194	0.434	0.571	0.433	0.296	0.283	46.540
副萼紫薇	0.377	0.525	0.148	0.186	0.432	0.576	0.429	0.292	0.277	45.635
川黔紫薇	0.376	0.526	0.149	0.187	0.432	0.571	0.431	0.294	0.275	45.743
屋久岛紫薇	0.379	0.529	0.146	0.184	0.431	0.571	0.431	0.290	0.272	45.659
多花紫薇	0.378	0.526	0.148	0.184	0.432	0.576	0.429	0.292	0.276	45.625
桂林紫薇	0.376	0.526	0.149	0.187	0.432	0.571	0.431	0.294	0.275	45.743
云南紫薇	0.379	0.526	0.140	0.191	0.431	0.571	0.431	0.290	0.275	45.340
福建紫薇	0.379	0.531	0.142	0.184	0.430	0.571	0.431	0.288	0.274	45.564
南洋紫薇	0.379	0.526	0.140	0.191	0.431	0.571	0.431	0.290	0.275	45.340
绒毛紫薇	0.377	0.525	0.148	0.186	0.432	0.576	0.429	0.292	0.277	45.635
西双紫薇	0.379	0.526	0.140	0.191	0.431	0.571	0.431	0.290	0.275	45.340
散沫花	0.379	0.536	0.151	0.171	0.429	0.569	0.435	0.282	0.276	45.264
千屈菜	0.389	0.535	0.138	0.173	0.428	0.576	0.433	0.275	0.285	45.007
石榴	0.381	0.518	0.153	0.184	0.436	0.578	0.437	0.294	0.275	46.153
圆叶节节菜	0.379	0.536	0.151	0.171	0.429	0.569	0.435	0.282	0.276	45.264
细果野菱	0.387	0.532	0.154	0.165	0.425	0.567	0.431	0.277	0.274	44.181
欧菱	0.387	0.532	0.154	0.165	0.426	0.569	0.431	0.277	0.274	44.029
虾子花	0.376	0.516	0.163	0.184	0.437	0.576	0.435	0.300	0.270	46.458

ENC 和 CAI 是衡量密码子使用偏好性程度的主要指标。ENC 从 20(氨基酸只由 1 种同义密码子编码) 至 61(同义密码子的使用没有偏好性)，越接近 20 偏好性越强。一般认为，ENC＜35 表示密码子的使用偏好性较强[19]。20 种千屈菜科植物 ENC 为 44.029~46.540，平均 45.493，分布范围较小且均远大于 35，表明 *rbcL* 基因整体偏好性不强。CAI 取值 0~1，越接近 1 密码子偏好性越强[20]。20 种植物 CAI 为 0.270~0.285，平均 0.276，同样说明偏好性强度不大。一般情况下，基因的密码子使用偏好性越强，在生物体内的表达水平越高[21]，可推测 *rbcL* 基因在千屈菜科植物中表达水平较低。

2.2　*rbcL* 基因同义密码子相对使用度分析

图 1 显示：在 25 个高频密码子 (RSCU＞1) 中，23 个以 A/U 结尾，仅 2 个由 C(AUC 和 AGC) 结尾。其中 RSCU 最高的 5 个密码子 (RSCU＞2) 末尾均为 U 碱基，表明 *rbcL* 基因 CDS 对于末端 A/U(T) 密码子具有的使用偏好性。

2.3　密码子碱基组成和使用偏好相关分析

相关分析 (表 3) 表明：ENC 和 GC、GC_{3s} 在 0.01 水平上显著相关 (Pearson 相关系数分别为 0.855 和 0.856)，表明碱基组成，尤其是密码子第 3 位点碱基类型对千屈菜科 *rbcL* 基因的密码子偏好性有明显影响。GC_{3s} 和 GC_{12} 相关不显著，说明不同位点组成上关联不大，碱基变异模式存在差异，*rbcL* 基因较保守，突变偏性较小。

2.4 ENC 绘图分析

图 2 显示了 *rbcL* 基因 ENC 和 GC_{3s} 的关系。所有散点分布在标准曲线下方一定距离处，表明千屈菜科植物 *rbcL* 基因的密码子偏好性除了受到碱基突变压力外，更主要受自然选择压力的约束；散点集中分布在较小范围内说明自然选择压力强度相近。

2.5 中性绘图分析

中性分析结果 (图 3) 显示：所有散点均落在直线 $y=x(GC_{12})$ 上方。GC_{3s} 与 GC_{12} 的回归曲线 (斜率为 0.069 4，R^2=0.036 1) 近似平行于 X 轴，表明千屈菜科植物 *rbcL* 基因密码子第 1、2 位点与第 3 位点碱基类型相差较大。结合表 3，GC_{3s} 与 GC_{12} 相关性较低 (Pearson 相关系数为 0.190)，说明碱基突变对于密码子第 3 位点的作用比第 1、2 位点弱，密码子偏好性主要受自然选择压力的作用，受突变压力的影响则较小。

2.6 奇偶偏差 (PR2) 分析

图 4 显示：当密码子偏好性只受碱基突变影响时，密码子第 3 位点上嘌呤和嘧啶含量应相同，即 $A_{3s}=T_{3s}$ 或 $C_{3s}=G_{3s}$[22]。所有散点均明显偏离交点 (0.50, 0.50)，且都分布在左下象限 [$G_{3s}/(G_{3s}+C_{3s})<0.5$，$A_{3s}/(A_{3s}+T_{3s})<0.5$]，密码子第 3 位点上嘧啶含量高于嘌呤 [$(A_{3s}+G_{3s})<(T_{3s}+C_{3s})$]。4 种碱基在密码子第 3 位点上分布不均匀，说明相较于碱基突变压力，自然选择压力对 *rbcL* 密码子偏好性有更强的影响。

2.7 基于 RSCU 和 CDS 的聚类分析

20 条 CDS 碱基替换未饱和 (I_{ss}=0.025 3，$I_{ss.c}$=0.785 2，P=0.000)，总体平均遗传距离为 0.2。系统聚类树状图和邻接树均将 20 种千屈菜科植物聚成了 4~5 个支系 (图 5)，说明不同支系的植物密码子使用特性存在一定区别。虽然两者在部分支系的内部结构上存在较大矛盾，但在支系水平 (属) 上，两者对 10 个紫薇属 *Lagerstroemia* 植物、散沫花和圆叶节节菜以及 2 个菱属 *Trapa* 植物之间的聚类结果相对一致，说明基于密码子 RSCU 的系统聚类能在某种程度上反映千屈菜科植物属间水平的亲缘关系，即不同植物密码子的使用偏好性与亲缘关系存在局部对应。

2.8 千屈菜科植物与模式物种密码子使用频率比较分析

从图 6 可以看出：与千屈菜科植物 *rbcL* 基因密码子平均使用频率相比，大肠埃希菌有 28 个密码子相差较大，最大值 5.76(AGA)；酵母有 26 个密码子相差较大，最大值 4.33(CGU)，说明酵母更适合作为千屈菜科植物 *rbcL* 基因异源表达的受体。拟南

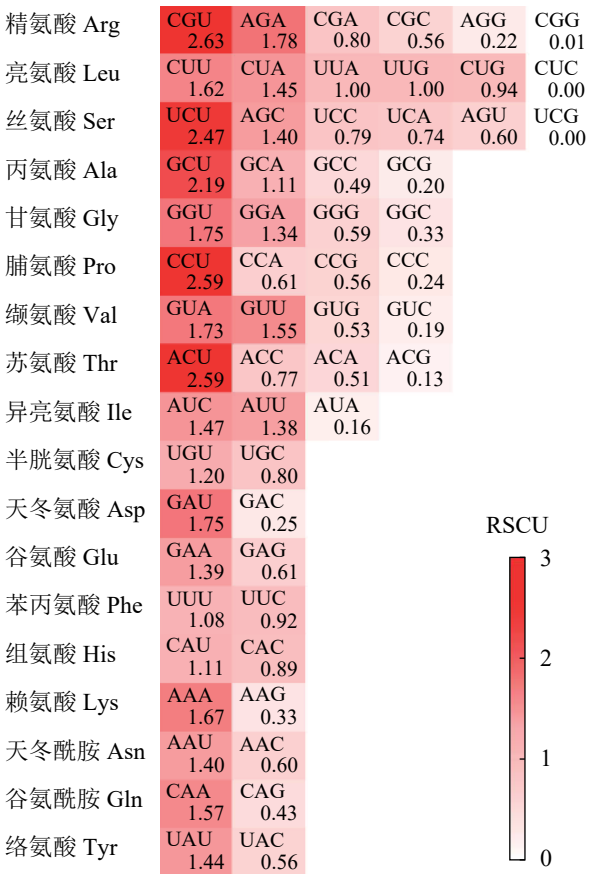


图 1 20 种千屈菜科植物 *rbcL* 基因同义密码子相对使用度

Figure 1 RSCU of *rbcL* genes from 20 Lythraceae species

表 3 碱基组成与密码子使用偏好相关性

Table 3 Correlation between base composition and codon usage bias

参数	CAI	ENC	GC	GC_{1s}	GC_{2s}	GC_{3s}
ENC	0.062					
GC	−0.136	0.855**				
GC_{1s}	0.138	0.403	0.712**			
GC_{2s}	0.029	0.229	0.348	0.314		
GC_{3s}	−0.264	0.856**	0.846**	0.324	−0.074	
GC_{12}	0.112	0.403	0.684**	0.869**	0.743**	0.190

说明：**表示在0.01水平上显著相关(双尾)

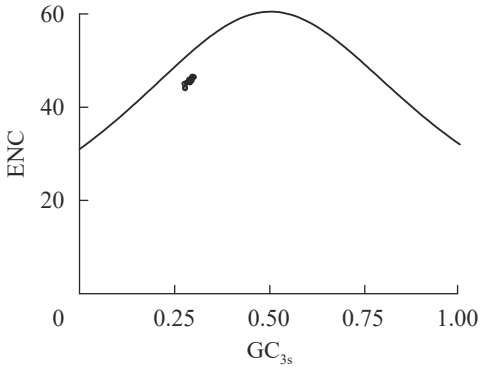


图 2 *rbcL* 基因 ENC-GC_{3s} 绘图分析

Figure 2 ENC-GC_{3s} plot analysis of *rbcL* genes

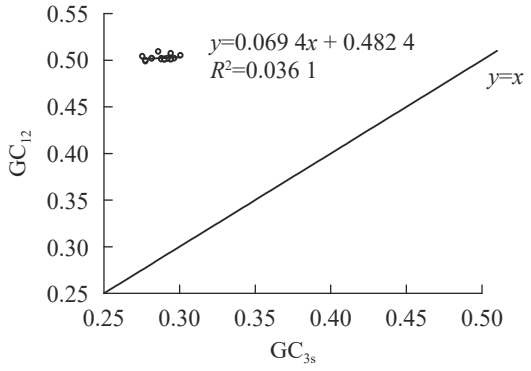


图 3　GC_{3s} 与 GC_{12} 的中性绘图

Figure 3　Neutral plot of GC_{3s} and GC_{12}

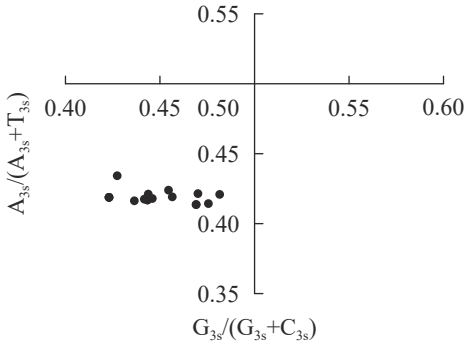


图 4　*rbcL* 基因密码子第 3 位点碱基奇偶偏好

Figure 4　PR2 plot of the 3rd sites in codons of *rbcL* genes

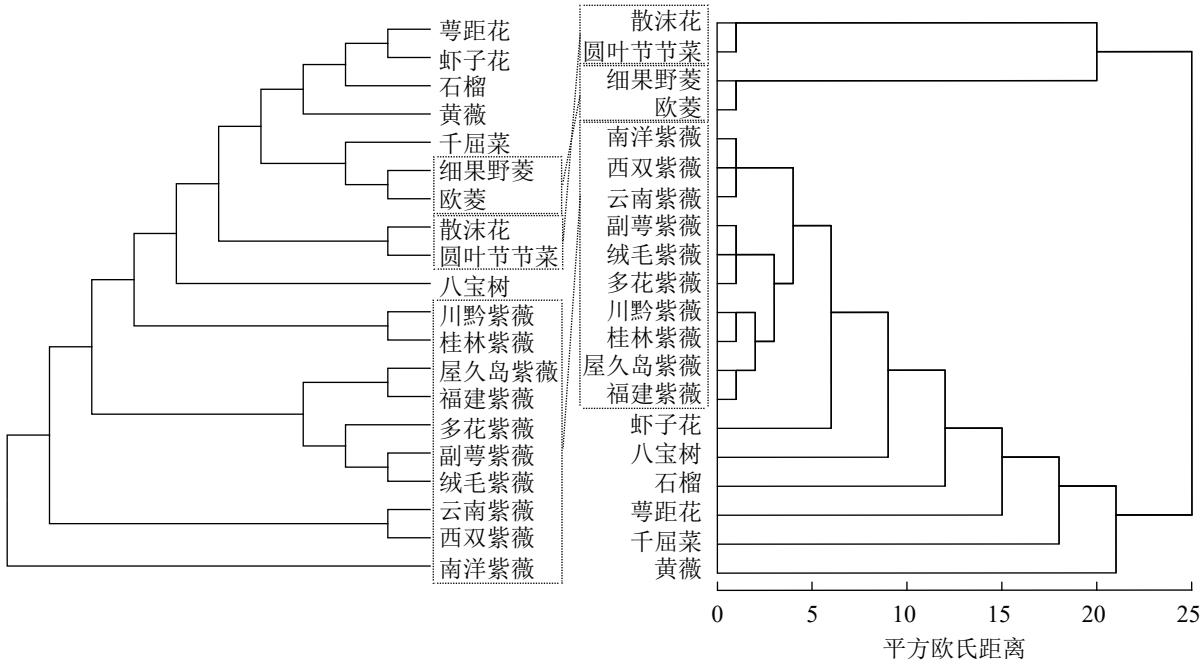


图 5　基于 *rbcL* 基因 CDS 的邻接树 (左) 和基于 59 个密码子 RSCU 的聚类树状图 (右)

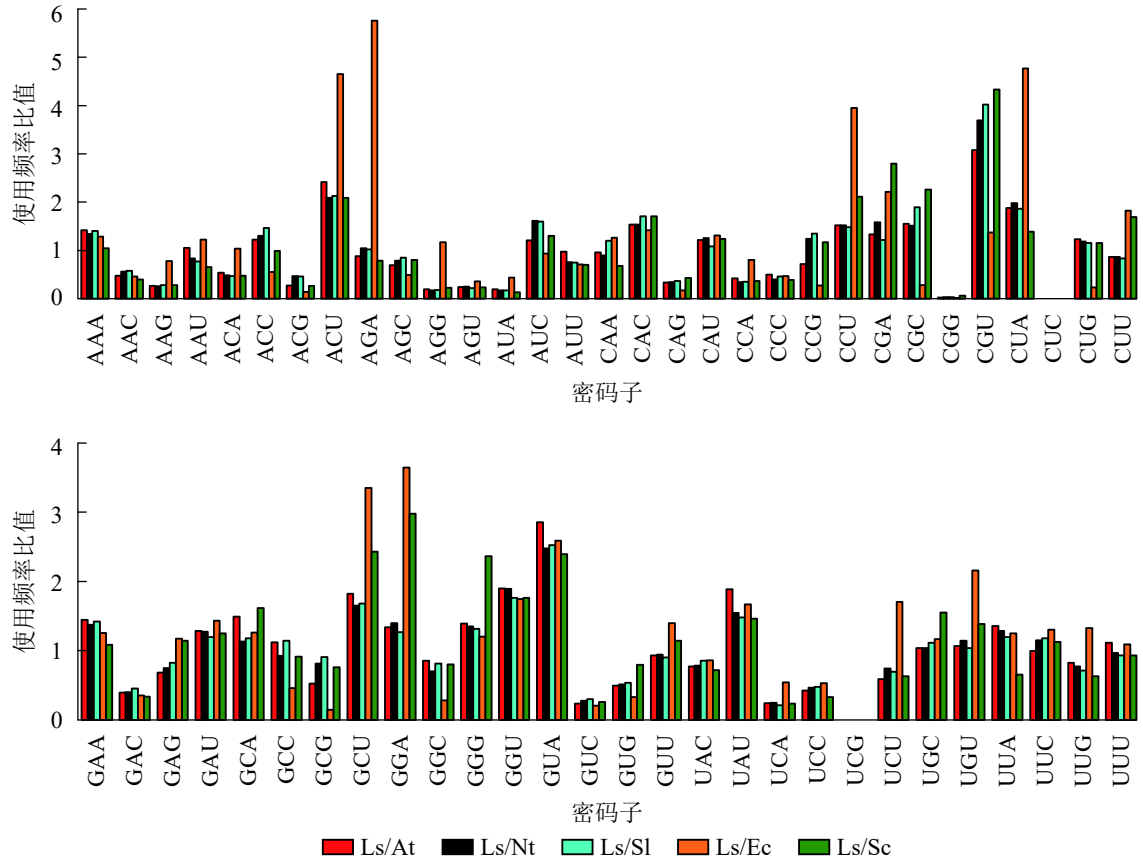
Figure 5　NJ tree based on CDS of *rbcL* genes (left) and cluster dendrogram based on RSCU of 59 codons (right)

芥、烟草和番茄分别存在 20、19 和 17 个使用频率相差较大的密码子，且最大值均出现在 CGU，初步说明相较于拟南芥和烟草，番茄更适合作为千屈菜科植物 *rbcL* 基因遗传转化的受体。

3　结论与讨论

特定的密码子使用偏好性是生物对环境变化适应性的体现，不同物种、不同功能基因的密码子偏好性存在明显差异。大部分双子叶植物密码子偏好 A/T 碱基结尾，单子叶植物则偏好 G/C 结尾[23]，与本研究中千屈菜科植物 *rbcL* 基因密码子 $A_{3s}+T_{3s}$ 远远大于 $G_{3s}+C_{3s}$ 的偏好性结果一致。李国灵等[13] 对红藻门 Rhodophyta 植物 *rbcL* 基因密码子偏好性研究也得到了类似结果，虽然红藻科和千屈菜科植物生活型、生理特性等相差较大，但千屈菜科也包括许多水生或湿生植物。两者研究结果显示：植物从水生向陆生过渡过程中，*rbcL* 基因密码子使用偏好性的变化可能较为稳定，这也许是 *rbcL* 基因受到强烈自然选择作用的结果。生物体内高表达的基因，其密码子偏好性也相对较强，反之亦然[24]。千屈菜科植物 *rbcL* 基因 ENC 较高，CAI 较低，说明千屈菜科植物 *rbcL* 基因整体的密码子使用偏好性不强，在植物体内表达水平也不高。但仍存在 CGU、CCU、ACU 等 13 个偏好性相对较强的密码子 (RSCU＞1.5)，其在氨基酸中残基含量也相对丰富。

密码子使用偏好性的影响因素包括碱基组成、突变、自然选择、漂变、基因长度、tRNA 丰度以及



Ls 表示千屈菜科物种; At 表示拟南芥; Nt 表示烟草; Sl 表示番茄; Ec 表示大肠埃希菌; Sc 表示酵母。无高度表示千屈菜科物种该密码子使用频率为 0

图 6 千屈菜科植物与模式生物密码子使用频率比值

Figure 6 Ratios of codon usage frequency of Lythraceae species to model organisms

基因表达水平的高低等，但最主要的压力来自于突变和自然选择[25]。本研究中，千屈菜科植物 *rbcL* 基因 GC_{3s} 和 GC、ENC 的相关性显著，表明密码子偏好性在一定程度上受到了碱基组成的影响，之前的研究也证明 GC_{3s} 和 GC 含量之间存在明显的线性关系[26]。但 GC_{3s} 与 GC_{12} 相关程度较低，且 GC_{3s} 集中分布在 0.275~0.300 内，KAWABE 等[23] 研究表明：密码子使用偏好性主要受自然选择的影响，而碱基突变的影响则较小，ENC 分析、中性分析、奇偶偏差分析也得出相同的结论。这可能是由于 *rbcL* 基因本身为叶绿体基因，分子进化速率相较于核基因更慢，且编码的二磷酸核酮糖羧化酶是参与光合作用的关键蛋白，相对比较保守，所以突变压力对其密码子使用偏好性的作用相对较弱；而正选择、协同进化等作用在陆生植物的 *rbcL* 基因中被证明广泛存在，也表明 *rbcL* 基因密码子使用偏好性可能广泛受到选择约束[27−28]。

与 RSCU 聚类分析结果相比，基于 CDS 的邻接树在理论上更接近真实的物种系统发育关系。两者相对一致的部分说明千屈菜科植物 *rbcL* 基因密码子使用特性与属间亲缘关系存在一定程度的对应；两者之间较为矛盾的分支可能是系统聚类仅选取单一 RSCU 数据分析导致的，结合密码子偏好性的其他参数，或许能获得更加一致的结果。由于单基因建树也可能会受到旁系同源基因干扰、水平基因转移等多种因素影响产生误差[29]，因此基于密码子偏好性的聚类分析也可对系统发生的研究内容进行一定补充。

转基因过程中，选择密码子使用偏好性相近的物种作为异源表达受体，有利于外源基因的高效表达[30]。千屈菜科植物多数都是木本植物，遗传转化体系尚未成熟，由于受限于同源物种生活史长、生长速度慢等因素，其基因功能研究十分依赖模式物种。通过与模式物种密码子使用频率的初步比较，酵母更适合作为千屈菜科植物 *rbcL* 基因的异源表达受体；与拟南芥、烟草相比，番茄的密码子使用频率与千屈菜科植物 *rbcL* 基因差异性最小，更适合作为 *rbcL* 基因功能验证的理想受体材料。但相对于番茄，拟南芥和烟草遗传转化体系建立相对较早，发展较为完善，已实现了多种木本植物叶绿体基因的遗传转

化，积累的技术经验较多，遗传转化的难度也相对较小[31]。在观赏植物研究中，番茄更多作为植物呈色相关基因的遗传转化受体，验证其在色素积累与代谢中的调控作用[32]。因此，密码子使用频率的比较结果仅能为千屈菜科植物 *rbcL* 基因异源表达受体选择提供初步的预测，受限于该科木本植物当前采样难度较大，且遗传转化体系尚未成熟建立等因素，最适的异源表达受体仍须在进一步的实验中进行深入研究和严格筛选。

4　参考文献

[1] WANG Liyuan, XIGN Huixian, YUAN Yanchao, *et al*. Genome-wide analysis of codon usage bias in four sequenced cotton species[J]. *PLoS One*, 2018, **13**(3): e0194372. doi: 10.1371/journal.pone.0194372.

[2] GUSTAFSSON C, GOVINDARAJAN S, MINSHULL J. Codon bias and heterologous protein expression [J]. *Trends Biotechnol*, 2004, **22**(7): 346 − 353.

[3] KAPRALOV M V, FILATOV D A. Widespread positive selection in the photosynthetic Rubisco enzyme[J]. *BMC Evol Biol*, 2007, **7**: 73. doi: 10.1186/1471-2148-7-73.

[4] SIQUEIRA A S, LIMA A R J, DALL' AGNOL L T. Comparative modeling and molecular dynamics suggest high carboxylase activity of the *Cyanobium* sp. CACIAM14 *rbcL* protein[J]. *J Mol Model*, 2016, **22**(3): 68. doi: 10.1007/s00894-016-2943-y.

[5] ANDERSSON I, BACKLUND A. Structure and function of Rubisco [J]. *Plant Physiol Biochem*, 2008, **46**(3): 275 − 291.

[6] 卞赛男, 常鹏杰, 王宁杭, 等. 氮素形态对喜树叶片生长、叶绿素荧光参数及叶绿体相关基因表达的影响[J]. 浙江农林大学学报, 2019, **36**(5): 908 − 916.
BIAN Sainan, CHANG Pengjie, WANG Ninghang, *et al*. Leaf growth, chlorophyll fluorescence characteristics, and expression of photosystem-related genes in *Camptotheca acuminata* with different N forms' fertilization [J]. *J Zhejiang A&F Univ*, 2019, **36**(5): 908 − 916.

[7] 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.
LI Donglin, JIN Yaqin, CUI Mengfan, *et al*. Growth, photosynthesis and ultrastructure of mesophyll cells for *Cercidiphyllum japonicum* seedlings with shading in summer [J]. *J Zhejiang A&F Univ*, 2020, **37**(3): 496 − 505.

[8] QIN Haining, SHIRLEY A G, MICHAEL G G. Lythraceae[M]//WU Zhengyi, PETER H R, HONG Deyuan. *Flora of China vol* 13. Beijing: Science Press, 2007: 274 − 290.

[9] HUANG Yelin, SHI Suhua. Phylogenetics of Lythraceae sensu lato: a preliminary analysis based on chloroplast *rbcL* gene, *psaA-ycf*3 spacer and nuclear rDNA internal transcribed spacer (ITS) sequences [J]. *Int J Plant Sci*, 2005, **163**(2): 215 − 225.

[10] ZHENG Gang, WEI Lingling, MA Li, *et al*. Comparative analysis of chloroplast genomes from 13 *Lagerstroemia* (Lythraceae) species: identification of highly divergent regions and inference of phylogenetic relationships [J]. *Plant Mol Biol*, 2020, **102**(6): 659 − 676.

[11] 赵洋, 杨培迪, 刘振, 等. 13 种植物 *actin* 基因的密码子使用特性分析[J]. 南方农业学报, 2016, **47**(4): 519 − 523.
ZHAO Yang, YANG Peidi, LIU Zhen, *et al*. Characterization of codon usage of *actin* genes for 13 species of plants [J]. *J Southern Agric*, 2016, **47**(4): 519 − 523.

[12] 朱沛煌, 陈妤, 朱灵芝, 等. 马尾松转录组密码子使用偏好性及其影响因素[J]. 林业科学, 2020, **56**(4): 74 − 81.
ZHU Peihuang, CHEN Yu, ZHU Lingzhi, *et al*. Codon usage bias and its influencing factors in *Pinus massoniana* transcriptome [J]. *Sci Silv Sin*, 2020, **56**(4): 74 − 81.

[13] 李国灵, 陶文, 高诗晨, 等. 红藻 *rbcL* 基因密码子偏爱性分析[J]. 分子植物育种, 2019, **18**(1): 109 − 117.
LI Guoling, TAO Wen, GAO Shichen, *et al*. The codon usage analysis in *rbcL* gene within Rhodophyta [J]. *Mol Plant Breed*, 2019, **18**(1): 109 − 117.

[14] 吴妙丽, 陈世品, 陈辉. 竹亚科叶绿体基因组的密码子使用偏性分析[J]. 森林与环境学报, 2019, **39**(1): 9 − 14.
WU Miaoli, CHEN Shipin, CHEN Hui. Condon preference of chloroplast genome of Bambusoideae [J]. *J For Environ*, 2019, **39**(1): 9 − 14.

[15] 李慧姬, 吉雪花, 朱冉冉, 等. 10 种植物 *PSY* 基因密码子使用偏好性分析[J]. 西北农业学报, 2020, **29**(2): 276 − 284.

LI Huiji, JI Xuehua, ZHU Ranran, *et al*. Codon usage bias analysis for octahydrolycopene synthase gene (*PSY*) from ten plant species [J]. *Acta Agric Boreali-Occident Sin*, 2020, **29**(2): 276 − 284.

[16] 巫伟峰, 陈明杰, 陈发兴. ‘皇冠李’苹果酸转运体基因 *ALMT*4、*ALMT*9 和 *tDT* 密码子偏好性分析[J]. 农业生物技术学报, 2020, **28**(1): 42 − 57.
WU Weifeng, CHEN Mingjie, CHEN Faxing. Codon bias analysis of *Prunus salicina* ‘Huangguan’ malate transporter *ALMT*4, *ALMT*9 and *tDT* genes [J]. *J Agric Biotechnol*, 2020, **28**(1): 42 − 57.

[17] 彭丽云, 王云, 孙雪丽, 等. 苋菜 *AmMYB*2 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.
PENG Liyun, WANG Yun, SUN Xueli, *et al*. Codon bias and evolutionary analysis of the *AmMYB*2 gene in *Amaranthus tricolor* L. [J]. *Chin J Appl Environ Biol*, 2019, **25**(3): 679 − 686.

[18] 晁岳恩, 吴政卿, 杨会民, 等. 11 种植物 *psbA* 基因的密码子偏好性及聚类分析[J]. 核农学报, 2011, **25**(5): 927 − 932.
CHAO Yueen, WU Zhengqing, YANG Huimin, *et al*. Cluster analysis and codon usage bias studies on *psbA* genes from 11 plant species [J]. *J Nucl Agric Sci*, 2011, **25**(5): 927 − 932.

[19] JIANG Yue, DENG Feng, WANG Hualin, *et al*. An extensive analysis on the global codon usage pattern of baculoviruses [J]. *Arch Virol*, 2008, **153**(12): 2273 − 2282.

[20] 李凌烜, 陈安琪, 黄凯, 等. 豆科植物 *dxr* 基因密码子偏好性分析[J]. 生物学杂志, 2020, **37**(1): 30 − 34.
LI Lingxuan, CHEN Anqi, HUANG Kai, *et al*. Analysis of codon bias of *dxr* gene in Leguminous plants [J]. *J Biol*, 2020, **37**(1): 30 − 34.

[21] SHARP P M, LI W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [J]. *J Mol Evol*, 1986, **24**(1/2): 28 − 38.

[22] 王云, 彭丽云, 苏立遥, 等. 龙眼 *Hsf* 基因家族密码子使用模式分析[J]. 分子植物育种, 2019, **17**(17): 5595 − 5603.
WANG Yun, PENG Liyun, SU Liyao, *et al*. Codon usage pattern analysis of *Dimocarpus longan* Lour. *Hsf* gene family [J]. *Mol Plant Breed*, 2019, **17**(17): 5595 − 5603.

[23] KAWABE A, MIYASHITA N T. Patterns of codon usage bias in three dicot and four monocot plant species [J]. *Genes Genet Syst*, 2003, **78**(5): 343 − 352.

[24] 王宇, 周俊良, 唐冬梅, 等. 阔叶猕猴桃叶绿体基因组特征及密码子偏好性分析[J]. 种子, 2020, **39**(5): 13 − 19.
WANG Yu, ZHOU Junliang, TANG Dongmei, *et al*. Analysis of chloroplast genome characteristics and codon preference in broad-leaf Kiwifruit [J]. *Seed*, 2020, **39**(5): 13 − 19.

[25] 原晓龙, 王毅, 张劲峰. 灰毛浆果楝叶绿体基因组密码子使用特征分析[J]. 森林与环境学报, 2020, **40**(2): 195 − 202.
YUAN Xiaolong, WANG Yi, ZHANG Jinfeng. Characterization of codon usage in *Cipadessa cinerascens* chloroplast genome [J]. *J For Environ*, 2020, **40**(2): 195 − 202.

[26] KUSUMI J, TACHIDA H. Compositional properties of green-plant plastid genomes [J]. *J Mol Evol*, 2005, **60**(4): 417 − 425.

[27] WANG Mingcong, KAPRALOV M V, ANISIMOVA M. Coevolution of amino acid residues in the key photosynthetic enzyme Rubisco[J]. *BMC Evol Biol*, 2011, **11**(1): 266. doi: 10.1186/1471-2148-11-266.

[28] LIU Lei, ZHAO Bo, ZHANG Yu, *et al*. Adaptive evolution of the *rbcL* gene in Brassicaceae [J]. *Biochem Syst Ecol*, 2012, **44**: 13 − 19.

[29] MIWA H, ODRZYKOSKI I J, MATSUI A, *et al*. Adaptive evolution of *rbcL* in *Conocephalum* (Hepaticae, bryophytes) [J]. *Gene*, 2009, **441**(1/2): 169 − 175.

[30] DANIEL H, LIN C S, YU Ming, *et al*. Chloroplast genomes: diversity, evolution, and applications in genetic engineering[J]. *Genome Biol*, 2016, **17**(1): 134. doi: 10.1186/s13059-016-1004-2.

[31] 刘佳音, 李儒剑, 齐双慧, 等. 叶绿体遗传转化系统及其应用进展[J]. 安徽农业科学, 2020, **48**(6): 16 − 19.
LIU Jiayin, LI Rujian, QI Shuanghui, *et al*. Chloroplast genetic transformation system and its application progress [J]. *J Anhui Agric Sci*, 2020, **48**(6): 16 − 19.

[32] SUI Xiaoming, WANG Yang, ZHAO Mingyuan, *et al*. Cloning and expression analysis of *RrGT*2 gene related to anthocyanin biosynthesis in *Rosa rugosa* [J]. *Am J Plant Sci*, 2018, **9**(10): 2008 − 2019.