

文章编号 1672-6634(2022)06-0096-08

DOI 10.19728/j.issn1672-6634.2022040017

锡金海棠叶绿体基因组结构及密码子偏好性分析

陈玲慧¹,王 慧²,刘 岩¹,钱关泽¹

(1.聊城大学 生命科学学院,山东 聊城 252059;2.聊城大学 学报编辑部,山东 聊城 252059)

摘 要 为研究锡金海棠叶绿体基因组结构、明确其叶绿体基因组密码子使用模式,通过 NOVOPlasty、geneious 和 codonW 等软件,组装出锡金海棠叶绿体基因组,在注释后分析叶绿体基因组结构,并筛选出 53 条蛋白编码序列(CDS)用于密码子偏好性分析。结果表明,锡金海棠叶绿体基因组全长 160 088 bp,GC 含量为 36.6%;不同碱基位置 GC 含量分别为 GC1(46.98%)>GC2(39.46%)>GC3(28.30%),表明其密码子末位碱基偏向以 A/U 结尾;中性图分析表明 GC12 与 GC3 未有明显相关性,即密码子第 1、2 位碱基与第 3 位碱基存在一定差异;根据 ENC-plot 和 PR2-plot 分析,影响密码子偏好性的主要因素是自然选择;最后确定 UUG、CUU 和 AUU 等 14 个密码子为锡金海棠叶绿体基因组的最优密码子。研究结果对今后研究锡金海棠亲缘关系、系统发育、物种鉴定提供了数据支持,对优化目标基因、种源的鉴定及异源表达基因的改造和系统发育分析提供了理论依据。

关键词 锡金海棠;叶绿体基因组;密码子偏好性

中国分类号 Q812

文献标识码 A

开放科学(资源服务)标识码(OSID)



Analysis of Structure and Codon Usage Bias in the Chloroplast Genome of *Malus sikkimensis*

CHEN Linghui¹, WANG Hui², LIU Yan¹, QIAN Guanze¹

(1.School of Life Science, Liaocheng University, Liaocheng 252059, China; 2.Editorial Department, Journal of Liaocheng University, Liaocheng 252059, China)

Abstract In order to study the chloroplast genome structure of *Malus sikkimensis* and determine the use pattern of its chloroplast genome codon, the chloroplast genome of *Malus sikkimensis* was assembled using NOVOPlasty, Geneious and codonW software, and the chloroplast genome structure was analyzed after annotation. The 53 protein coding sequences (CDS) were screened for codon bias analysis. The results showed that the chloroplast genome of *Malus sikkimensis* was 160088bp with GC content of 36.6%. The GC contents of different base positions were GC1 (46.98%)>GC2(39.46%)>GC3(28.30%), indicating that their codon terminal bases were biased to end with A/U. Neutrality plot analysis showed none significant correlation between GC12 and GC3, that is, there were some differences between the first and second bases of the codon and the third base. ENC-plot analysis and PR2-plot analysis show that codon preference is mainly affected by natural selection. Finally, 14 codons, including UUG, CUU

收稿日期:2022-04-17

基金项目:国家自然科学基金项目(31170178)资助

通讯作者:钱关泽,男,汉族,博士,教授,研究方向:种子植物分类学, E-mail:qianguanze@lcu.edu.cn.

and AUU, were identified as the optimal codons for the chloroplast genome of *Malus sikkimensis*. The results of this study will provide data support for future studies on the kinship, phylogeny, and species identification of *Malus sikkimensis*, and provide a theoretical basis for optimizing target genes, seed identification, and heterologous expression gene modification.

Key words *malus sikkimensis*; chloroplast genome; codon preference

0 引言

叶绿体是植物进行光合作用的细胞器,为植物生活提供能量,叶绿体基因组(cpDNA)通常大小为 110~170 kb^[1]。叶绿体基因组在进行物种鉴定与亲缘关系分析时通常有以下优势:其基因组较小、序列简单,较易通过测序组装后获得;同时其基因序列高度保守、且属于母系遗传,使得叶绿体基因组适用于亲缘关系与系统发育分析^[2]。目前 NCBI 数据库中的叶绿体基因组数据不断被补充,包括但不限于苹果属(*Malus*)^[3]、蕨麻属(*Argentina*)^[4]、淫羊藿属(*Epimedium*)^[5]等属内植物,这也从侧面表明叶绿体基因组的广泛应用。

蛋白质由氨基酸组成,20 种氨基酸又由 61 个密码子和 3 个终止密码子所编码^[6],编码同一种氨基酸的密码子为同义密码子。在蛋白质翻译的过程中,尽管几种不同密码子可以编码同种氨基酸,但是会出现某一氨基酸频繁被同一密码子编码的现象,即密码子偏好性^[7],而不同密码子会对基因的功能及表达产生影响^[8]。不同物种叶绿体基因组密码子偏好性具有差异,一般基因表达越强,其密码子偏好性也越强,反之其偏好性则低^[9]。此前研究表明,密码子偏好性主要通过自然选择与突变来进行平衡^[10],因此对密码子偏好性的研究在基因组水平,如分子进化等方面具有重要意义。现已在云南油杉(*Keteleeria evelyniana*)^[11]、澳洲坚果光壳种(*Macadamia integrifolia*)^[12]、菠萝(*Ananas comosus*)^[13]、睡莲属(*Nymphaea*)^[14]等多种植物中展开密码子偏好性分析。

锡金海棠(*Malus sikkimensis*)是苹果属(*Malus*)植物,属于蔷薇科(Rosaceae)苹果亚科(Subfam. Maloideae Weber),主要分布于我国云南、四川西部及西藏地区。锡金海棠在《中国珍稀濒危植物名录》中被列为国家二级保护物种,同时还被 IUCN(国家自然保护联盟)分别评为易危物种,以及被《中国物种红色名录》收录^[15]。梁国鲁^[16]对锡金海棠进行带型分析后发现其以着丝粒带(C)为主,华利源^[17]在染色体层面进行核型分析后发现锡金海棠染色体数目为 68 条,为四倍体。到目前锡金海棠叶绿体基因组结构及密码子偏好性的研究还未见报道,因此,本文对锡金海棠的叶绿体基因组结构及密码子偏好性进行分析,筛选出最优密码子,为今后对锡金海棠叶绿体基因组、系统发育研究及开发应用提供科学依据和理论参考。

1 材料和方法

1.1 材料数据获取

从 NCBI 数据库中下载锡金海棠(*M. sikkimensis*)原始测序数据, SRR 登录号为 SRR15691192。通过 NO-VOPlasty 4.3.1^[18]组装, PGA 注释后获得其叶绿体全基因组。

1.2 叶绿体基因组特征分析

将所获数据导入 geneious 2021.2.2^[19],根据注释结果,统计锡金海棠(*M. sikkimensis*)的蛋白编码基因(CDS)数目、tRNA 数目、rRNA 数目以及 GC 含量。利用 Chloroplot 软件绘制其叶绿体基因组物理图谱。

1.3 密码子偏好性分析

使用 geneious 2021.2.2 筛选出 53 条长度大于 300 bp 的 CDS 序列。通过在线软件 CUSP(<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>)统计各基因密码子的第 1 位(GC1)、第二位(GC2)及第三位(GC3)的 GC 含量。使用 codonW(<http://codonw.sourceforge.net>)获得 GC 含量、相对同义密码子使用度(Relative Synonymous Codon Usage, RSCU)以及有效密码子数(Effective Number of Codon, ENC)。当 RSCU 值大于 1 时,表明密码子使用频率较其他密码子高,为高频密码子,并且偏好性强,反之偏好性则弱;当 RSCU 值等于 1 时,该密码子无偏好性^[20]。最终结果使用 Excel 和 SPSS 进行统计分析。

以 GC3 为横坐标、GC12 为纵坐标进行中性绘图分析,散点图中的每一个点均代表一个基因,当其沿对角线分布时则代表碱基未有显著差异,此时的偏好性是由基因突变所导致,反之则是由选择压力所导致^[11]。以

GC3 为横坐标、ENC 值为纵坐标做 ENC-plot 绘图分析,并以公式 $ENC=2+GC3+29/[GC32+(1-GC3)2]$ 绘制 ENC 期望值标准曲线。该标准曲线代表在没有自然选择时的 ENC 值^[12]。以 $G3/(G3+C3)$ 为横坐标、 $A3/(A3+T3)$ 为纵坐标进行 PR2-plot 分析,其平面中心代表密码子无偏倚情况出现,即 $A=T, C=G$ ^[13]。

为确定锡金海棠(*M.sikkimensis*)的最优密码子,根据 53 条 CDS 序列的 ENC 值,提取最低 10%与最高值 10%的基因,分别建立高、低表达基因库,计算出二者的 Δ_{RSCU} 值,当 $\Delta_{RSCU} \geq 0.08$ 时的密码子被定义为高表达优越密码子,若该密码子同时为高频密码子,则将其定义为最优密码子^[21]。

2 结果与分析

2.1 叶绿体基因组基本特征

通过对锡金海棠组装后分析,其 cpDNA 呈典型的环状四分体结构,包括两个反向重复区(IRA 和 IRB, 26 358 bp),一个大单拷贝区(LSC,88 188 bp)和一个小单拷贝区(SSC,19 184 bp),全长为 160 088 bp,总的 GC 含量为 36.6%(图 1)。注释结果显示,锡金海棠叶绿体基因组含有 132 个基因,包括 87 个蛋白编码基因,37 个 tRNA 基因以及 8 个 rRNA 基因,其中有重复基因 18 个(7 个为蛋白编码基因,7 个为 tRNA 基因,4 个为 rRNA 基因)。

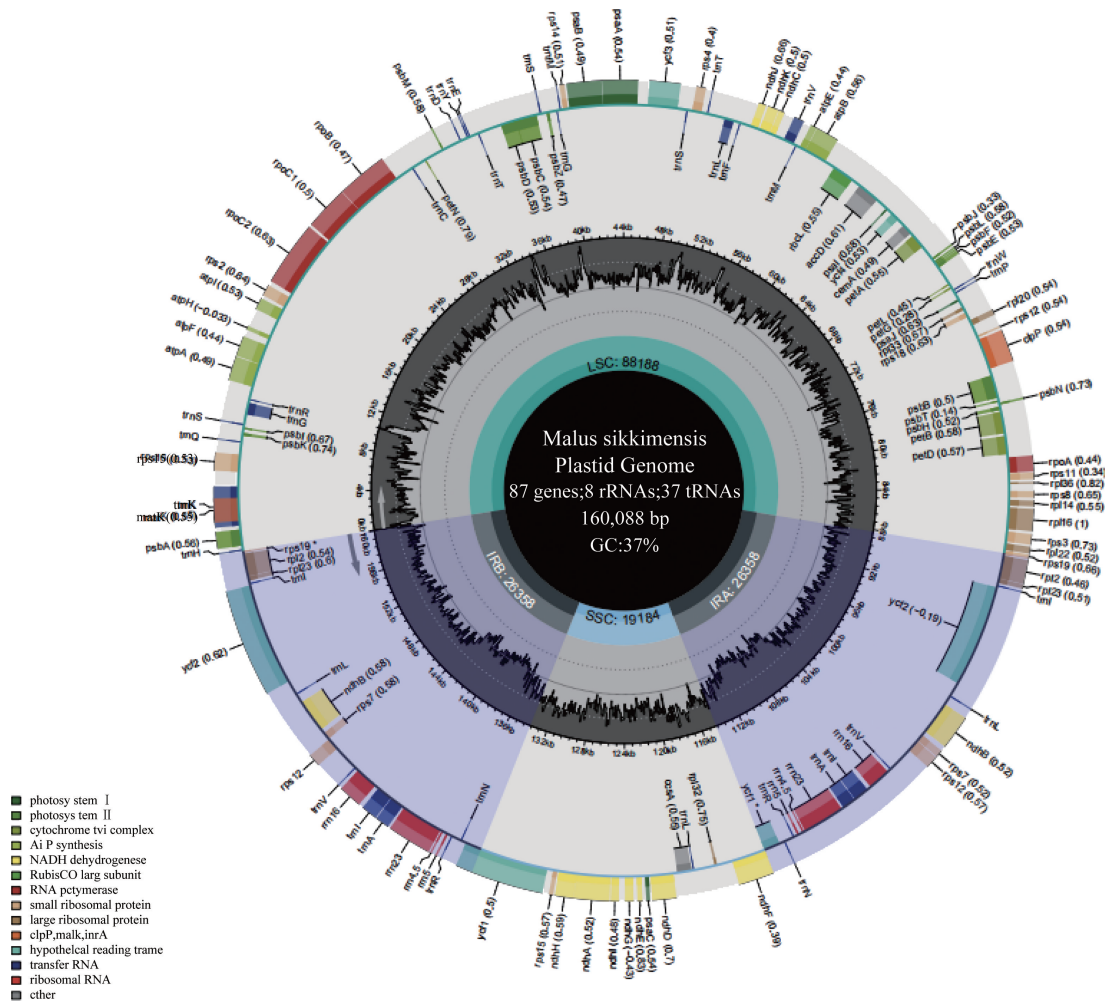


图 1 锡金海棠叶绿体基因组物理图谱

2.2 密码子偏好性分析

2.2.1 密码子组成分析。通过 CodonW 分析锡金海棠 53 条 CDS 序列的 GC 含量及 ENC 值(表 1)发现,其叶绿体基因组各基因 GC 含量平均为 38.25%,其中第一位碱基 GC 平均含量为 46.98%、第二位 GC 平均含量为 39.46%、第三位 GC 平均含量为 28.30%。GC3 平均含量最低,这表明第三位碱基更倾向于以 A/U 结尾。

ENC 值可以用于衡量同义密码子的偏好程度,取值范围通常在 20~61 之间,以 ENC 值为 45 作为分界点,

大于 45 说明密码子偏好性低,反之则说明其偏好性较高。锡金海棠 ENC 值取值范围为 34.30~56.39,平均值为 47.51,且 53 条 CDS 序列中存在 41 条 ENC 值大于 45,表明锡金海棠的密码子偏好性偏弱。

表 1 锡金海棠叶绿体基因组 53 个 CDS 密码子的 GC 含量及 ENC 值

基因	GC1	GC2	GC3	GC	ENC	基因	GC1	GC2	GC3	GC	ENC
accD	39.23%	35.57%	28.66%	34.49%	45.63	psbB	54.81%	46.37%	30.26%	43.81%	48.30
atpA	54.72%	40.35%	25.79%	40.29%	46.91	psbC	53.59%	46.20%	32.91%	44.23%	47.07
atpB	57.40%	42.80%	26.57%	42.26%	45.62	psbD	51.98%	43.22%	33.05%	42.75%	46.23
atpE	50.75%	40.30%	29.85%	40.30%	48.62	rbcL	58.19%	43.49%	27.73%	43.14%	46.82
atpF	45.95%	34.05%	34.59%	38.20%	45.96	rpl2	51.27%	47.64%	31.64%	43.52%	53.23
atpI	49.19%	36.69%	28.23%	38.04%	48.58	rpl14	55.28%	38.21%	28.46%	40.65%	48.07
ccsA	34.16%	36.65%	25.47%	32.09%	46.67	rpl16	50.00%	54.41%	25.74%	43.38%	34.30
cemA	37.39%	28.70%	30.43%	32.17%	49.54	rpl20	38.14%	40.68%	23.73%	34.18%	43.98
clpP	55.84%	38.07%	29.44%	41.12%	55.65	rpl22	41.55%	35.21%	28.17%	34.98%	50.73
matK	39.41%	30.30%	27.13%	32.28%	48.31	rpoA	45.48%	30.72%	26.51%	34.24%	49.70
ndhA	43.96%	37.64%	23.63%	35.07%	44.59	rpoB	50.33%	38.07%	28.25%	38.88%	48.43
ndhB	41.68%	39.33%	31.31%	37.44%	48.44	rpoC1	50.15%	38.12%	28.15%	38.81%	50.16
ndhC	47.11%	33.88%	25.62%	35.54%	50.69	rpoC2	45.27%	37.48%	28.75%	37.17%	50.53
ndhD	40.52%	38.12%	28.94%	35.86%	47.26	rps2	44.30%	42.19%	29.11%	38.54%	50.71
ndhE	39.22%	34.31%	31.37%	34.97%	51.83	rps3	48.86%	32.88%	23.74%	35.16%	48.72
ndhF	37.50%	35.77%	22.21%	31.83%	43.71	rps4	51.98%	37.13%	28.22%	39.11%	49.98
ndhG	43.50%	32.20%	29.38%	35.03%	50.04	rps7	52.56%	44.87%	23.72%	40.38%	44.11
ndhH	51.02%	36.04%	28.68%	38.58%	49.99	rps8	40.74%	39.26%	22.22%	34.07%	35.74
ndhI	43.71%	35.93%	25.15%	34.93%	49.36	rps11	51.80%	56.12%	25.18%	44.36%	49.92
ndhJ	50.94%	37.74%	28.93%	39.20%	51.89	rps12	51.61%	47.58%	29.03%	42.74%	43.26
ndhK	45.37%	46.70%	25.11%	39.06%	47.44	rps14	43.56%	49.50%	32.67%	41.91%	39.46
petA	54.52%	36.14%	32.40%	41.02%	49.86	rps18	37.50%	44.23%	23.08%	34.94%	39.10
petB	49.07%	41.67%	30.56%	40.43%	44.39	ycf1	35.32%	28.19%	24.79%	29.43%	46.32
petD	50.93%	38.51%	26.09%	38.51%	42.07	ycf2	41.58%	34.17%	36.58%	37.44%	52.73
psaA	52.86%	43.41%	32.09%	42.79%	51.46	ycf3	47.34%	37.87%	29.59%	38.26%	56.39
psaB	48.98%	42.99%	30.48%	40.82%	48.30	ycf4	42.16%	40.54%	29.19%	37.30%	50.14
psbA	49.72%	42.94%	31.36%	41.34%	41.24	平均值	46.98%	39.46%	28.30%	38.25%	47.51

对密码子 GC 含量、ENC 值进行相关性分析后发现(表 2),GC1、GC2、GC3 均与 GC 呈极显著相关,GC1 与 GC2 也呈极显著相关,但 GC1、GC2 均为表现出与 GC3 的显著相关性。这表明密码子第 3 位碱基组成有差异,锡金海棠密码子偏好性受自然选择影响较大。而 ENC 值与 GC2、GC3 呈极显著相关关系,表明密码子第 2、3 位对偏好性有一定影响。

表 2 锡金海棠密码子的 GC 含量及各参数相关性分析

因素	GC1	GC2	GC3	GC
GC2	0.441 **	-	-	-
GC3	0.222	0.022	-	-
GC	0.850 **	0.775 **	0.424 **	-
ENC	0.148	-0.361 **	0.375 **	0.002

注: ** 表示在 0.01 水平相关性极显著; * 表示在 0.05 水平相关性显著。

进一步对锡金海棠 53 条 CDS 序列的密码子(除甲硫氨酸与色氨酸)进行 RSCU 分析后显示(表 3),有 29 个密码子的 RSCU 值大于 1.00,其中 15 个以 U 结尾,13 个以 A 结尾,1 个以 G 结尾,这表明更多密码子偏好以 A 和 U 结尾。

表 3 锡金海棠叶绿体基因组氨基酸相对同义密码子使用度

氨基酸	密码子	数目	相对密码子使用度	氨基酸	密码子	数目	相对密码子使用度
Phe	UUU	899	1.30	Tyr	UAU	751	1.61
	UUC	489	0.70		UAC	180	0.39
Leu	UUA	836	1.94	His	CAU	464	1.54
	UUG	535	1.24		CAC	137	0.46
	CUU	555	1.29	Gln	CAA	684	1.53
	CUC	161	0.37		CAG	211	0.47
	CUA	337	0.78	Asn	AAU	930	1.52
	CUG	162	0.38		AAC	290	0.48
Ile	AUU	1032	1.46	Lys	AAA	981	1.49
	AUC	407	0.58		AAG	333	0.51
	AUA	678	0.96	Asp	GAU	852	1.62
Val	GUU	485	1.46		GAC	198	0.38
	GUC	141	0.43	Glu	GAA	968	1.48
	GUA	510	1.54		GAG	342	0.52
	GUG	190	0.57	Cys	UGU	205	1.47
Ser	UCU	528	1.68		UGC	74	0.53
	UCC	308	0.98	Arg	CGU	316	1.28
	UCA	371	1.18		CGC	106	0.43
	UCG	182	0.58		CGA	336	1.36
Pro	CCU	387	1.54	Ser	CGG	115	0.46
	CCC	193	0.77		AGU	379	1.20
	CCA	291	1.16	Arg	AGC	120	0.38
	CCG	136	0.54		AGA	453	1.83
Thr	ACU	497	1.59	Gly	AGG	158	0.64
	ACC	230	0.74		GGU	541	1.30
	ACA	388	1.24		GGC	178	0.43
	ACG	135	0.43		GGA	674	1.62
Ala	GCU	597	1.84	Ter	GGG	269	0.65
	GCC	203	0.62		UAA	34	1.76
	GCA	364	1.12		UAG	14	0.72
	GCG	136	0.42		UGA	10	0.52

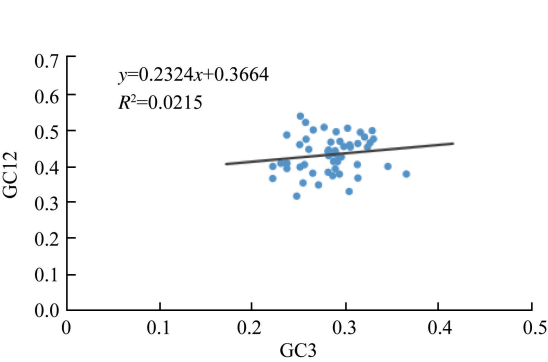


图 2 锡金海棠叶绿体基因组中性绘图分析

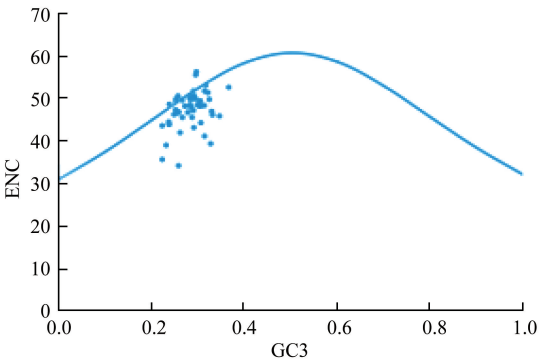


图 3 锡金海棠叶绿体基因的 ENC-plot 分析

2.2.2 中性绘图分析。中性绘图分析(图2)表明,锡金海棠密码子 GC12 的值介于 0.317 6~0.539 6 之间,GC3 的值介于 0.222 1~0.365 8 之间。其相关系数为 0.146 8,未呈现显著相关性,回归系数为 0.021 5,表明 GC12 与 GC3 未有明显相关性,密码子第 1、2 位碱基与第 3 位碱基存在一定差异,说明自然选择对锡金海棠叶绿体基因密码子偏好有较大影响。

2.2.3 ENC-plot 绘图分析。ENC-plot 绘图分析(图3)显示,部分基因的实际 ENC 值与预期 ENC 值趋于一致,但也有部分基因的实际 ENC 值偏离标准曲线,位于标准曲线下方。为比较实际 ENC 值与预期 ENC 值之间的差异,计算 ENC 比值并得到以下结果(表4)。ENC 值在 $-0.05 \sim 0.05$ 区间内的基因有 23 个,在 $-0.05 \sim 0.05$ 区间外的基因有 30 个,这表明大部分基因与预期 ENC 值存在一定差异,表明锡金海棠叶绿体基因密码子偏好更受自然选择影响。

表4 ENC 比值频数分布

组限	组中值	频数	频率
$-0.15 \sim -0.05$	-0.1	2	0.04
$-0.05 \sim 0.05$	0.0	23	0.43
$0.05 \sim 0.15$	0.1	20	0.38
$0.15 \sim 0.25$	0.2	6	0.11
$0.25 \sim 0.35$	0.3	2	0.04
合计		53	1

2.2.4 PR2-plot 分析。通过 PR2-plot 分析编码基因中密码子 A、G 和 T、C 的使用频率关系。结果如图4所示,锡金海棠叶绿体基因多数分布于右下象限,这表明密码子碱基的使用频率有差异且 $T > A$ 、 $G > C$,这也说明锡金海棠密码子偏好性不仅受自然选择影响,同时也可能受其他因素影响。

2.2.5 最优密码子的确定。通过分析锡金海棠叶绿体基因组氨基酸相对同义密码子使用度,有 29 个密码子 RSCU 值大于 1,被确定为高频密码子。结合表5,筛选出 22 个 $\Delta RSCU \geq 0.08$ 的密码子,被定为高表达密码子,其中以 A 结尾的密码子有 5 个,以 U 结尾的密码子有 8 个,以 C 结尾的密码子有 6 个,以 G 结尾的密码子有 3 个。当密码子满足 $RSCU > 1$ 且 $\Delta RSCU \geq 0.08$ 条件时,我们将其确定为最优密码子。通过结合表3与表5分析,共确定 14 个最优密码子,分别为 UAU、UUG、CUU、CAA、AUU、GUU、GAA、UCU、CGU、CGA、AGA、GGU、ACA、GCU,其中有 8 个以 U 结尾,有 5 个以 A 结尾,1 个以 G 结尾。

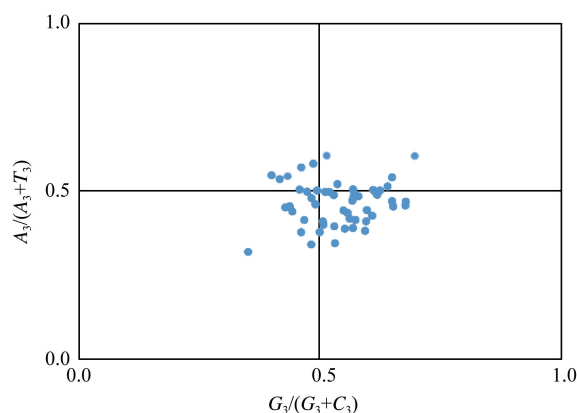


图4 锡金海棠叶绿体基因的 PR2-plot 分析

3 结论与讨论

cpDNA 较小、结构保守,易于测序和组装,同时叶绿体属于单亲遗传,其基因组包含丰富的核苷酸信息,因此通常被认为是研究植物进化关系、系统发育的理想模型。密码子在基因组与蛋白质的联系中有重要作用,在物种进化过程中,其通常会形成特有的使用模式,因此不同物种的密码子在使用偏性方面存在一定差异。密码子的第 3 位具有兼并性特点,通常第 3 位碱基所受到的选择压力较前两位碱基要小,因此,分析密码子的第 3 位碱基对研究密码子偏好性具有重要意义。

本研究通过对锡金海棠的 cpDNA 结构和密码子使用模式进行分析,发现锡金海棠叶绿体基因组呈环状四分体结构,包括两个反向重复区(IRA 和 IRB)、一个大单拷贝区(LSC)和一个小单拷贝区(SSC),叶绿体基因组大小为 160 088 bp,GC 含量为 36.6%。锡金海棠共含有 132 个基因,包括 87 个蛋白编码基因,37 个 tRNA 基因和 8 个 rRNA 基因。

表 5 锡金海棠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		Δ RSCU	AA	密码子	高表达基因		低表达基因		Δ RSCU
		数量	RSCU	数量	RSCU				数量	RSCU	数量	RSCU	
Phe	UUU	18	1.42	100	1.40	0.02	Tyr	<u>UAU*</u>	18	1.80	97	1.57	0.23
	UUC	21	0.58	84	0.60	-0.02		UAC	6	0.20	25	0.43	-0.23
Leu	UUA	22	1.74	65	2.05	-0.31	His	CAU	10	0.88	68	1.61	-0.73
	<u>UUG*</u>	17	1.41	77	1.15	0.26		CAC**	8	0.72	16	0.39	0.33
	<u>CUU***</u>	18	2.20	75	1.24	0.96	Gln	<u>CAA**</u>	21	1.87	83	1.43	0.44
	CUC	0	0.00	28	0.36	-0.36		CAG	1	0.13	35	0.57	-0.44
	CUA	11	0.61	41	0.61	0	Asn	AAU	26	1.40	134	1.57	-0.17
	CUG	1	0.04	28	0.59	-0.55		AAC*	13	0.60	43	0.43	0.17
Ile	<u>AUU*</u>	35	1.34	97	1.18	0.16	Lys	AAA	33	1.37	122	1.54	-0.17
	AUC	15	0.46	59	0.62	-0.16		AAG	6	0.23	61	0.46	-0.23
	AUA	25	1.21	87	1.20	0.01	Asp	GAU	8	0.69	143	1.71	-1.02
Met	AUG	26	1.00	68	1.00	0		GAC***	8	1.31	30	0.29	1.02
Val	<u>GUU***</u>	21	2.01	41	1.27	0.74	Glu	<u>GAA*</u>	33	1.50	116	1.36	0.14
	GUC	1	0.11	22	0.59	-0.48		GAG	10	0.50	62	0.64	-0.14
	GUA	17	0.98	37	1.49	-0.51	Cys	UGU	4	1.00	25	1.17	-0.17
	GUG	2	0.30	23	0.65	-0.35		UGC	2	0.60	13	0.83	-0.23
Ser	<u>UCU***</u>	26	2.30	72	1.61	0.69	Trp	UGG	16	0.80	56	1.00	-0.20
	UCC	7	0.52	54	1.03	-0.51		<u>CGU*</u>	17	1.29	28	1.14	0.15
	UCA	5	0.47	60	0.97	-0.5		CGC*	8	0.58	14	0.38	0.20
	UCG	5	0.56	32	0.59	-0.03		<u>CGA*</u>	22	1.55	44	1.28	0.27
Pro	CCU	18	1.62	47	1.83	-0.21		CGG	4	0.28	23	0.55	-0.27
	CCC	5	0.60	23	0.61	-0.01		Ser AGU	14	1.24	44	1.54	-0.3
	CCA	9	1.06	35	1.02	0.04	Arg	AGC***	11	0.91	15	0.26	0.65
	CCG*	5	0.72	20	0.54	0.18		<u>AGA*</u>	29	2.02	65	1.74	0.28
Thr	ACU	16	1.21	38	1.15	0.06	Gly	AGG	4	0.28	35	0.91	-0.63
	ACC***	13	1.46	25	0.88	0.58		<u>GGU***</u>	33	1.51	42	0.99	0.52
	<u>ACA*</u>	10	1.33	42	1.19	0.14		GGC	5	0.23	15	0.49	-0.26
	ACG	0	0.00	20	0.78	-0.78		GGA	20	1.52	75	1.91	-0.39
Ala	<u>GCU***</u>	39	2.68	46	1.55	1.13	Ter	GGG*	7	0.75	34	0.61	0.14
	GCC	2	0.05	21	0.88	-0.83		UAA	3	1.80	3	1.80	0
	GCA	15	1.12	32	1.09	0.03		UAG	1	0.60	1	0.60	0
	GCG	3	0.16	13	0.47	-0.31		UGA	1	0.60	1	0.60	0

注：*： Δ RSCU ≥ 0.08 ；**： Δ RSCU ≥ 0.3 ；***： Δ RSCU ≥ 0.5 ，加下划线为最优密码。

在密码子偏好性方面,发现锡金海棠叶绿体基因组密码子第 3 位碱基的 GC3 含量为 28.30%,远低于 GC1 和 GC2,即 GC1(46.98%)>GC2(39.46%)>GC3(28.30%),这与西藏凹乳芹(*Vicatia thibetica* de Boiss)^[22]、滇楸(*Catalpa fargesii* Bur.f.*duclouxii*)^[23]植物一致,说明不同物种密码子的使用特征具有相似性。中性绘图分析显示,GC12 与 GC3 未有明显相关性,说明自然选择对其偏性影响更大,这与大山樱(*Prunus sargentii*)^[24]所得结果一致。在 PR2-plot 分析中发现碱基使用频率 T>A、G>C,这表明基因在受到选择影响的同时还受到其他因素作用,这与杧果(*Mangifera indica*)^[25]叶绿体基因组碱基使用频率一致。在 ENC-plot 分析结果中显示,有

30个基因的ENC频数比值在-0.05~0.05区间之外,说明锡金海棠的密码子偏好性的产生更受由自然选择影响。结合锡金海棠叶绿体基因组最优密码子分析,最终确定UAU、UUG、CUU、CAA、AUU、GUU、GAA、UCU、CGU、CGA、AGA、GGU、ACA、GCU共14个密码子为锡金海棠叶绿体基因组的最优密码子。

本研究发现其叶绿体基因组结构与大多数被子植物的结构一致,在对密码子偏好性进行分析后发现,锡金海棠密码子偏好性的产生主要受自然选择影响,同时也筛选出锡金海棠叶绿体最优密码子。叶绿体基因组数据对今后研究锡金海棠亲缘关系、系统发育、物种鉴定提供了数据支持,此后可通过结合属内其它物种叶绿体基因组数据构建系统发育树,分析锡金海棠在属内种间的亲缘关系与系统发育关系。最优密码子的确定也为优化目标基因、种源鉴定以及异源表达基因改造提供了理论依据。

参 考 文 献

- [1] 张俊焱,曾阳,李锦萍,等.2种药用樟牙菜叶绿体基因组密码子偏好性分析[J].中国中医药信息杂志,2022,29(1):96-102.
- [2] 王武超.中国蕨属叶绿体基因组比较分析及种间亲缘关系的研究[D].北京:中国科学院大学(中国科学院武汉植物园),2021.
- [3] 王帅.中国苹果属植物系统发育基因组学研究[D].郑州:河南农业大学,2021.
- [4] 敖干.中国蕨麻属(*Argentina*)植物叶绿体基因组研究[D].内蒙古:内蒙古师范大学,2021.
- [5] 刘翔.基于叶绿体基因组和 rDNA 序列的淫羊藿属系统进化研究[D].北京:北京协和医学院,2021.
- [6] NIE X J, DENG P C, FENG K W. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family[J]. Plant Molecular Biology Reporter, 2014, 32(4): 828-840.
- [7] ROMERO H, ZAVALA A, MUSTO 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] 陆奇丰,骆文华,黄至欢.两种梧桐叶绿体基因组密码子使用偏性分析[J].广西植物,2020,40(2):173-183.
- [9] ZELASKO S, PALARIA A, DAS A. Optimizations to achieve high-level expression of cytochrome P450 proteins using *Escherichia coli* expression systems[J]. Protein Expression and Purification, 2013, 92(1): 77-87.
- [10] MORTON B. R. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA[J]. J Mol Evol, 2003, 56(5): 616-629.
- [11] 李江飞,王瑜,颜廷雨,等.云南油杉叶绿体基因组密码子偏好性分析[J].中南林业科技大学学报,2022,42(4):30-39.
- [12] 蔡元保,杨祥燕.澳洲坚果壳种叶绿体基因组的密码子使用偏好性及其影响因素分析[J].植物科学学报,2022,40(2):229-239.
- [13] 杨祥燕,蔡元保,谭秦亮,等.菠萝叶绿体基因组密码子偏好性分析[J].热带作物学报,2022,43(3):439-446.
- [14] 毛立彦,黄秋伟,龙凌云,等.7种睡莲属植物叶绿体基因组密码子偏好性分析[J].西北林学院学报,2022,37(2):98-107.
- [15] 王大江,肖艳宏,高源,等.我国苹果属植物野生资源收集、保存和利用研究现状[J].中国果树,2021,2021(10):6-11.
- [16] 梁国鲁.锡金海棠 Giemsa C 带诱导及带型分析[J].西南农业大学学报,1997(2):95-97.
- [17] 华利源,王轩,钱关泽.苹果属植物的核型分析[J].聊城大学学报(自然科学版),2019,32(6):97-110.
- [18] DIERCKXSENS N, MARDULYN P, SMITS G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data[J]. Nucleic Acids Research, 2017, 45(4): e18.
- [19] KEARSE M, MOIR R, WILSON A. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data[J]. Bioinformatics, 2012, 28(12): 1647-1649.
- [20] 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.
- [21] 杨国锋,苏昆龙,赵怡然,等.蕹菜苜蓿叶绿体密码子偏好性分析[J].草业学报,2015,24(12):171-179.
- [22] 关云会,姚文燕,杨青淑,等.西藏凹乳芹叶绿体基因组密码子偏好性分析[J].分子植物育种,2021,19(23):7771-7781.
- [23] 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.
- [24] 李蒙,叶琦,宋炎峰,等.大山樱叶绿体基因组特征及密码子偏好性分析[J/OL].分子植物育种,2022-09-15.
- [25] 辛雅萱,黎若竹,李鑫,等.杧果叶绿体基因组密码子使用偏好性分析[J].中南林业科技大学学报,2021,41(9):148-156+165.

(责任编辑:刘庆松)