



油点草叶绿体基因组密码子使用偏好性分析

李红,周梦丽,王霞,郭蕾,王瑞红

(浙江理工大学生命科学与医药学院,杭州 310018)

摘要: 为分析油点草(*Tricyrtis macropoda*)叶绿体基因组的密码子使用偏好性,探究其密码子使用模式及影响因素,运用CodonW1.4.2、CUSP和SPSS20.0软件计算油点草叶绿体基因组密码子的相关参数并绘制相应的统计图。结果表明:油点草叶绿体中53个蛋白质编码序列(Coding sequence, CDS)的3个密码子位点GC平均值由大到小分别为46.63%(GC1)、39.74%(GC2)和29.05%(GC3),有效密码子数目(Effective number of codon, ENC)平均值为46.96,远大于35.00;密码子适应指数(Codon adaptation index, CAI)在0.11~0.31之间,密码子偏好性受到多种因素的影响,但自然选择是主要驱动力;CUU、UAU、CGA等7个密码子被确定为最优密码子,且都偏好以A或U结尾。该论文结果可为油点草属系统发育关系和叶绿体基因组中密码子的演化研究提供科学依据。

关键词: 油点草;叶绿体基因组;密码子偏好性;最优密码子;自然选择

中图分类号: Q943.2

文献标志码: A

文章编号: 1673-3851(2025)05-0377-08

引文格式: 李红,周梦丽,王霞,等. 油点草叶绿体基因组密码子使用偏好性分析[J]. 浙江理工大学学报(自然科学), 2025, 53(3): 377-384.

Reference Format: LI Hong, ZHOU Mengli, WANG Xia, et al. Analysis of codon usage bias in the chloroplast genome of *Tricyrtis macropoda* [J]. Journal of Zhejiang Sci-Tech University, 2025, 53(3): 377-384.

Analysis of codon usage bias in the chloroplast genome of *Tricyrtis macropoda*

LI Hong, ZHOU Mengli, WANG Xia, GUO Lei, WANG Ruihong

(College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: To deeply analyze the codon usage bias in the chloroplast genome of *Tricyrtis macropoda* and explore its codon usage patterns and influencing factors, CodonW1.4.2, CUSP and SPSS20.0 were used to calculate the relevant parameters and produce statistical graphs. The results showed that the average GC content values at the three codon sites of 53 protein-coding sequences (CDS) were 46.63% (GC1), 39.74% (GC2) and 29.05% (GC3), respectively. The average number of the effective number of codon (ENC) was 46.96, much higher than 35.00, while the codon adaptation index (CAI) ranged between 0.11 and 0.31. Codon usage bias was influenced by a variety of factors, but natural selection was the main driver. A total of seven codons including CUU, UAU and CGA were identified as the optimal codons, all of which showed a preference for A or U at the third position. The results provide a scientific basis for understanding the phylogenetic relationship of *Tricyrtis macropoda* and the evolution of codons in the chloroplast genome.

Key words: *Tricyrtis macropoda*; chloroplast genome; codon usage bias; optimal codon; natural selection

收稿日期: 2024-08-14 网络出版日期: 2024-11-04

基金项目: 浙江省基础公益研究计划(LGN22H280005); 浙江理工大学基本科研业务费(24042128-Y); 国家自然科学基金项目(31600183); 浙江省教育厅一般科研项目(Y202353132)

作者简介: 李红(2001—), 女, 湖南张家界人, 硕士研究生, 主要从事植物系统进化与基因组学方面的研究。

通信作者: 王瑞红, Email: ruihong8977@zstu.edu.cn

0 引言

密码子在核酸和蛋白质之间的生物体遗传信息传递过程中起着至关重要的作用,大多数氨基酸都由至少 2 个同义密码子编码^[1]。密码子使用偏好(Codon usage bias, CUB)是指同义密码子的优先和非随机应用^[2]。受到突变、自然选择、基因组大小、基因中密码子位置、mRNA 折叠、蛋白质结构以及 tRNA 丰度等多种因素的影响,密码子偏好性不仅在物种、科或界内群体之间存在差异,而且在生物体内的基因之间也存在差异^[3-4]。密码子使用偏好性对基因的转录、蛋白质的翻译折叠等生命过程都会造成影响,可以反映物种或基因的起源和对环境的适应性^[5]。密码子使用偏好分析不仅有助于理解进化过程中基因的表达规律和物种对环境的适应,而且在基因工程的载体选择和基因的表达优化及优良品种选育等方面具有重要的指导意义^[6-8]。

叶绿体将光能转化为化学能,是植物光合作用不可或缺的细胞器,含有自身的一套基因组,以母系遗传方式进行遗传信息传递^[9];具有体积小、遗传相对保守且基因组结构简单等特点,已经被证实可以帮助解决物种进化和系统发育等问题^[10]。近年来,高通量测序技术在生物技术领域得到广泛应用,植物叶绿体基因组被测序,极大程度丰富了叶绿体基因组数据库,目前已有超过 2000 多种植物的叶绿体基因组被发布^[11-12]。

油点草(*Tricyrtis macropoda*),又名紫海葱,是一种百合科油点草属的多年生草本植物。油点草主要分布在中国华北和秦岭以南各省市,常生长在海拔 800 m 左右的林下、石头裂缝和山坡沟边等阴凉的环境下^[13]。花被片披针形,内有多个紫红色斑点,外轮花被开花后前端向内卷,内轮花被狭条形,因其花型精致迷人而具有很高的园艺观赏价值,已培育出许多园艺品种^[14-15]。油点草含有黄酮类、酚酸类、甾体类、菇类及其苷类、鞋质、有机酸等成分^[16-17],具有良好的止咳、补虚、消积、延长凝血时间以及抗血栓形成等多种功效^[18-19]。

目前,关于油点草叶绿体基因组密码子使用偏好性和影响偏好性形成的主要驱动力的系统研究尚未见报道。本文对油点草叶绿体基因组密码子使用偏好进行分析,并探讨了其影响因素,确定了油点草叶绿体基因组的最佳密码子,以期为油

点草系统进化以及资源开发利用等研究提供理论依据。

1 材料与方法

1.1 实验材料

利用课题组此前测序组装并注释的油点草叶绿体基因组(GenBank: MG599475)进行分析,序列长度为 155778 bp,GC 含量为 37.4%,蛋白质编码序列(Coding sequence, CDS)84 条^[20]。为了减少误差,剔除长度小于 300 bp 且存在终止密码子和起始密码子异常的重复基因序列,最终得到可以用于分析密码子偏性的蛋白质编码序列 53 条。

1.2 方法

1.2.1 密码子组成分析

利用 CodonW1.4.2 软件计算油点草叶绿体基因组蛋白质编码序列的有效密码子数(Effective number of codon, ENC)、密码子适应指数(Codon adaptation index, CAI)以及同义密码子的相对使用频率(Relative synonymous codon usage, RSCU)等;并用 CUSP 在线软件和 Excel 2019 分析统计每个蛋白质编码序列在第三位碱基组成上的 A、T、G、C 含量,即 A3、T3、G3、C3,以及密码子第一个碱基(GC1)、第二个碱基(GC2)、第三个碱基(GC3)的 GC 含量,总 GC 含量(GCAll)以及同义密码子第三位的 G+C 含量(GC3_s)。

1.2.2 中性绘图分析

以 GC12(GC1 和 GC2 的平均值)作为纵坐标,以 GC3 作为横坐标,绘制相应散点图。当回归曲线的斜率接近 1 时,说明该基因受突变压力的影响较大;当斜率接近 0 时,选择压力则被认为是造成密码子使用偏性的主要驱动力^[21]。

1.2.3 ENC-plot 分析

ENC 是基因中密码子偏好程度的量度,ENC 独立于基因长度和氨基酸组成^[22]。ENC 值介于 20~61 之间,小于等于 35 代表基因有着显著的密码子偏差,ENC 值越高则代表密码子偏性越弱^[23]。以 ENC 值作为纵坐标,以第三位碱基 GC 含量(GC3)作为横坐标,制作散点图,并在图中绘制 ENC 值期望曲线($n_{ENC} = 2 + n_{GC3} + 29/[n_{GC32} + (1 - n_{GC3})^2]$)。当实际 ENC 值与期望值相接近时,说明突变是影响密码子偏好的主要因素;当实际 ENC 值与期望值相差较远时,则说明自然选择是影响该基因密码子偏性的主要因素^[24]。

1.2.4 PR2-plot 分析

PR2-plot 分析用来指示选择和突变压力对密码子偏移的影响,以 A3/(A3+T3)作为纵坐标,以 G3/(G3+C3)作为横坐标,绘制二维散点矩阵图。图中的十字交叉中心点(A=T 且 G=C),表示偏好性完全受突变影响,其余点与中心点之间的矢量距离表示其位置基因的偏移方向和程度^[25]。

1.2.5 最优密码子分析

将油点草 53 条蛋白编码基因的有效密码子数值进行排序,选择前 10%的基因构建高表达基因库,选择后 10%的基因构建低表达基因库,计算出两库的同义密码子相对使用度 RSCU 和 ΔRSCU 值。RSCU 大于 1 则表示该密码子为高频密码子,将 ΔRSCU 大于或等于 0.08 的密码子定义为高表达密码子,同时满足这 2 个条件的密码子就称其为最优密码子^[26]。

2 结果与分析

2.1 密码子组成分析

对油点草叶绿体基因组编码序列进行碱基分析,结果如表 1 所示。表 1 显示:GCall 的平均含量为 38.47%,GC1、GC2、GC3 的平均值分别为 46.63%、39.74%、29.05%,表明油点草叶绿体基因组的 GC 含量在密码子三个位置的分布存在差异,GC1 与 GC2 的含量相差不大,GC3 的平均值最小,表明油点草叶绿体基因组的密码子第三位碱基组成多为 A 和 T;油点草 ENC 在 38.26~59.23 之间,除了 *petD*、*ndhG* 基因外,其余基因的 ENC 值都大于 40;CAI 的理论范围在 0~1 之间,其值越小表明密码子使用偏向性越弱^[27],油点草 CAI 在 0.11~0.31 之间,平均值为 0.17。以上结果表明,油点草叶绿体基因组不仅密码子偏好性较弱,基因表达水平也较低。

表 1 油点草叶绿体基因组密码子碱基相关参数

基因	GC1 /%	GC2 /%	GC3 /%	GC3s /%	GCall /%	ENC	CAI	基因	GC1 /%	GC2 /%	GC3 /%	GC3s /%	GCall /%	ENC	CAI
<i>rps14</i>	45.54	45.54	30.69	31.60	40.59	41.77	0.14	<i>ndhK</i>	43.41	43.02	27.13	27.50	37.86	47.57	0.16
<i>ndhE</i>	40.20	31.37	26.47	27.30	32.68	45.91	0.15	<i>rpl2</i>	51.10	49.26	30.88	31.60	43.75	49.59	0.14
<i>rps18</i>	36.27	41.18	24.51	24.70	33.99	42.68	0.11	<i>petA</i>	53.89	35.83	30.22	29.90	39.98	49.46	0.16
<i>rpl22</i>	42.86	40.18	24.11	25.20	35.71	46.09	0.18	<i>ccsA</i>	32.92	38.20	26.71	27.00	32.61	44.67	0.14
<i>ndhC</i>	49.59	34.71	28.93	28.30	37.74	47.06	0.19	<i>rpoA</i>	44.71	34.41	27.65	26.80	35.59	48.22	0.16
<i>rpl20</i>	37.19	42.15	29.75	30.60	36.36	51.84	0.11	<i>psbA</i>	48.87	43.79	33.05	33.10	41.90	40.94	0.31
<i>rpl14</i>	55.28	37.40	23.58	23.50	38.75	43.45	0.16	<i>psbD</i>	51.98	43.50	35.03	35.00	43.50	46.71	0.25
<i>rps8</i>	42.11	40.60	26.32	26.60	36.34	44.98	0.13	<i>ndhA</i>	42.58	39.29	24.45	24.40	35.44	42.82	0.13
<i>rps12</i>	33.58	47.76	43.28	47.00	41.54	59.23	0.15	<i>ndhH</i>	51.02	36.29	28.17	28.60	38.49	48.02	0.15
<i>atpE</i>	50.37	41.48	31.11	31.00	40.99	47.85	0.17	<i>psbC</i>	53.25	46.33	35.22	34.90	44.93	43.60	0.18
<i>rps11</i>	56.83	54.68	22.30	22.60	44.60	41.75	0.15	<i>rbcL</i>	57.17	43.24	31.56	31.50	43.99	48.67	0.27
<i>rpl16</i>	52.14	52.86	31.43	32.80	45.48	45.36	0.15	<i>accD</i>	38.06	37.65	25.51	25.50	33.74	43.60	0.19
<i>rps7</i>	51.92	45.51	21.79	22.40	39.74	44.74	0.18	<i>atpB</i>	56.31	41.48	30.66	31.20	42.82	47.22	0.19
<i>ndhJ</i>	47.17	37.74	29.56	29.90	38.16	47.39	0.16	<i>ndhD</i>	40.24	37.25	30.28	30.40	35.92	48.91	0.13
<i>ycf3</i>	47.37	39.18	33.33	33.50	39.96	58.03	0.17	<i>atpA</i>	56.30	39.57	25.20	25.80	40.35	43.74	0.21
<i>petD</i>	49.43	38.64	25.00	24.70	37.69	38.26	0.16	<i>psbB</i>	54.03	46.56	32.61	33.10	44.40	46.87	0.19
<i>ndhG</i>	43.50	35.59	20.90	21.10	33.33	39.93	0.14	<i>matK</i>	42.00	31.60	26.59	26.70	33.40	47.13	0.17
<i>ndhI</i>	40.56	37.22	25.00	24.70	34.26	48.04	0.19	<i>ndhB</i>	42.75	40.27	30.92	31.20	37.98	46.07	0.16
<i>ycf4</i>	47.57	41.62	31.35	31.80	40.18	44.83	0.17	<i>rpoC1</i>	50.37	38.51	28.40	29.00	39.09	47.65	0.15
<i>atpF</i>	47.37	35.79	29.47	29.00	37.54	44.06	0.14	<i>psaB</i>	48.44	42.99	34.69	35.00	42.04	49.10	0.18
<i>rps4</i>	47.03	38.61	26.73	27.20	37.46	46.38	0.15	<i>ndhF</i>	36.47	37.95	24.23	24.50	32.88	45.27	0.14
<i>clpP</i>	58.05	35.61	32.20	31.70	41.95	56.19	0.18	<i>psaA</i>	52.20	43.41	33.56	33.60	43.05	48.66	0.19
<i>rps3</i>	45.66	33.79	25.57	26.20	35.01	45.43	0.15	<i>rpoB</i>	50.33	37.98	28.34	28.30	38.88	48.06	0.15
<i>rps2</i>	43.75	42.41	29.91	30.70	38.69	51.81	0.20	<i>rpoC2</i>	45.83	37.43	28.81	28.80	37.36	47.86	0.15
<i>petB</i>	48.26	41.74	33.48	33.20	41.16	48.78	0.19	<i>ycf1</i>	35.82	29.59	25.37	25.60	30.26	46.55	0.18
<i>cemA</i>	41.56	25.97	32.90	33.20	33.48	50.96	0.19	<i>ycf2</i>	41.20	34.88	36.87	37.10	37.65	52.04	0.16
<i>atpI</i>	49.19	36.69	27.82	27.80	37.90	47.17	0.17	平均值	46.63	39.74	29.05	29.33	38.47	49.96	0.17

表 2 为油点草叶绿体基因组碱基参数相关性分析结果,GC1 与 GC2、GC3、GCall 和 CAI 均呈正相关,其中 GC1 与 GC2 呈显著相关,相关系数为 0.338;GC3 与 GC1、GC2 之间相关性不显著,表明

第一位碱基与第二位碱基组成相似,但与第三位碱基之间有较大差异;ENC 与 GC1、GC2 之间呈负相关,与 GC3 呈极显著正相关,表明第三位碱基对密码子偏好性的影响最大。

表 2 油点草叶绿体基因组密码子中碱基参数相关性

指标	GC1	GC2	GC3	GC3 _s	GCall	ENC	CAI
GC2	0.338*						
GC3	0.085	0.230					
GC3 _s	0.038	0.272*	0.990**				
GCall	0.765**	0.757**	0.535**	0.525**			
ENC	−0.079	−0.082	0.620**	0.631**	0.150		
CAI	0.419**	0.092	0.299*	0.257	0.396**	0.005	

注: * 表示 $P<0.05$, ** 表示 $P<0.01$ 。

油点草叶绿体基因组整体同义密码子相对使用度分析结果如表 3 所示。由表 3 可知:油点草编码序列包含 64 种密码子,其中,编码 Ile 的密码子 AUU 数量最多,有 867 个,编码 Ter 的密码子 UAG 数量最少,为 15 个;油点草叶绿体基因组 RSCU>1 的密码子有 30 个,说明其使用频率较高,除了编码 Leu 的 UUG 密码子以 G 结尾,其余的 29 个均已 A 或 U 结尾,油点草叶绿体基因组偏好以 A 或 U 结尾。

2.2 中性绘图分析

油点草叶绿体基因组中性绘图分析结果如图 1 所示。图 1 表明:GC3 取值在 0.209~0.4328 区间,GC12 取值在 0.3271~0.5576 区间,除了一个基因其余都在对角线上方; $R^2=0.041$, $r=0.202$,GC12 与 GC3 组成差异较大;曲线斜率为 0.23,表明油点草进化中所受到的突变压力占比为 23%,自然选择占比为 77%,油点草叶绿体基因组密码子偏性主要受到自然选择的影响。

2.3 ENC-plot 分析

油点草叶绿体基因组 ENC-plot 分析结果如图 2 所示,ENC 的取值范围在 38.26~59.23 之间,6%的基因分布于曲线上方或沿标准曲线分布,94%的基因偏离标准曲线分布于曲线下方。

通过对 ENC 比值($ENC_{exp}-ENC_{obs}$)/ ENC_{exp} 的计算,可更准确地比较 ENC 实际值与 ENC 预期值之间的差异,结果表 4 所示。表 4 显示:分布在 −0.05~0.05 区间的有 11 个,即有 21%基因的 ENC 实际值与期望值相接近;而分布在 −0.05~0.05 区间之外的有 42 个,即有 79%基因的 ENC 实际值与期望值相差较大,进一步验证了自然选择对油点草叶绿体基因组密码子偏好性的影响程度更大。

表 3 油点草叶绿体基因组同义密码子相对使用度

氨基酸	密码子	数目	RSCU	氨基酸	密码子	数目	RSCU
Phe	UUU	769	1.31	Ala	GCU	521	1.79
	UUC	404	0.69		GCC	182	0.62
Leu	UUA	722	2.00		GCA	332	1.14
	UUG	434	1.20		GCG	131	0.45
	CUU	455	1.26	His	CAU	408	1.58
	CUC	144	0.40		CAC	109	0.42
	CUA	274	0.76	Arg	CGU	286	1.92
	CUG	135	0.37		CGC	88	0.53
Ile	AUU	867	1.44		CGA	268	1.92
	AUC	368	0.61		CGG	98	0.53
	AUA	573	0.95		AGA	401	1.92
	AUG	492	1.00		AGG	110	0.53
Met	AUG	492	1.00	Asn	AAU	774	1.57
Val	GUU	433	1.48		AAC	212	0.43
	GUC	144	0.49	Lys	AAA	781	1.48
	GUA	445	1.52		AAG	277	0.52
	GUG	149	0.51	Asp	GAU	679	1.59
Ser	UCU	451	1.65		GAC	176	0.41
	UCC	271	0.99	Glu	GAA	858	1.51
	UCA	331	1.21		GAG	281	0.49
	UCG	146	0.53	Cys	UGU	177	1.53
	AGU	345	1.26		UGC	54	0.47
	AGC	94	0.34	Trp	UGG	380	1.00
Pro	CCU	332	1.53		Gly	GGU	472
	CCC	189	0.87	GGC		165	0.45
	CCA	248	1.14		GGA	563	1.53
	CCG	98	0.45		GGG	269	0.73
Thr	ACU	430	1.59	Gln	CAA	546	1.49
	ACC	196	0.73		CAG	187	0.51
	ACA	335	1.24	TER	UAA	28	1.40
	ACG	119	0.44		UAG	15	0.75
Tyr	UAU	649	1.61		UGA	17	0.85
	UAC	158	0.39				

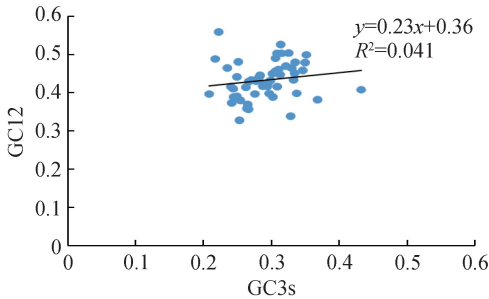


图 1 油点草叶绿体基因组中性绘图

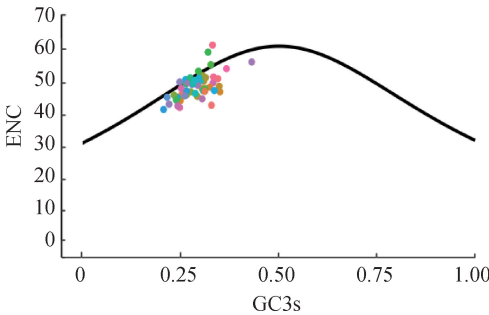


图 2 油点草叶绿体基因组 ENC-plot

表 4 油点草叶绿体基因组 ENC 比值频数分布

组限	组中值	频数/个	频率
−0.15~(−0.05)	−0.1	1	2
−0.05~0.05	0.0	11	21
0.05~0.15	0.1	35	66
0.15~0.25	0.2	6	11
0.25~0.35	0.3	0	0
合计		53	100

2.4 PR2-plot 分析

油点草叶绿体基因组 PR2-plot 分析结果如图 3 显示,从图中可以看出:基因并没有分布在原点上,且在 4 个象限不均衡分布;62%的油点草叶绿体基因位于平面的下方,是在左下角,显示第三位碱基 T 的使用频率高于 A、C 的使用频率高于 G,油点草叶绿体基因组密码子的第三位碱基在选择上有一定偏好性,其偏好性受到选择压力和其他因素的共同影响,但自然选择为主要影响因素。

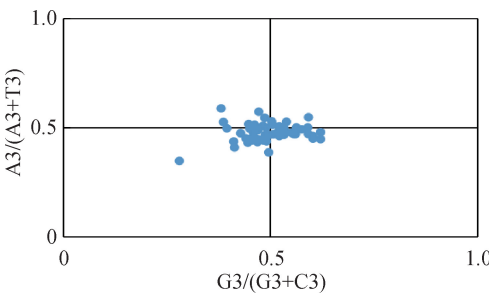


图 3 油点草叶绿体基因组 PR2-plot

2.5 最优密码子分析

最优密码子分析结果如表 5 所示,ΔRSCU≥0.08 的密码子有 CUU、CUC、CUG 等共 28 个(43%);最终确定了 CUU、UAU、UAA、AAU、CGA、AGA 和 GGA 等 7 个密码子作为最优密码子,并且全部都由 A 或 U 结尾。

表 5 油点草叶绿体基因组最优密码子

氨基酸	密码子	高表达基因		低表达基因		ΔRSCU
		次数	RSCU	次数	RSCU	
Phe	UUU	89	1.01	27	1.06	−0.05
	UUC	87	0.99	24	0.94	0.05
Leu	UUA	53	1.14	32	1.90	−0.76
	UUG	68	1.46	25	1.49	−0.03
	CUU	75	1.61	24	1.43	0.18
	CUC	24	0.51	3	0.18	0.33
	CUA	36	0.77	14	0.83	−0.06
	CUG	24	0.51	3	0.18	0.33
Ile	AUU	102	1.32	43	1.57	−0.25
	AUC	55	0.71	17	0.62	0.09
	AUA	74	0.96	22	0.80	0.16
Met	AUG	63	1.00	25	1.00	0.00
Val	GUU	43	1.58	29	1.63	−0.05
	GUC	15	0.55	5	0.28	0.27
	GUA	31	1.14	34	1.92	−0.78
	GUG	20	0.73	3	0.17	0.56
Ser	UCU	71	1.54	20	1.69	−0.15
	UCC	58	1.26	11	0.93	0.33
	UCA	59	1.28	17	1.44	−0.16
	UCG	31	0.67	1	0.08	0.59
	AGU	42	0.91	17	1.44	−0.53
	AGC	15	0.33	5	0.42	−0.09
Pro	CCU	35	1.23	22	1.73	−0.50
	CCC	25	0.88	8	0.63	0.25
	CCA	37	1.30	19	1.49	−0.19
	CCG	17	0.60	2	0.16	0.44
Thr	ACU	43	1.35	22	1.69	−0.34
	ACC	21	0.66	9	0.69	−0.03
	ACA	36	1.13	19	1.46	−0.33
	ACG	27	0.85	2	0.15	0.70
Ala	GCU	41	1.74	36	1.97	−0.23
	GCC	14	0.60	7	0.38	0.22
	GCA	23	0.98	26	1.42	−0.44
	GCG	16	0.68	4	0.22	0.46
Tyr	UAU	85	1.52	19	1.41	0.11
	UAC	27	0.48	8	0.59	−0.11
TER	UAA	9	2.25	3	1.80	0.45
	UAG	2	0.50	2	1.20	−0.70
	UGA	1	0.25	0	0.00	0.25

续表5

氨基酸	密码子	高表达基因		低表达基因		Δ RSCU
		次数	RSCU	次数	RSCU	
His	CAU	54	1.44	16	1.52	-0.08
	CAC	21	0.56	5	0.48	0.08
Gln	CAA	69	1.33	15	1.30	0.03
	CAG	35	0.67	8	0.70	-0.03
Asn	AAU	129	1.50	29	1.29	0.21
	AAC	43	0.50	16	0.71	-0.21
Lys	AAA	104	1.21	22	1.57	-0.36
	AAG	68	0.79	6	0.43	0.36
Asp	GAU	125	1.66	21	1.68	-0.02
	GAC	26	0.34	4	0.32	0.02
Glu	GAA	106	1.30	25	1.43	-0.13
	GAG	57	0.70	10	0.57	0.13
Cys	UGU	21	1.31	7	1.75	-0.44
	UGC	11	0.69	1	0.25	0.44
Trp	UGG	52	1.00	19	1.00	0.00
Arg	CGU	26	0.77	25	2.73	-1.96
	CGC	13	0.38	4	0.44	-0.06
	CGA	42	1.24	9	0.98	0.26
	CGG	24	0.71	2	0.22	0.49
	AGA	70	2.07	13	1.42	0.65
Gly	AGG	28	0.83	2	0.22	0.61
	GGU	31	0.96	36	1.82	-0.86
	GGC	16	0.50	10	0.51	-0.01
	GGA	51	1.58	24	1.22	0.36
	GGG	31	0.96	9	0.46	0.50

3 讨 论

不同物种独特的密码子使用偏性是在长期进化和在适应环境的过程中形成^[28],第三位碱基的改变通常不会造成编码氨基酸的改变,其低选择压力使得 GC3 常用于密码子使用模式分析^[29-30]。在油点草叶绿体基因组中,GC1、GC2、GC3 的平均含量分别为 46.63%、39.74%、29.05%,GC3 含量明显低于 GC1 和 GC2,第三位密码子偏好以 A/T 结尾,GC1 与 GC2 呈显著正相关,GC3 与 GC1、GC2 之间相关性均不显著,与 *Bupleurum falcatum*^[31]、杜梨^[32]、西南桦^[33]等植物基因组密码子 GC 含量分布相似,这表明不同物种密码子偏好性存在相似性。PR2-plot 分析结果显示,T 和 C 的使用频率要高于 A 和 G,这与大花香水月季^[34]、陆地棉^[35]等植物的研究结果一致,而在百合^[36]和东亚仙女木^[37]中第三位碱基 G 的使用频率大于 C、T 的使用频率大于

A,这说明叶绿体基因组密码子组成在不同物种间存在差异。

自然选择和突变是影响密码子偏好性的主要因素。通过中性绘图、ENC-plot 分析以及 PR2-plot 分析,表明在油点草进化过程中,其密码子偏好性主要受到自然选择的影响,与大花香水月季^[34]、*Delphinium grandiflorum*^[38]的密码子偏好性相一致;而在西南桦^[33]、油茶^[39]等植物研究中,其偏好性主要受到突变的影响,植物叶绿体基因组密码子偏好性受到多种因素的影响且在不同物种间其影响因素有所不同。

本文中满足 $RSCU > 1$ 且 $\Delta RSCU \geq 0.08$ 的密码子共 7 个,即 CUU、UAU、UAA、AAU、CGA、AGA、GGA,被确定为油点草叶绿体基因组的最优密码子,且偏好以 A 或 U 结尾,本研究结果与百合^[36]、柔毛郁金香^[40]最优密码子分析相一致,说明亲缘关系相近的不同物种之间在进化过程中叶绿体基因组相对保守,符合“高等植物密码子倾向于 A 或 U 结尾”的假设^[41],但与先前单子叶植物常以 G 或 C 结尾,双子叶植物常以 A 或 U 结尾的研究结论相反^[42],对于造成这种进化模式的差异未来还需要进一步研究。

4 结 语

本文以油点草叶绿体基因组的 53 条蛋白编码序列作为研究对象,对密码子使用偏好性进行分析,主要结论如下:

a)油点草叶绿体基因组密码子三个位置的 GC 含量分布存在差异,第一位碱基组成与第二位碱基组成相似,与第三位碱基组成差异较大;ENC 取值范围在 30.26~59.23 之间,CAI 取值范围在 0.11~0.31 之间,表明了油点草叶绿体基因组密码子偏好性和基因表达水平较弱。

b)中性绘图、ENC-plot 和 PR2-plot 分析都表明了油点草叶绿体基因组密码子偏好性受到多种因素的影响,但是自然选择是主要影响因素。

c)满足 $RSCU > 1$ 和 $\Delta RSCU \geq 0.08$ 的密码子有七个,分别是 CUU、UAU、UAA、AAU、CGA、AGA、GGA,被确定为油点草叶绿体基因组的最优密码子,且都偏好以 A 或 U 结尾。

密码子偏好性的形成因素相对复杂,目前人们对于其认知还不足,还需要多方面深入研究。未来的研究可以聚焦于以下几个方向:a)从转录、mRNA 稳定性和翻译效率等多个层面研究密码子偏好性如

何影响基因的表达效率;b)探索密码子偏好性在植物适应不同环境压力中的作用,以及如何通过密码子优化来提高植物的抗逆性;c)利用对密码子偏好性的理解,通过基因工程手段改造外源基因,提高其在植物体内的表达量,改善植物性状等。这些方面的研究可为植物科学和农业可持续发展带来新的视角。

参考文献:

[1] Zhang Y, Shen Z, Meng X, et al. Codon usage patterns across seven Rosales species[J]. BMC Plant Biology, 2022, 22(1): 65.

[2] Chakraborty S, Yengkhom S, Uddin A. Analysis of codon usage bias of chloroplast genes in *Oryza* species: Codon usage of chloroplast genes in *Oryza* species[J]. Planta, 2020, 252(4): 67.

[3] Tyagi S, Kabade P G, Gnanapragasam N, et al. Codon usage provide insights into the adaptation of rice genes under stress condition[J]. International Journal of Molecular Sciences, 2023, 24(2): 1098.

[4] Parvathy S T, Udayasuriyan V, Bhadana V. Codon usage bias [J]. Molecular Biology Reports, 2022; 49(1): 539-565.

[5] Fages-Lartaud M, Hundvin K, Hohmann-Marriott M F. Mechanisms governing codon usage bias and the implications for protein expression in the chloroplast of *Chlamydomonas reinhardtii* [J]. The Plant Journal: for Cell and Molecular Biology, 2022, 112(4): 919-945.

[6] Bu Y, Wu X, Sun N, et al. Codon usage bias predicts the functional MYB10 gene in *Populus* [J]. Journal of Plant Physiology, 2021, 265: 153491.

[7] Yang Q, Xin C, Xiao Q S, et al. Codon usage bias in chloroplast genes implicate adaptive evolution of four *ginger* species[J]. Frontiers in Plant Science, 2023, 14: 1304264.

[8] 高守舆, 李钰莹, 杨志青, 等. 白羊草叶绿体基因组密码子使用偏好性分析[J]. 草业学报, 2023, 32(7): 85-95.

[9] Tang D, Wei F, Cai Z, et al. Analysis of codon usage bias and evolution in the chloroplast genome of *Mesona chinensis Benth* [J]. Development Genes and Evolution, 2021, 231(1/2): 1-9.

[10] Shen L, Chen S, Liang M, et al. Comparative analysis of codon usage bias in chloroplast genomes of ten medicinal species of Rutaceae[J]. BMC Plant Biology, 2024, 24(1): 424.

[11] 杨林, 谭萍, 姜月影, 等. 3 种木莲属濒危药用植物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(8): 93-99.

[12] Hu X, Li Y, Meng F, et al. Analysis of chloroplast genome characteristics and codon usage bias in 14 species of Annonaceae [J]. Functional & Integrative Genomics, 2024, 24(3): 109.

[13] 半夏. 浙西夏秋植物赏:上[J]. 花卉, 2019(23): 50-53.

[14] 任保青, 马骥, 周哲峰. 山西省百合科植物新资料[J]. 山西林

业科技, 2015, 44(2): 10-12.

[15] 周百黎. 药草花园的花园故事:十四:秋日杜鹃:油点草[J]. 园林, 2014(8): 84-85.

[16] 闫蕊. 黄花油点草主要有效成分分析研究[D]. 汉中:陕西理工大学, 2018: 3-6.

[17] Ren L, Wang Y Z, Zhang W, et al. Triculata A, a novel compound from *Tricyrtis maculata* (D. Don) J. F. Macbr. with biological properties [J]. Natural Product Research, 2021, 35(21): 3729- 3737.

[18] Yu S, Guo Q, Jia T, et al. Mechanism of action of nicotiflorin from *Tricyrtis maculata* in the treatment of acute myocardial infarction: from network pharmacology to experimental pharmacology[J]. Drug Design Development and Therapy, 2021, 15: 2179- 2191.

[19] 杨楠, 贾晓斌, 张振海, 等. 黄酮类化合物抗肿瘤活性及机制研究进展[J]. 中国中药杂志, 2015, 40(3): 373-381.

[20] Wang R H, Gao J, Wu X, et al. Characterization of the complete chloroplast genome of toad lily *Tricyrtis Macropoda* (Liliaceae)[J]. Mitochondrial DNA Part B: Resources, 2018, 3(1): 145- 146.

[21] Li Y, Hu X, Xiao M, et al. An analysis of codon utilization patterns in the chloroplast genomes of three species of *Coffea* [J]. BMC Genomic Data, 2023, 24(1): 42.

[22] Li T, Ma Z, Ding T, et al. Codon usage bias and phylogenetic analysis of chloroplast genome in 36 gracilariaceae species[J]. Functional & Integrative Genomics, 2024, 24(2): 45.

[23] 冉然, 张晓明, 宛涛, 等. 密码子偏好性的相关研究及方法分析[J]. 草原与草业, 2022, 34(3): 5-10.

[24] Zhang P, Xu W, Lu X, et al. Analysis of codon usage bias of chloroplast genomes in *Gynostemma* species[J]. Physiology and Molecular Biology of Plants: An International Journal of Functional Plant Biology, 2021, 27(12): 2727- 2737.

[25] Yang M, Liu J, Yang W, et al. Analysis of codon usage patterns in 48 *Aconitum* species[J]. BMC Genomics, 2023, 24(1): 703.

[26] 原晓龙, 李云琴, 张劲峰, 等. 降香黄檀叶绿体基因组密码子偏好性分析[J]. 广西植物, 2021, 41(4): 622- 630.

[27] Li G, Zhang L, Xue P. Codon usage pattern and genetic diversity in chloroplast genomes of *Panicum* species[J]. Gene, 2021, 802: 145866.

[28] 石元豹, 汪贵斌, 杨晓明, 等. 银杏 WRKY 家族基因密码子使用偏向性分析[J]. 分子植物育种, 2019, 17(5): 1503-1511.

[29] 吴彦庆, 赵大球, 陶俊. 芍药花色调控基因的密码子使用模式及其影响因素分析[J]. 中国农业科学, 2016, 49(12): 2368- 2378.

[30] Gao Y, Lu Y, Song Y, et al. Analysis of codon usage bias of WRKY transcription factors in *Helianthus annuus* [J]. BMC Genomic Data, 2022, 23(1): 46.

[31] Gao M, Huo X, Lu L, et al. Analysis of codon usage patterns in *Bupleurum falcatum* chloroplast genome [J]. Chinese Herbal Medicines, 2023, 15(2): 284- 290.

[32] 辛雅萱, 董章宏, 瞿绍宏, 等. 杜梨叶绿体基因组密码子偏好

性分析[J]. 河北农业大学学报, 2020, 43 (6): 51- 59.

[33] 原晓龙,李云琴,王毅,等. 西南桦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, 2020, 39 (12): 5758- 5764.

[34] 牛元, 徐琼, 王嵩德, 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.

[35] 尚明照, 刘方, 华金平, 等. 陆地棉叶绿体基因组密码子使用偏性的分析[J]. 中国农业科学, 2011, 44(2): 245-253.

[36] 代国娜, 尚明越, 王嘉乐, 等. 百合属药用植物叶绿体基因组密码子偏好性及系统发育研究[J]. 中草药, 2024, 55(11): 3835-3844.

[37] Ling L, Zhang S, Yang T. Analysis of codon usage bias in chloroplast genomes of *Dryas octopetala* var. *asiatica* (Rosaceae) [J]. Genes(basel), 2024, 15(7): 899.

[38] Duan H, Zhang Q, Wang C, et al. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints [J]. PeerJ, 2021, 9: e10787.

[39] 王鹏良, 杨利平, 吴红英, 等. 普通油茶叶绿体基因组密码子偏好性分析[J]. 广西植物, 2018, 38(2): 135-144.

[40] 秦斗文, 徐庭亮, 闫京艳, 等. 柔毛郁金香叶绿体基因组密码子偏好性分析[J]. 江苏农业科学, 2023, 51(22): 41-47.

[41] Campbell WH, Gowri G. Codon usage in higher plants, green algae, and cyanobacteria[J]. Plant Physiology, 1990; 92(1): 1-11.

[42] Wang Y Z, Jiang D C, Guo K, et al. Comparative analysis of codon usage patterns in chloroplast genomes of ten *Epimedium* species[J]. BMC Genomic Data, 2023, 24(1): 3.

(责任编辑:张会巍)