

胡晓艳, 孟文瑞, 李昊炆, 等. 太行菊叶绿体基因组密码子使用偏性分析[J]. 山西农业科学, 2025, 53(4):91-99.

HU X Y, MENG W R, LI H Y, et al. Codon usage bias in *Opisthopappus taihangensis*(Y. Ling) C. Shih chloroplast genome[J]. Journal of Shanxi Agricultural Sciences, 2025, 53(4):91-99.

doi:10.26942/j.cnki.issn.1002-2481.2025.04.11

太行菊叶绿体基因组密码子使用偏性分析

胡晓艳¹, 孟文瑞², 李昊炆², 冯宇奇², 杜淑辉²

(1. 山西农业大学 食品科学与工程学院, 山西 晋中 030801; 2. 山西农业大学 林学院, 山西 晋中 030801)

摘 要: 太行菊为我国特有物种, 分布范围狭窄, 具有重要的观赏和药用价值。为研究太行菊叶绿体基因组密码子使用偏性, 以太行菊叶绿体基因组 53 条长度大于 300 bp 的蛋白编码序列为对象, 采用 CodonW 等软件对太行菊叶绿体基因组密码子使用偏性的关键因素和最优密码子进行分析。结果表明, 太行菊叶绿体基因组密码子 3 个位点的 GC 含量分布差异较大, $GC_1(38.82\%) > GC_2(37.87\%) > GC_3(36.35\%)$, 且密码子均偏向使用 A/T 结尾。有效密码子数(ENC)平均值为 50.95, 表明太行菊叶绿体基因组密码子使用的随机性较高而偏性较弱。相关性分析结果显示, ENC 仅和 GC_3 有极显著相关性, 表明 GC_3 对密码子使用偏性的影响较大, 这与植物中非同义突变的发生概率普遍较低有关。中性绘图、ENC-plot 和 PR2-plot 分析结果表明, 影响太行菊叶绿体基因组密码子使用偏性的因素有自然选择和突变, 但是自然选择发挥主要作用, 这可能与太行菊经历的特定进化历史及采取的环境适应策略有关。根据 ENC 值建立高低表达基因库, 筛选出 UUU、UUA、CUU、AUU 等 22 个太行菊叶绿体基因组最优密码子, 且这些密码子全部以 A 或 U 结尾, 这可能与特定基因的转录活性调控机制有关。

关键词: 太行菊; 自然选择; 叶绿体基因组; 密码子偏性; 最优密码子

中图分类号: S682.1+1

文献标识码: A

文章编号: 1002-2481(2025)04-0091-09

Codon Usage Bias in *Opisthopappus taihangensis*(Y. Ling) C. Shih Chloroplast Genome

HU Xiaoyan¹, MENG Wenrui², LI Haoyang², FENG Yuqi², DU Shuhui²

(1. College of Food Science and Engineering, Shanxi Agricultural University, Jinzhong 030801, China;

2. College of Forestry, Shanxi Agricultural University, Jinzhong 030801, China)

Abstract: *Opisthopappus taihangensis*(Y. Ling) C. Shih is native to China with restricted distribution area, important ornamental and medicinal value. In order to study codon usage bias in *O. taihangensis*(Y. Ling) C. Shih chloroplast genome, in this study, 53 protein coding sequences with the length longer than 300 bp in *O. taihangensis* chloroplast genome were taken as the object, the factors influencing codon usage bias and the optimal codon in *O. taihangensis* chloroplast genome were analyzed by softwares such as Codon W. The results showed that GC content was different at three sites of codon in the *O. taihangensis* chloroplast genome with $GC_1(38.82\%) > GC_2(37.87\%) > GC_3(36.35\%)$, and all the condon preferred the ends with A/T. The average value of effective number of codon(ENC) was 50.95, indicating the codon usage randomness was strong and bias was weak in *O. taihangensis* chloroplast genome. Correlation analysis indicated that ENC had only significant correlation with GC_3 , showing that GC_3 had an important influence on codon usage bias and it was related to the generally lower probability of non synonymous mutations in plants. The results of neutral plotting, ENC-plot, and PR2-plot analysis showed that the factors influencing codon usage bias in *O. taihangensis* chloroplast genome were natural selection and mutation, but natural selection played the major role, this was attributed to the specific evolutionary history and environmental adaption strategy of *O. taihangensis*. In addition, based on ENC values, the high and low expression gene library was constructed and 22 codons including UUU, UUA, CUU and AUU etc. were defined as the optimal codons and all had the ends with A or U, which may be attributed to the specific transcription regulatory mechanism.

收稿日期: 2024-11-04

基金项目: 山西农业大学博士科研启动项目(2017YJ22)

作者简介: 胡晓艳, 实验师, 博士, 主要从事功能油料树种品质形成研究, Email: dshhxy1985@163.com

通信作者: 杜淑辉, 副教授, 博士, 主要从事功能油料树种品质形成研究, Email: agas231@126.com

Keywords: *Opisthopappus taihangensis* (Y. Ling) C. Shih; natural selection; chloroplast genome; codon usage bias; optimal codon

翻译过程中编码同一氨基酸的若干同义密码子使用频率不均等,这是普遍存在的密码子使用偏性,与植物所处的系统进化位置及其进化历史有着密不可分的关系^[1]。不同基因有密码子使用偏性的差异,起初学者们对其的研究主要集中于低等真核生物和原核生物中,认为基因突变、随机漂变、自然选择等因素是影响不同物种间密码子使用偏性的 3 个主要原因^[2]。随着研究的不断深入和技术方法的巨大进步,逐渐发现密码子使用偏性还和基因表达水平、tRNA 丰度、氨基酸序列组成等因素有关。目前,国内外学者普遍认为,自然选择和突变是影响密码子使用偏性的 2 个主要因素^[3]。

叶绿体是植物重要的半自主细胞器,在碳同化、次生代谢产物的生物合成等代谢途径中起着非常重要的作用^[4]。目前,国内外学者对叶绿体基因组的研究已涉及植物种质资源开发、基因功能调控、分子育种、遗传改良、基因表达分析等方面,而对叶绿体基因组密码子使用偏性的研究是上述研究开展的重要基础^[5-6]。近年来,植物叶绿体基因组密码子使用偏性的研究发展较为迅速,已在经济作物、药用植物、观赏植物中广泛开展^[7-9]。已有研究发现,不同植物中叶绿体基因组密码子组成、使用特点、影响因素等存在差异。叶绿体基因组密码子使用偏性的研究对经济社会发展同样起到了重要的推动作用,如高效密码子的选择与使用会增强蛋白质的翻译水平,从而使其功能更完善^[10]。因此,对不同植物类群叶绿体基因组密码子使用偏性及其影响因素进行研究可以为筛选和构建合适的基因表达载体、推动生物育种工程的开展奠定坚实的基础。

太行菊(*Opisthopappus taihangensis* (Y. Ling) C. Shih) 是菊科(Compositae)太行菊属(*Opisthopappus*)多年生草本植物,在我国仅分布于山西和河南太行山区的峭壁裂缝中,具有良好的观赏价值和逆境适应性。由于太行菊分布区狭小、生存环境独特且繁殖能力弱,现已濒临灭绝,被列为国家第二批珍稀濒危保护植物^[11]。目前,学者们对太行菊的研究主要集中在形态多样性、逆境适应性、功能成分、景观遗传学等方面。如郭彦宏等^[12]研究发现,太行菊在干旱胁迫下仍具有较高的叶绿素

含量和较强的光合作用,比其他菊属植物表现出更高的耐旱性。薛淑娟等^[13]研究发现,太行菊挥发性油主要含有醇类、萜烯类、酮类及酯类成分,具有很高的药用价值。目前,国内外对太行菊叶绿体基因组密码子使用偏性的研究还没有开展。

本研究以太行菊叶绿体基因组为对象,通过生物信息学分析明确决定其密码子使用偏性的关键因素以及最优密码子,旨在为太行菊和太行菊属植物叶绿体密码子的优化、基因功能等研究奠定坚实的科学基础。

1 材料和方法

1.1 试验材料

从 NCBI 数据库获取太行菊叶绿体基因组序列(NC_042787)^[14],并使用在线网站 OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>)绘制基因组图谱。为了保证后续分析的可靠性,从中去除长度小于 300 bp 和重复的蛋白质编码序列(Coding sequence, CDS),最终得到 53 条 CDS 序列,用于后续分析。

1.2 试验方法

1.2.1 密码子使用偏性分析 运用 CodonW 1.4.2 软件分析太行菊叶绿体基因组的 53 条 CDS 序列,得出密码子 3 个碱基位点的 GC 含量(GC₁、GC₂和 GC₃),并计算 GC₁和 GC₂的平均值 GC₁₂,GC₁、GC₂和 GC₃的平均值 GC),末位碱基鸟嘌呤(A₃)、胸腺嘧啶(T₃)、胞嘧啶(C₃)、腺嘌呤(G₃)含量,有效密码子数(ENC, Effective number of codon),相对同义密码子使用度(RSCU, Relative synonymous codon usage);再运用 SPSS 20.0 软件分析各参数间的相关性。

1.2.2 中性绘图分析 以各 CDS 序列的 GC₃、GC₁₂含量作为横纵坐标进行中性绘图分析,以判断影响密码子使用偏性的关键因素。GC₁₂和 GC₃显著相关,表明密码子 3 个位点的碱基变异模式相同,突变是密码子使用偏性的影响因素;若不相关,则表明 3 个位点的碱基变异模式差异显著,自然选择是影响密码子使用偏性的关键因素^[10]。

1.2.3 ENC-plot 绘图分析 以每个 CDS 序列的 GC₃为横坐标、ENC 值为纵坐标进行 ENC-plot 绘

图,通过散点图分析不同序列间的差异,探讨密码子使用偏性的影响因素。根据以下公式判断ENC比值在频数上的分布情况。

$$ENC_{\text{比值}} = (ENC_{\text{期望}} - ENC_{\text{实际}}) / ENC_{\text{期望}} \quad (1)$$

$$ENC_{\text{期望}} = 2 + GC_3 + \frac{29}{(GC_3^2 + (1 - GC_3)^2)} \quad (2)$$

对比ENC比值的大小和频数可进一步判断影响密码子使用偏性的因素。如果ENC比值较小且频数较高,则说明突变是影响密码子使用偏性的主要因素;反之则表明自然选择是主要因素。

1.2.4 PR2-plot分析 以 $A_3/(A_3+T_3)$ 为纵坐标、 $G_3/(G_3+C_3)$ 为横坐标进行PR2-plot绘图,分析密码子第3位碱基组成差异。散点图中心原点位置($A=T$ 、 $G=C$)代表密码子使用不具有偏性,距离中心点的距离代表使用偏性的方向和程度。

1.2.5 最优密码子分析 将53条CDS序列的ENC值从小到大排序,并在前后各选取10%的序列以构建高(ENC值较小)、低(ENC值较大)表达库。将表达库中RSCU值 >1 的密码子定义为高频

密码子, $\Delta RSCU \geq 0.08$ ($\Delta RSCU = RSCU_{\text{高表达}} - RSCU_{\text{低表达}}$) 表示高表达密码子,将同时符合RSCU值 >1 和 $\Delta RSCU \geq 0.08$ 的密码子选为最优密码子^[11]。

2 结果与分析

2.1 太行菊叶绿体基因组密码子的组成特征

使用OGDRAW网站绘制了太行菊叶绿体基因组示意图(图1),太行菊叶绿体基因组全长151 089 bp,为典型的四分体结构,共包含87条CDS序列。由表1可知, T_3 、 G_3 、 A_3 和 C_3 频率的平均值分别为39%、22%、40%和26%,表明密码子第3位碱基组成有差异。总GC含量的变化范围为30.00%~46.90%,平均为39%,且 $GC_1(38.82\%) > GC_2(37.87\%) > GC_3(36.35\%)$,表明GC含量在密码子3个位置上分布不均匀且偏向使用A/T结尾。ENC的平均值为50.95,53条太行菊叶绿体CDS序列中有48条的ENC值大于45,表明其密码子使用的随机性较高,密码子使用偏性较弱。

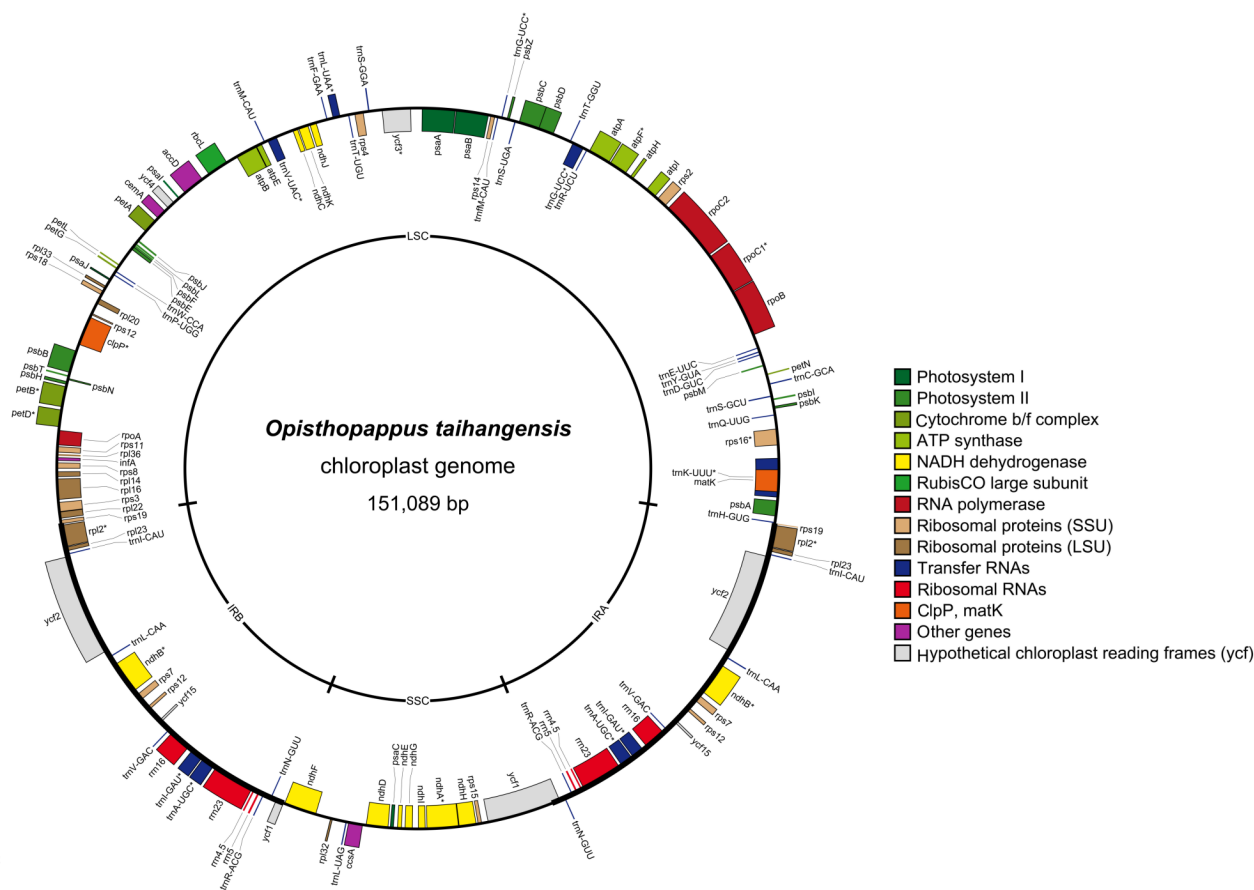


图1 太行菊叶绿体基因组

Fig. 1 The chloroplast genome of *O. taihangensis*

表 1 太行菊叶绿体基因组 CDS 序列密码子碱基组成分析

Tab.1 Analysis of basic group composition in CDS sequence codons in *O. taihangensis* chloroplast genome

基因	Gene	T ₃ /%	C ₃ /%	A ₃ /%	G ₃ /%	GC ₁ /%	GC ₂ /%	GC ₃ /%	GC ₁₂ /%	GC/%	ENC
<i>rps12</i>		19.35	33.33	38.64	37.50	49.59	30.08	54.00	39.84	46.30	52.16
<i>psbA</i>		41.96	29.37	30.21	25.71	32.86	50.14	42.00	41.50	43.80	55.37
<i>matK</i>		51.91	14.35	42.33	20.45	38.93	32.61	25.30	35.77	33.00	46.94
<i>rpoB</i>		29.89	28.44	38.65	36.95	37.08	26.89	46.90	31.99	38.90	57.18
<i>rpoC1</i>		38.25	21.49	41.70	24.40	27.87	49.78	35.60	38.83	39.60	52.92
<i>rpoC2</i>		30.62	28.08	39.44	33.71	37.35	29.34	44.80	33.35	39.10	57.52
<i>rps2</i>		47.37	14.21	47.40	16.56	43.04	41.35	23.20	42.20	37.10	46.48
<i>atpI</i>		31.77	25.52	33.96	35.86	37.65	27.94	46.80	32.80	39.90	57.90
<i>atpF</i>		41.30	20.29	45.83	23.01	33.70	46.20	32.50	39.95	39.00	42.84
<i>atpA</i>		49.38	15.21	42.66	14.58	56.39	39.49	23.50	47.94	40.40	46.13
<i>psbD</i>		41.20	28.24	27.49	25.79	30.59	52.41	42.40	41.50	43.40	57.06
<i>psbC</i>		45.68	18.52	39.27	14.89	53.38	46.20	27.70	49.79	43.70	45.18
<i>rps14</i>		33.33	34.72	43.75	25.42	49.00	36.00	43.50	42.50	44.00	50.70
<i>psaB</i>		52.08	14.26	38.11	17.21	48.16	42.99	24.50	45.58	40.20	47.14
<i>psaA</i>		45.65	17.55	39.96	17.60	52.73	43.68	28.00	48.21	42.90	49.02
<i>ycf3</i>		29.41	27.94	44.55	21.36	32.14	47.62	40.30	39.88	41.70	53.50
<i>rps4</i>		35.22	22.64	37.76	27.56	26.87	53.23	39.20	40.05	41.60	47.26
<i>ndhJ</i>		51.56	16.41	38.53	20.83	50.94	36.48	27.50	43.71	39.90	45.43
<i>ndhK</i>		45.79	16.84	46.71	13.25	43.81	44.25	24.00	44.03	38.40	51.83
<i>ndhC</i>		53.47	8.91	43.68	14.86	44.63	33.06	17.90	38.85	33.90	49.89
<i>atpE</i>		45.37	14.81	44.95	14.89	52.99	41.04	23.40	47.02	40.40	52.41
<i>atpB</i>		22.87	25.22	34.68	43.85	41.26	28.66	53.20	34.96	43.40	60.59
<i>rbcL</i>		51.79	18.88	37.43	16.32	56.79	42.59	27.70	49.69	43.40	48.13
<i>accD</i>		38.70	23.16	46.92	24.34	30.74	41.32	34.90	36.03	37.90	49.40
<i>ycf4</i>		38.24	22.79	34.15	34.55	41.30	28.80	42.30	35.05	39.30	61.00
<i>cemA</i>		36.57	25.71	40.41	25.81	28.38	29.69	38.50	29.04	34.10	51.04
<i>petA</i>		47.66	19.92	38.43	19.28	52.34	36.76	29.90	44.55	40.20	50.19
<i>rps18</i>		46.84	11.39	40.23	23.08	37.25	43.14	27.30	40.20	36.60	48.21
<i>rpl20</i>		26.37	27.47	51.72	24.32	46.03	26.98	38.40	36.51	39.10	47.57
<i>clpP</i>		25.74	29.41	30.71	49.14	37.76	28.06	56.70	32.91	42.60	45.49
<i>psbB</i>		38.19	26.97	30.32	28.98	30.71	54.33	43.40	42.52	44.70	54.12
<i>petB</i>		33.13	24.54	26.87	40.32	41.86	25.58	50.00	33.72	41.40	55.31
<i>petD</i>		47.73	31.82	20.79	17.44	25.00	51.25	40.40	38.13	40.60	41.11
<i>rpoA</i>		29.96	27.31	45.45	29.95	31.34	30.15	41.30	30.75	36.30	54.81
<i>rps11</i>		15.79	30.53	41.84	41.94	57.35	24.26	54.80	40.81	46.90	57.00
<i>rps8</i>		43.00	18.00	40.51	38.16	26.87	40.30	38.50	33.59	36.80	51.99
<i>rpl14</i>		18.52	28.40	37.04	46.67	38.52	25.41	56.30	31.97	41.30	46.98
<i>rpl16</i>		30.56	32.41	24.14	37.50	21.48	51.85	54.60	36.67	44.90	52.90
<i>rps3</i>		39.52	20.36	49.64	24.60	24.31	45.87	32.70	35.09	35.30	51.73
<i>rpl22</i>		31.13	19.81	45.87	33.33	40.65	25.81	38.50	33.23	37.10	54.59
<i>rpl2</i>		31.12	29.08	38.73	32.28	32.36	50.18	45.80	41.27	45.90	56.52
<i>ycf2</i>		41.34	24.55	42.03	19.82	36.72	42.11	33.80	39.42	38.90	51.12
<i>ndhB</i>		43.05	21.64	41.87	13.25	41.68	38.55	28.40	40.12	37.50	49.10
<i>rps7</i>		29.60	32.00	38.18	27.27	23.23	53.55	45.90	38.39	42.10	42.87
<i>ndhF</i>		54.30	11.11	41.88	15.53	36.66	35.58	19.90	36.12	32.20	44.42

续表 1 太行菊叶绿体基因组 CDS 序列密码子碱基组成分析

Tab.1(Continued) Analysis of basic group composition in CDS sequence codons in *O. taihangensis* chloroplast genome

基因	Gene	T ₃ /%	C ₃ /%	A ₃ /%	G ₃ /%	GC ₁ /%	GC ₂ /%	GC ₃ /%	GC ₁₂ /%	GC/%	ENC
	<i>ccsA</i>	44.35	18.83	43.15	25.31	37.65	24.07	31.00	30.86	33.50	50.46
	<i>ndhD</i>	40.48	21.69	39.13	29.39	37.52	28.34	37.10	32.93	37.30	54.41
	<i>ndhE</i>	55.13	26.92	30.00	21.95	17.82	41.58	34.10	29.70	32.60	50.71
	<i>ndhG</i>	46.88	14.37	40.60	17.59	45.20	35.03	24.60	40.12	35.80	54.59
	<i>ndhI</i>	36.00	23.20	33.91	36.08	36.14	25.90	43.20	31.02	36.30	56.20
	<i>ndhA</i>	48.08	11.54	47.44	9.41	44.51	39.29	17.20	41.90	34.80	42.75
	<i>ndhH</i>	48.87	12.62	48.43	15.85	50.76	36.29	21.20	43.53	37.80	48.19
	<i>ycf1</i>	36.87	23.95	52.37	21.48	27.59	24.96	31.80	26.28	30.00	51.79
平均值	Mean	39.23	22.39	39.63	25.80	38.82	37.87	36.35	38.34	39.32	50.95

太行菊叶绿体基因组碱基组成各指标相关性分析结果如表 2 所示,GC₁与 GC₂、GC₃均呈显著负相关($P<0.05$),而 GC₂与 GC₃相关性不显著;GC 和 GC₂呈显著相关($P<0.05$),GC 和 GC₃呈极显著正相关($P<0.01$),GC 与 ENC 相关性不显著。上述结果表明,太行菊叶绿体基因 GC₁和 GC₂的碱基组成较一致,而与 GC₃的碱基组成有很大差异。ENC 与 GC₃呈极显著正相关($P<0.01$),表明密码子第 3 位碱基组成对密码子使用偏性有较大影响。

表 2 太行菊叶绿体基因组 CDS 序列碱基组成参数间相关性分析

Tab.2 Correlation analysis among base composition parameters of CDS sequences in *O. taihangensis* chloroplast genome

参数	Parameter	GC ₁	GC ₂	GC ₃	GC	ENC
	GC ₁	1.000				
	GC ₂	-0.321*	1.000			
	GC ₃	-0.313*	-0.188	1.000		
	GC	0.256	0.326*	0.559**	1.000	
	ENC	-0.076	-0.299*	0.461**	0.180	1.000

注:*代表不同参数间显著相关($P<0.05$);**代表不同参数间极显著相关($P<0.01$)。

Note: * indicated significant correlation among different parameters($P<0.05$), ** indicated extremely significant correlation among different parameters($P<0.01$).

2.2 中性绘图分析

中性绘图结果如图 2 所示,GC₁₂和 GC₃之间 $R^2=0.187$ 且 $r=0.135$,表明太行菊叶绿体基因组 CDS 序列密码子 GC₁₂和 GC₃之间联系较弱且差异较明显。回归曲线的斜率为-0.226 1,表明密码子使用偏性主要受自然选择的影响。

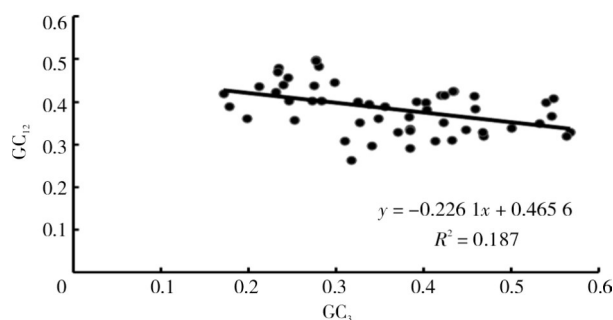


图 2 太行菊叶绿体基因组 CDS 序列密码子中性绘图分析
Fig.2 Neutrality plot analysis of CDS sequence codons in *O. taihangensis* chloroplast genome

2.3 ENC-plot 绘图分析

ENC-plot 绘图分析结果如图 3 所示,图中散点分布呈现出一定的分散性,大多数 CDS 的 ENC 值明显偏离预期。根据 ENC 比值频率分布可以发现(表 3),只有 20 个 CDS 的组限为-0.05~0.05,与 ENC 期望值相近或相同,而另外 33 个 CDS 的组限全部位于这个区间之外,与 ENC 的期望值相差较远。说明太行菊叶绿体基因组密码子使用偏性明显受到了自然选择的影响。

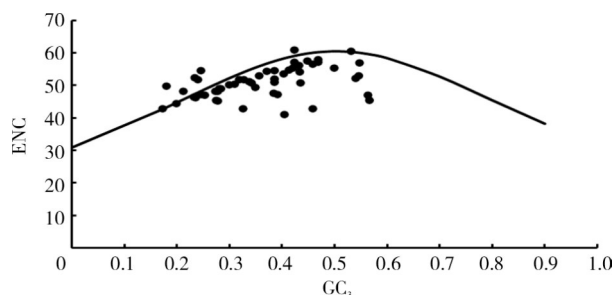


图 3 太行菊叶绿体基因组 CDS 序列 ENC-plot 绘图分析
Fig.3 ENC-plot analysis of CDS sequence in *O. taihangensis* chloroplast genome

表 3 ENC 比值频率分布

Tab.3 Frequency distribution of ENC ratio

组限 Range	组中值 Mid-value	频数 Number	频率 Frequency
-0.15~-0.05	-0.1	5	0.09
-0.05~0.05	0	20	0.38
0.05~0.15	0.1	21	0.40
0.15~0.25	0.2	5	0.09
0.25~0.35	0.3	2	0.04

2.4 PR2-plot 分析

从图 4 可以看出,PR2-plot 散点在坐标轴平面的 4 个区域中呈现出不均匀的分布状态,其中,右上方的散点数量最多,说明太行菊叶绿体基因组密码子末位碱基使用频次 $A>T$ 、 $G>C$,在密码子第 3 位碱基上存在使用偏性。结果表明,太行菊叶绿体基因组密码子使用偏性受到自然选择和突变的双重影响,但自然选择的影响更大。

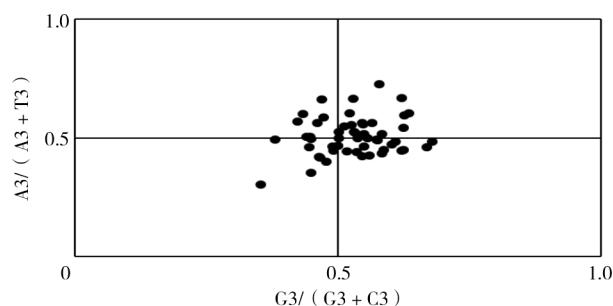


图 4 太行菊叶绿体基因组 CDS 序列 PR2-plot 绘图分析

Fig.4 PR2-plot analysis of CDS sequence in *O. taihangensis* chloroplast genome

2.5 最优密码子分析

太行菊叶绿体基因组 CDS 序列密码子 RSCU 分析结果如表 4 所示,UUU、UUA、CUU、AUU、GUU、UCU、CCU、ACU、GCU、UAU、UAA、CAU、CAA 等 22 个为最优密码子,其中,15 个以 U 结尾,7 个以 A 结尾。

表 4 太行菊叶绿体基因组最优密码子分析

Tab.4 Analysis of optimal codons for *O. taihangensis* chloroplast genome

氨基酸 Amino acid	密码子 Codon	高表达基因 High expressed gene		低表达基因 Low expressed gene		Δ RSCU	基因组 Genome	
		数目 Number	RSCU	数目 Number	RSCU		数目 Number	RSCU
Phe	UUU***	95	1.57	110	1.04	0.53	860	1.25
	UUC	26	0.43	102	0.96	-0.53	521	0.75
Leu	UUA***	80	2.65	85	1.06	1.60	613	1.62
	UUG	27	0.90	118	1.47	-0.57	479	1.26
	CUU***	49	1.62	85	1.06	0.57	450	1.19
	CUC	2	0.07	54	0.67	-0.60	207	0.55
	CUA	18	0.60	73	0.91	-0.31	315	0.83
	CUG	5	0.17	68	0.84	-0.68	211	0.56
Ile	AUU***	73	1.52	94	0.96	0.56	788	1.38
	AUC	22	0.46	86	0.88	-0.42	401	0.70
	AUA	49	1.02	113	1.16	-0.14	528	0.92
Met	AUG	35	1.00	106	1.00	0.00	430	1.00
Val	GUU***	38	2.08	51	1.10	0.99	336	1.35
	GUC	9	0.49	39	0.84	-0.35	160	0.64
	GUA	23	1.26	63	1.35	-0.09	354	1.42
	GUG	3	0.16	33	0.71	-0.55	149	0.60
Ser	UCU***	43	1.43	33	0.94	0.50	400	1.26
	UCC	26	0.87	44	1.25	-0.38	285	0.90
	UCA	33	1.10	39	1.11	-0.01	383	1.20
	UCG	18	0.60	25	0.71	-0.11	204	0.64
Pro	CCU***	27	1.96	21	0.72	1.24	247	1.33
	CCC	4	0.29	23	0.79	-0.50	122	0.66
	CCA	15	1.09	32	1.10	-0.01	213	1.14
	CCG	9	0.65	40	1.38	-0.72	163	0.88

续表4 太行菊叶绿体基因组最优密码子分析

Tab.4(Continued) Analysis of optimal codons for *O. taihangensis* chloroplast genome

氨基酸 Amino acid	密码子 Codon	高表达基因 High expressed gene		低表达基因 Low expressed gene		Δ RSCU	基因组 Genome	
		数目 Number	RSCU	数目 Number	RSCU		数目 Number	RSCU
Thr	<u>ACU</u> ***	29	1.32	14	0.55	0.77	267	1.16
	ACC	16	0.73	20	0.78	-0.06	187	0.81
	<u>ACA</u>	31	1.41	35	1.37	0.04	305	1.32
	ACG	12	0.55	33	1.29	-0.75	165	0.71
Ala	<u>GCU</u> ***	29	1.73	9	0.61	1.12	299	1.54
	GCC	9	0.54	16	1.08	-0.55	147	0.75
	<u>GCA</u>	19	1.13	16	1.08	0.05	227	1.17
	GCG	10	0.60	18	1.22	-0.62	106	0.54
Tyr	<u>UAU</u> ***	55	1.57	66	1.02	0.55	636	1.39
	UAC	15	0.43	63	0.98	-0.55	281	0.61
TER	<u>UAA</u> ***	18	1.71	70	0.77	0.94	418	1.10
	UAG	3	0.29	111	1.23	-0.94	343	0.90
His	<u>CAU</u> ***	18	1.71	42	1.11	0.61	339	1.35
	CAC	3	0.29	34	0.89	-0.61	165	0.65
Gln	<u>CAA</u> ***	31	1.59	62	1.00	0.59	482	1.35
	CAG	8	0.41	62	1.00	-0.59	232	0.65
Asn	<u>AAU</u> **	56	1.42	78	1.05	0.36	598	1.32
	AAC	23	0.58	70	0.95	-0.36	309	0.68
Lys	<u>AAA</u> ***	63	1.70	126	1.02	0.68	836	1.36
	AAG	11	0.30	121	0.98	-0.68	391	0.64
Asp	<u>GAU</u> *	36	1.53	48	1.32	0.22	453	1.48
	GAC	11	0.47	25	0.68	-0.22	158	0.52
Glu	<u>GAA</u> **	34	1.58	94	1.20	0.38	561	1.37
	GAG	9	0.42	63	0.80	-0.38	257	0.63
Cys	<u>UGU</u> *	17	1.17	25	0.93	0.25	237	1.04
	UGC	12	0.83	29	1.07	-0.25	220	0.96
TER	UGA	13	1.00	71	1.00	0.00	396	1.00
Trp	UGG	34	1.00	76	1.00	0.00	461	1.00
Arg	<u>CGU</u> ***	14	1.60	16	0.70	0.90	171	1.12
	CGC	4	0.46	14	0.61	-0.15	81	0.53
	<u>CGA</u> *	11	1.26	24	1.04	0.21	219	1.43
	CGG	6	0.69	38	1.65	-0.97	142	0.93
Ser	<u>AGU</u> ***	39	1.39	11	0.73	0.66	234	1.15
	AGC	17	0.61	19	1.27	-0.66	173	0.85
Arg	<u>AGA</u> **	28	1.33	48	1.02	0.31	413	1.28
	AGG	14	0.67	46	0.98	-0.31	231	0.72
	GGU***	36	1.22	15	0.50	0.72	291	0.99
Gly	GGC	18	0.61	19	0.63	-0.02	173	0.59
	<u>GGA</u>	36	1.22	38	1.27	-0.05	410	1.39
	<u>GGG</u>	28	0.95	48	1.60	-0.65	302	1.03

注:下划线代表RSCU>1;*代表该密码子 Δ RSCU $\geq$ 0.08,**代表 Δ RSCU $\geq$ 0.3,***代表 Δ RSCU $\geq$ 0.5;加粗标注的密码子为最优密码子。

Note: Underline indicated RSCU>1; * indicated Δ RSCU $\geq$ 0.08, ** indicated Δ RSCU $\geq$ 0.3, *** indicated Δ RSCU $\geq$ 0.5; bold indicated the optimal codons.

3 结论与讨论

本研究针对太行菊叶绿体基因组 53 条 CDS 序列,对其密码子使用模式与影响因素进行了分析。密码子碱基组成分析发现,ENC 的均值为 50.95,表明太行菊叶绿体基因组密码子使用偏性较弱,同时又解释了最优密码子数量(22 个)较多的原因,与在酸枣(*Ziziphus jujuba* var. *spinosa*)^[15]、翅果油树(*Elaeagnus mollis*)^[16]、紫苏(*Perilla frutescens*)^[17]等上的研究结果一致,较高的叶绿体密码子使用偏性可能会对叶绿体正常功能产生不良影响。太行菊叶绿体基因组碱基组成相关性分析表明,ENC 仅和 GC₃呈极显著相关,表明 GC₃对密码子使用偏性的影响较大。这与绝大多数物种叶绿体基因组密码子使用偏性的研究结果一致,密码子第 3 位碱基的变化会直接影响翻译出的氨基酸种类。因此,植物体内非同义突变的发生概率普遍较低,从而间接影响密码子使用偏性^[18-19]。

通过中性绘图、ENC-plot 和 PR2-plot 分析,发现自然选择对太行菊叶绿体基因组密码子使用有显著影响,而突变的影响较弱,这与椭圆叶花锚(*Halenia elliptica*)^[20]、白毛银露梅(*Potentilla glabra* var. *mandshurica*)^[21]和吊灯扶桑(*Hibiscus schizopetalus*)^[22]等的研究结果一致。但也有研究发现了其他影响叶绿体基因组密码子使用偏性的关键因素,如赵森等^[23]研究发现,突变是秋茄(*Kandelia candel*)叶绿体基因组密码子使用偏性的主要影响因素。从上述分析可以看出,不同物种叶绿体基因组密码子使用偏性受不同因素的影响,且不同影响因素间的差异比较明显,这可能与不同植物经历的进化历史及采取的环境适应策略不同有关。第四纪冰期-间冰期循环对现存绝大多数植物类群的系统进化产生了重要影响,不同植物采取不同的策略适应环境的剧烈变化,从而对其叶绿体基因组密码子使用偏性产生了影响^[24-25]。根据 ENC 值建立高低表达基因库,最终筛选出 UUU、UUA、CUU、AUU 等 22 个太行菊叶绿体基因组的最优密码子,它们全部以 A 或 U 结尾,表明太行菊叶绿体基因组密码子末位碱基使用存在明显偏向,符合双子叶植物密码子末位碱基偏好使用 A 和 U 的规律^[26]。虽然叶绿体基因组相对保守,但是不同物种间最优密码子仍存在一定的差异,这可能与特定基因的转录表达活性相关^[27]。

本研究通过对太行菊叶绿体基因组 53 条 CDS 序列密码子使用模式的研究,发现自然选择是影响其密码子使用偏性的关键因素,其最优密码子数量较多且全部以 A 或 U 结尾,太行菊独特的进化历史和环境适应策略可能是上述特征产生的根本原因。本研究的开展将为太行菊属物种适应性、分子系统进化以及叶绿体基因工程的开展奠定重要的理论基础。但是本研究只对太行菊叶绿体基因组密码子使用偏性进行了分析,没有涉及同属物种裂叶太行菊及菊属其他物种,无法判断其密码子使用偏性与近缘物种是否相同,同时有关太行菊叶绿体密码子使用偏性形成的分子机制及其因素还需要进一步研究。后续可针对叶绿体基因组密码子使用偏性深入开展比较分析,以明确影响不同物种密码子使用偏性的因素。

参考文献:

- [1] 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 open access[J]. Nucleic Acids Research, 2000, 28(10): 2084-2090.
- [2] 张琰, 汪成忠, 岳建华. 火炬花叶绿体基因组特征及系统发育[J]. 东北林业大学学报, 2024, 52(10): 52-58.
ZHANG Y, WANG C Z, YUE J H. Chloroplast genome features and phylogenetics of *Kniphofia uvaria* (L.) oken[J]. Journal of Northeast Forestry University, 2024, 52(10): 52-58.
- [3] 李亚麒, 黄家雄, 姜予强, 等. 小粒咖啡铁皮卡叶绿体基因组密码子偏好性分析[J]. 西北林学院学报, 2023, 38(2): 92-99.
LI Y Q, HUANG J X, LOU Y Q, et al. Codon usage bias of the chloroplast genome in *Coffea Arabica* 'typica'[J]. Journal of Northwest Forestry University, 2023, 38(2): 92-99.
- [4] DU S H, HU X Y, GUO Y T, et al. A comparative plastomic analysis of *Ziziphus jujuba* var. *spinosa* (Bunge) Hu ex H. F. Chow and implication of the origin of Chinese jujube[J]. AoB Plants, 2023, 15(2): plad006.
- [5] 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.
MAO L Y, HUANG Q W, LONG L Y, et al. Comparative analysis of codon usage bias in chloroplast genomes of seven *Nymphaea* species[J]. Journal of Northwest Forestry University, 2022, 37(2): 98-107.
- [6] 张加强, 刘慧春, 王杰, 等. 金银花大毛花叶绿体基因组密码子的偏好性分析[J]. 浙江农业学报, 2023, 35(4): 821-830.
ZHANG J Q, LIU H C, WANG J, et al. Analysis on codon usage patterns of chloroplast genome of *Lonicera japonica* cv. Damaohua[J]. Acta Agriculturae Zhejiangensis, 2023, 35(4): 821-830.
- [7] 丁祥青, 陈丝雨, 陈佳婷, 等. 11 种金茶花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), 2023, 52(4): 473-479.

- DING X Q, CHEN S Y, CHEN J T, et al. Codon bias analysis of 11 yellow *Camellia* chloroplast genome[J]. Journal of Fujian Agriculture and Forestry University(Natural Science Edition), 2023, 52(4):473-479.
- [8] SHI S L, LIU Y Q, XIA R X, et al. Comprehensive analysis of codon usage in *Quercus chloroplast* genome and focus on *psbA* gene[J]. Genes, 2022, 13(11):2156.
- [9] LING L Z, ZHANG S D, YANG T. Analysis of codon usage bias in chloroplast genomes of *Dryas octopetala* var. *asiatica* (rosaceae)[J]. Genes, 2024, 15(7):899.
- [10] 吴宪明, 吴松峰, 任大明, 等. 密码子偏性的分析方法及相关研究进展[J]. 遗传, 2007, 29(4):420-426.
- WU X M, WU S F, REN D M, et al. The analysis method and progress in the study of codon bias[J]. Hereditas, 2007, 29(4):420-426.
- [11] 张光明. 太行菊[J]. 生物学通报, 2019, 54(2):37.
- ZHANG G M. Taihang *Chrysanthemum*[J]. Bulletin of Biology, 2019, 54(2):37.
- [12] 郭彦宏, 张晶星, 杨永娟, 等. 六种野生广义菊属植物对干旱胁迫的生理响应[J]. 浙江农业学报, 2018, 30(8):1349-1354.
- GUO Y H, ZHANG J X, YANG Y J, et al. Response of six germplasms of *Chrysanthemum* and related genera to drought stress[J]. Acta Agriculturae Zhejiangensis, 2018, 30(8):1349-1354.
- [13] 薛淑娟, 杨江凯, 陈随清. GC-MS结合保留指数法解析太行菊挥发油的化学成分[J]. 中国实验方剂学杂志, 2022, 28(15):120-128.
- XUE S J, YANG J K, CHEN S Q. Analysis of chemical constitutions of volatile oil in *Opisthopappus taihangensis* by GC-MS combined with retention index[J]. Chinese Journal of Experimental Traditional Medical Formulae, 2022, 28(15):120-128.
- [14] GU H H, YANG Y, XING M H, et al. Physiological and transcriptome analyses of *Opisthopappus taihangensis* in response to drought stress[J]. Cell & Bioscience, 2019, 9(1):56.
- [15] 胡晓艳, 许艳秋, 韩有志, 等. 酸枣叶绿体基因组密码子使用偏性分析[J]. 森林与环境学报, 2019, 39(6):621-628.
- HU X Y, XU Y Q, HAN Y Z, et al. Codon usage bias analysis of the chloroplast genome of *Ziziphus jujuba* var. *spinosa* [J]. Journal of Forest and Environment, 2019, 39(6):621-628.
- [16] 郭媛婷, 张敏, 杨玉洁, 等. 翅果油树叶体基因组密码子偏好性分析[J]. 山西农业科学, 2023, 51(3):233-240.
- GUO Y T, ZHANG M, YANG Y J, et al. Analysis on codon usage bias in chloroplast genome of *Elaeagnus mollis*[J]. Journal of Shanxi Agricultural Sciences, 2023, 51(3):233-240.
- [17] 梁璇, 李鹏, 许冬梅, 等. 紫苏叶绿体全基因组特征分析[J]. 山西农业科学, 2021, 49(3):265-272.
- LIANG X, LI P, XU D M, et al. Characteristics analysis of whole chloroplast genome in *Perilla frutescen*[J]. Journal of Shanxi Agricultural Science, 2021, 49(3):265-272.
- [18] SEO T K, KISHINO H, THORNE J L. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species Divergences[J]. Molecular Biology and Evolution, 2004, 21(7):1201-1213.
- [19] VIHINEN M. Nonsynonymous synonymous variants demand for a paradigm shift in genetics[J]. Current Genomics, 2023, 24(1):18-23.
- [20] 陈雯清, 刘芝华, 汪明金, 等. 椭圆叶花锚叶绿体基因组密码子偏好性分析[J/OL]. 分子植物育种, 2024:1-16[2025-06-19]. <http://kns.cnki.net/kcms/detail/46.1068.S.20240812.1658.014.html>.
- CHEN W Q, LIU Z H, WANG M J, et al. Analysis of codon usage bias in chloroplast genome of *Halenia elliptica*[J/OL]. Molecular Plant Breeding, 2024:1-16[2025-06-19]. <http://kns.cnki.net/kcms/detail/46.1068.S.20240812.1658.014.html>.
- [21] 杨蕾, 董泽阳, 王洋, 等. 白毛银露梅叶绿体全基因组密码子偏好性分析[J]. 分子植物育种, 2022, 20(4):1095-1103.
- YANG L, DONG Z Y, WANG Y, et al. Analysis on codon usage bias of chloroplast genome in *Potentilla glabra* var. *mandshurica*[J]. Molecular Plant Breeding, 2022, 20(4):1095-1103.
- [22] 周全, 王璐瑶, 袁庆, 等. 吊灯扶桑叶绿体基因组特征与木槿属系统发育研究[J]. 草地学报, 2025, 33(3):748-765.
- ZHOU Q, WANG L Y, YUAN Q, et al. Chloroplast genomic characteristics of *Hibiscus schizopetalus* and phylogenetic relationships of *Hibiscus*[J]. Acta Agrestia Sinica, 2025, 33(3):748-765.
- [23] 赵森, 邓力华, 陈芬. 秋茄叶绿体基因组密码子使用偏好性分析[J]. 森林与环境学报, 2020, 40(5):534-541.
- ZHAO S, DENG L H, CHEN F. Codon usage bias of chloroplast genome in *Kandelia obovata*[J]. Journal of Forest and Environment, 2020, 40(5):534-541.
- [24] DU S H, WANG Z S, INGVARSSON P K, et al. Multilocus analysis of nucleotide variation and speciation in three closely related *Populus* (Salicaceae) species[J]. Molecular Ecology, 2015, 24(19):4994-5005.
- [25] 吴民华, 吴子健, 叶晓霞, 等. 鸡矢藤叶绿体基因组分[J]. 河南农业科学, 2024, 53(1):70-77.
- WU M H, WU Z J, YE X X, et al. Analysis on chloroplast genome of *Paederia scandens*[J]. Journal of Henan Agricultural Science, 2024, 53(1):70-77.
- [26] 刘德浩, 陈智涛, 王高锋, 等. 博罗红豆叶绿体基因组测序与特征分析[J]. 林业与环境科学, 2024, 40(4):105-113.
- LIU D H, CHEN Z T, WANG G F, et al. Sequencing and characterization of chloroplast genome of *Ormosia boluoensis* [J]. Forestry and Environmental Science, 2024, 40(4):105-113.
- [27] ZHANG Y R, NIE X J, JIA X O, et al. Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family [J]. Australian Journal of Botany, 2012, 60(5):461.