

万露露,范敦锦,王中煊,等. 杜鹃叶绿体基因组特征及密码子偏好性分析[J]. 江苏农业科学,2023,51(13):54-59.
doi:10.15889/j.issn.1002-1302.2023.13.008

杜鹃叶绿体基因组特征及密码子偏好性分析

万露露^{1,2}, 范敦锦¹, 王中煊^{1,2}, 龚娜¹, 丁祥青², 陈世品^{1,2}, 苏小青¹

(1. 福建农林大学林学院, 福建福州 350002;

2. 福建农林大学园林学院/兰科植物保护与利用国家林业和草原局重点实验室, 福建福州 350002)

摘要:为了确定杜鹃(*Rhododendron simsii*)叶绿体基因组密码子的使用模式,为杜鹃花科叶绿体基因组学研究提供参考依据,绘制了杜鹃叶绿体基因组图谱,对三大功能基因进行分类,统计各蛋白编码基因密码子的3位碱基的GC含量、有效密码子数、同义密码子相对使用度,开展中性绘图分析和PR2-plot等分析,并系统分析影响杜鹃密码子偏好性的各种因素,从而筛选出了最优密码子。结果显示,杜鹃叶绿体基因组全长206 912 bp,大单拷贝区为110 234 bp,小单拷贝区为2 606 bp,反向重复区为47 036 bp。蛋白编码序列GC含量为36%,密码子第3位偏好使用A/U,ENC值范围为37.48~57.13,说明其密码子使用偏好性较弱;中性绘图分析显示GC₁₂和GC₃的相关系数为0.255,拟合的回归系数为0.345;PR2-plot分析显示G>C,T>A;有30个密码子的RSCU值大于1,以U或A结尾的有29个,影响其密码子使用偏好性的最主要为自然选择压力;最终鉴定出最优密码子8个,分别为GCA、GAU、UUG、CCA、CGU、UCA、GUA、UAU。可见,杜鹃叶绿体基因组具典型的双链环状构造,密码子偏好性较弱,主要受到自然和突变的共同影响。

关键词:杜鹃;叶绿体基因组;密码子偏好性;最优密码子

中图分类号:S685.210.32 **文献标志码:**A **文章编号:**1002-1302(2023)13-0054-06

杜鹃(*Rhododendron simsii*),隶属杜鹃花科(Ericaceae Juss)杜鹃属(*Rhododendron*)映山红亚属(subg. *Tsutsusi*),其花冠为漏斗形,美艳异常,极具观赏价值,且花期较长,是重要的育种资源。对杜鹃的研究主要集中在组培快繁、杂交育种、重要基因家族分析等方面^[1-3],尚未对其密码子的使用偏好性进行分析。叶绿体基因组具有结构稳定、进化保守的优点,特定物种更倾向于使用特定的密码子,对特定基因的功能和表达水平产生一定的影响^[4-6]。本研究分析杜鹃密码子使用偏好的影响因素,鉴定其最优密码子,揭示杜鹃叶绿体密码子的使用模式,旨在为杜鹃叶绿体基因组在良种选育中的应用提供参考,为其后续的育种研究以及品种优化提供帮助。

1 材料与方法

1.1 叶绿体基因组密码子基因序列获取

从NCBI数据库中,下载杜鹃叶绿体基因组序

列(登录号:MT239364.1),参照Wang等的研究^[7],按照国际惯例对CDS进行筛选,以便用于后续叶绿体密码子使用偏好性的分析。

1.2 叶绿体基因组密码子的相关参数分析

计算密码子的各种相关参数。具体计算方法参照江淑珍等的研究^[8]。

1.3 叶绿体基因组密码子绘图分析

以(GC₃,GC₁₂)坐标点绘制散点图以及添加趋势线。中性绘图与PR2-plot偏倚分析具体方法参照龙凌云等的研究^[9]。

1.4 叶绿体基因组最优密码子分析

将44条基因的ENC值按由大至小降序排列,构建高、低表达库。采用两端筛选的各5条基因,并按照李江飞等的方法^[10],定义最优密码子。

2 结果与分析

2.1 叶绿体基因组的基因组成

杜鹃叶绿体的基因组全长206 912 bp,具典型的双链环状构造,大单拷贝区为110 234 bp,小单拷贝区为2 606 bp,反向重复区分别为47 036 bp(图1)。相比于其他被子植物,杜鹃叶绿体基因组较大,IR区显著扩张而SSC区显著缩小,2个IR区相互对称,且被LSC区和SSC区隔开,IR区的扩张是

收稿日期:2022-10-04

基金项目:国家重点研发计划(编号:2019YFD1001000)。

作者简介:万露露(1998—),女,江西进贤人,硕士研究生,主要从事森林培育与分子生物学研究。E-mail:wan201920@163.com。

通信作者:苏小青,博士,副教授,主要从事森林培育和森林生态学研究。E-mail:fjsxqsxq@126.com。

导致其叶绿体基因组较大的主要因素。叶绿体的 GC 含量为 36%, 与其他绿色植物的叶绿体含量比较接近。一共预测注释到 136 个基因, 包括 90 个蛋白编码基因、38 个 tRNA 基因、8 个 rRNA 基因。在大多数情况下, AUG 被用作 CDS 的起始密码子, 尽

管在某些植物叶绿体基因组的某些 CDS 的起始密码子中有时会观察到 RNA 编辑。在本研究中, *ndhD* 以 GUG 为起始密码子, *atpB* 和 *atpE* 存在 4 bp 重叠, *psbC* 和 *psbD* 存在 53 bp 重叠, *rps3* 和 *rpl22* 存在 16 bp 重叠^[11]。

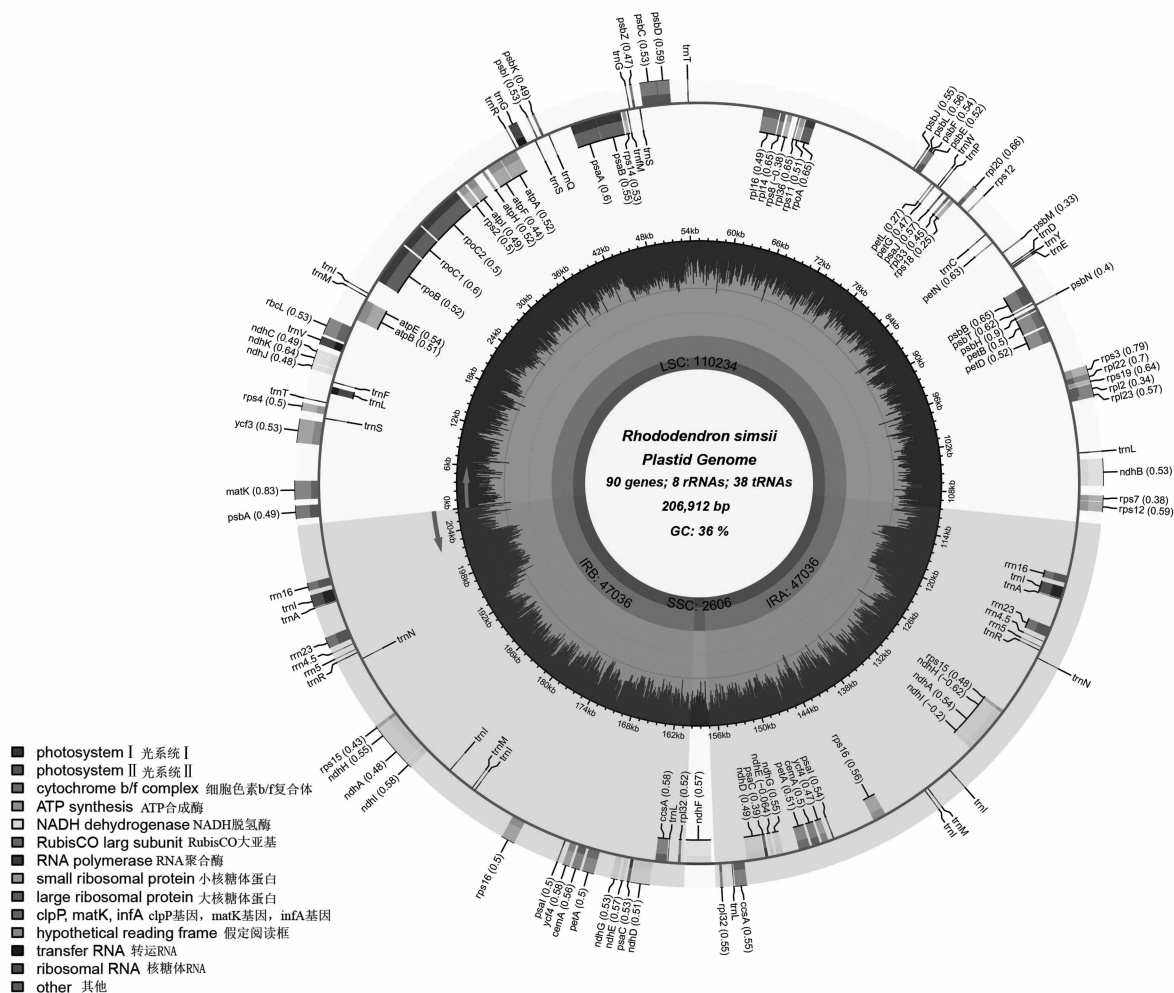


图1 杜鹃叶绿体基因组图谱

将叶绿体基因的不同功能分为 3 个大类, 依次为光合作用能力、自我复制能力、其他和功能未知。光合作用的相关基因分为 6 个小类(7 个光系统 I 的基因、15 个光系统 II 的基因、17 个 NADH - 脱氢酶基因、7 个细胞色素 b/f 复合物基因、6 个 ATP 合酶基因、1 个羧化酶基因); 自我复制的相关基因分为 5 个小类(10 个核糖体大亚基的基因、14 个核糖体小亚基的基因、4 个 RNA 聚合酶基因、38 个 tRNA 基因、8 个 rRNA 基因); 其他基因分为 5 个小类(表 1)。叶绿体基因组中包含 2 个内含子(如 *rps12*、*ycf3*), 有些基因具有 2 个拷贝(如 *psaC*、*psaI* 等), 均符合杜鹃属叶绿体基因组的规律。

2.2 密码子主要相关指数分析

密码子组成分析表明, 杜鹃叶绿体基因组蛋白编码基因密码子 GC_1 (46.78%) > GC_2 (39.58%) > GC_3 (27.75%), 说明 GC 在不同位置的分布是不均匀的, 但多以碱基 A、T 结尾。各基因有效密码子数 (ENC) 介于 37.48 ~ 57.13 之间, 平均值 (46.63) 远大于 35, 且有 32 条基因的有效密码子数值大于 45, 表明其使用偏性较弱(表 2)。

对密码子各参数之间的相关性进行了研究, 结果(表 3)表明: ENC 与密码子数目极显著相关, 相关系数为 1.000, 说明基因序列大小显著影响密码子偏好性; GC_{all} 与 GC_1 、 GC_2 、 GC_3 之间极显著相关

表 1 叶绿体基因组基因分类

类别	家族	基因
光合作用	光系统 I 的亚基	<i>psaA</i> 、 <i>psaB</i> 、 <i>psaC^a</i> 、 <i>psaI^a</i> 、 <i>psaJ</i>
	光系统 II 的亚基	<i>psbA</i> 、 <i>psbB</i> 、 <i>psbC</i> 、 <i>psbD</i> 、 <i>psbE</i> 、 <i>psbF</i> 、 <i>psbH</i> 、 <i>psbI</i> 、 <i>psbJ</i> 、 <i>psbK</i> 、 <i>psbL</i> 、 <i>psbM</i> 、 <i>psbN</i> 、 <i>psbT</i> 、 <i>psbZ</i>
	NADH－脱氢酶的亚基	<i>ndhA^{ab}</i> 、 <i>ndhB^b</i> 、 <i>ndhC</i> 、 <i>ndhD^a</i> 、 <i>ndhE^a</i> 、 <i>ndhF</i> 、 <i>ndhG^a</i> 、 <i>ndhH^a</i> 、 <i>ndhI^a</i> 、 <i>ndhJ</i> 、 <i>ndhK</i>
	细胞色素 b/f 复合物的亚基	<i>petA^a</i> 、 <i>petB^b</i> 、 <i>petD^b</i> 、 <i>petG</i> 、 <i>petL</i> 、 <i>petN</i>
	ATP 合酶亚基	<i>atpA</i> 、 <i>atpB</i> 、 <i>atpE</i> 、 <i>atpF^b</i> 、 <i>atpH</i> 、 <i>atpI</i>
	二磷酸核酮糖氧合酶/羧化酶亚基	<i>rbcL</i>
自我复制	核糖体大亚基	<i>rpl14</i> 、 <i>rpl16^b</i> 、 <i>rpl2</i> 、 <i>rpl20</i> 、 <i>rpl22</i> 、 <i>rpl23</i> 、 <i>rpl32^a</i> 、 <i>rpl33</i> 、 <i>rpl36</i>
	核糖体小亚基	<i>rps11</i> 、 <i>rps12^c</i> 、 <i>rps14</i> 、 <i>rps15a</i> 、 <i>rps16^{ab}</i> 、 <i>rps18</i> 、 <i>rps19</i> 、 <i>rps2</i> 、 <i>rps3</i> 、 <i>rps4</i> 、 <i>rps7</i> 、 <i>rps8</i>
	RNA 聚合酶	<i>rpoA</i> 、 <i>rpoB^b</i> 、 <i>rpoC1</i> 、 <i>rpoC2</i>
	rRNA	<i>rrn16^a</i> 、 <i>rrn23^a</i> 、 <i>rrn4</i> 、 <i>5^a</i> 、 <i>rrn5^a</i>
	tRNA	<i>trnA</i> － <i>UGC^a</i> 、 <i>trnC</i> － <i>GCA</i> 、 <i>trnD</i> － <i>GUC</i> 、 <i>trnE</i> － <i>UUC</i> 、 <i>trnF</i> － <i>GAA</i> 、 <i>trnG</i> － <i>GCC</i> 、 <i>trnG</i> － <i>UCC^b</i> 、 <i>trnH</i> － <i>GUG</i> 、 <i>trnI</i> － <i>CAU^a</i> 、 <i>trnI</i> － <i>GAU^{ab}</i> 、 <i>trnL</i> － <i>CAA</i> 、 <i>trnL</i> － <i>UAA^b</i> 、 <i>trnL</i> － <i>UAG^a</i> 、 <i>trnM</i> － <i>CAU^a</i> 、 <i>trnN</i> － <i>GUU^a</i> 、 <i>trnP</i> － <i>UGG</i> 、 <i>trnQ</i> － <i>UUG</i> 、 <i>trnR</i> － <i>ACG^a</i> 、 <i>trnR</i> － <i>UCU^a</i> 、 <i>trnS</i> － <i>GCU</i> 、 <i>trnS</i> － <i>GGA</i> 、 <i>trnS</i> － <i>UGA</i> 、 <i>trnT</i> － <i>GGU</i> 、 <i>trnT</i> － <i>UGU</i> 、 <i>trnV</i> － <i>GAC^a</i> 、 <i>trnV</i> － <i>UAC^b</i> 、 <i>trnW</i> － <i>CCA</i> 、 <i>trnY</i> － <i>GUA</i> 、 <i>trnfM</i> － <i>CAU</i>
	其他和	
功能未知	包膜蛋白基因	<i>cemA^a</i>
	成熟酶	<i>matk</i>
	蛋白酶	<i>clpP</i>
	c 型细胞色素合成基因	<i>ccsA^a</i>
	假定叶绿体阅读框	<i>ycf3c</i> 、 <i>ycf4a</i>

注：a：表示有 2 个拷贝基因；b：表示包含 1 个内含子的基因；c：表示包含 2 个内含子的基因；

表 2 叶绿体基因组的主要参数

基因	GC ₁	GC ₂	GC ₃	GC _{all}	ENC	基因	GC ₁	GC ₂	GC ₃	GC _{all}	ENC
<i>rps12</i>	50.81	46.77	27.42	41.67	44.22	<i>ccsA</i>	35.20	37.38	27.41	33.33	47.88
<i>ycf3</i>	47.93	37.87	29.59	38.46	49.36	<i>ndhE</i>	40.20	34.31	23.53	32.68	39.82
<i>rps4</i>	46.89	38.76	25.36	37.00	47.18	<i>ndhG</i>	41.57	33.71	23.03	32.77	47.49
<i>ndhJ</i>	49.69	36.48	29.56	38.57	46.51	<i>petA</i>	52.02	36.45	31.46	39.98	49.10
<i>ndhK</i>	44.25	42.48	24.34	37.02	49.77	<i>cemA</i>	39.57	28.26	32.61	33.48	46.46
<i>ndhC</i>	47.11	31.40	28.10	35.54	57.13	<i>ycf4</i>	48.11	39.46	27.57	38.38	48.81
<i>rbcL</i>	58.40	43.28	28.99	43.56	47.34	<i>accD</i>	54.92	29.31	30.61	38.28	48.88
<i>atpB</i>	56.31	41.68	27.66	41.88	46.20	<i>psbD</i>	51.98	43.22	32.20	42.47	45.77
<i>atpE</i>	48.51	38.06	28.36	38.31	45.29	<i>psbC</i>	54.01	46.41	31.86	44.09	45.55
<i>rpl14</i>	54.89	36.84	26.32	39.35	48.59	<i>rps14</i>	47.52	48.51	32.67	42.90	42.24
<i>rps8</i>	39.85	39.10	22.56	33.83	39.64	<i>psaB</i>	48.44	42.99	31.16	40.86	48.35
<i>rps11</i>	48.20	53.96	21.58	41.25	51.88	<i>psaA</i>	52.06	43.14	32.36	42.52	50.17
<i>rpoA</i>	43.64	35.64	24.00	34.42	43.54	<i>atpA</i>	54.33	40.16	25.59	40.03	44.40
<i>rps18</i>	33.63	41.59	24.78	33.33	37.48	<i>atpF</i>	45.79	34.21	32.63	37.54	45.22
<i>rpl20</i>	37.72	47.37	28.07	37.72	43.00	<i>atpI</i>	49.19	37.10	30.24	38.84	46.12
<i>psbB</i>	54.81	45.97	31.24	44.01	48.56	<i>rps2</i>	41.77	40.51	29.96	37.41	50.51
<i>petD</i>	48.57	40.57	24.57	37.90	44.34	<i>rpoC1</i>	48.99	36.96	27.68	37.87	48.18
<i>rps3</i>	44.92	34.32	25.00	34.75	49.38	<i>rpoC2</i>	45.45	36.58	28.58	36.87	49.61
<i>rpl22</i>	39.35	38.71	24.52	34.19	41.98	<i>ndhI</i>	42.61	36.93	28.41	35.98	48.05
<i>rpl2</i>	47.76	48.88	30.22	42.29	50.45	<i>ndhA</i>	41.21	39.56	21.98	34.25	41.77
<i>ndhB</i>	42.03	38.05	31.87	37.32	47.03	<i>ndhH</i>	51.02	36.55	25.38	37.65	49.27
<i>rps7</i>	51.28	42.31	25.64	39.74	46.54	平均值	46.78	39.58	27.75	38.00	46.63
<i>ndhF</i>	35.83	35.16	24.20	31.73	42.78						

表 3 基因密码子各参数之间的相关性分析

变量	相关系数				
	GC ₁	GC ₂	GC ₃	GC _{all}	ENC
GC ₂	0.220				
GC ₃	0.354 *	0.022			
GC _{all}	0.821 **	0.655 **	0.543 **		
ENC	0.400 **	-0.062	0.285	0.297	
密码子数	0.400 **	-0.062	0.285	0.297	1.000 **

注: * 表示相关性达到显著水平 ($P < 0.05$); ** 表示相关性达到极显著水平 ($P < 0.01$)。

($P < 0.01$), 这表明其叶绿体基因组密码子碱基组成的相似度较高; GC₁ 与 ENC 极显著相关 ($P < 0.01$), 说明 GC₁ 对密码子偏好性存在较强影响。

对各氨基酸同义密码子相对使用度进行数据分析(表 4), 得出: 氨基酸 RSCU > 1 的密码子一共有 30 个, 其中 29 个密码子以碱基 A、U 结尾, 所占比例较高(97%)。表明杜鹃叶绿体基因组偏好使用以 A/U 结尾的密码子; 相反, RSCU < 1 的密码子, 则偏好使用 G/C 结尾。

2.3 中性绘图分析

对密码子进行中性绘图分析, 结果(图 2)显示: 纵坐标 GC₁₂ 介于 0.34 ~ 0.54, 横坐标 GC₃ 介于 0.22 ~ 0.51。所有基因均处于坐标轴对角线上方, 密码子 GC₁₂ 与密码子 GC₃ 的相关性表现为不显著, 相关系数为 0.255, 拟合的回归系数为 0.345, 说明密码子的 3 个碱基可能有着一定差异。由此可以推断, 杜鹃叶绿体基因组的密码子偏好性受到自然选择的影响较大。

2.4 PR2 - plot 分析

偏倚分析(图 3)表明: 杜鹃在 4 个区域叶绿体基因分布是不均衡的, 且大多数基因散布在整个平面图的右下方, 这表明 4 个碱基有不一样的使用频率, 即 $G > C, T > A$; 因此, 可以推断在杜鹃叶绿体基因组密码子使用模式上, 其偏好性很可能会受到突变和自然选择的共同作用。

2.5 最优密码子分析

根据 $\Delta RSCU > 0.08$ 为高表达密码子, 已筛选出杜鹃叶绿体基因组有 27 个, 其高表达密码子中有 5 个是以碱基 A 结尾, 3 个是以碱基 U 结尾, 9 个是以碱基 G 结尾, 10 个是以碱基 C 结尾(表 5)。综合考虑杜鹃叶绿体基因 RSCU 的分析结果(表 4), 最终得到最优密码子 8 个, 分别为 GCA、GAU、UUG、CCA、CGU、UCA、GUA、UAU, 其中以碱基 A 结尾的

表 4 相对同义密码子使用度

氨基酸	密码子	数量	RSCU	氨基酸	密码子	数量	RSCU
Ter	UAA *	22	1.500 0	Met	AUG	318	1.000 0
	UAG	8	0.545 5	Asn	AAC	124	0.413 3
	UGA	14	0.954 5		AAU *	476	1.586 7
Ala	GCA *	255	1.108 7	Pro	CCA *	173	1.140 0
	GCC	150	0.652 2		CCC	110	0.724 9
	GCG	103	0.447 8		CCG	67	0.441 5
Cys	GCU *	412	1.791 3		CCU *	257	1.693 6
	UGC	39	0.475 6	Gln	CAA *	402	1.588 9
	UGU *	125	1.524 4		CAG	104	0.411 1
Asp	GAC	112	0.390 9	Arg	AGA *	239	1.691 0
	GAU *	461	1.609 1		AGG	66	0.467 0
	GAA *	554	1.528 3		CGA *	213	1.507 1
Glu	GAG	171	0.471 7		CGC	50	0.353 8
	UUC	243	0.639 5		CGG	52	0.367 9
	UUU *	517	1.360 5		CGU *	228	1.613 2
Gly	GGA *	404	1.508 9	Ser	AGC	64	0.403 8
	GGC	114	0.425 8		AGU *	199	1.255 5
	GGG	184	0.687 2		UCA *	183	1.154 6
His	GGU *	369	1.378 2		UCC	150	0.946 4
	CAC	78	0.472 7		UCG	69	0.435 3
	CAU *	252	1.527 3		UCU *	286	1.804 4
Ile	AUA	370	0.898 8	Thr	ACA *	220	1.178 0
	AUC	243	0.590 3		ACC	152	0.813 9
	AUU *	622	1.510 9		ACG	66	0.353 4
Lys	AAA *	557	1.569 0		ACU *	309	1.654 6
	AAG	153	0.431 0	Val	GUA *	304	1.470 4
Leu	CUA	198	0.780 6		GUC	101	0.488 5
	CUC	94	0.370 6		GUG	115	0.556 2
	CUG	82	0.323 3		GUU *	307	1.484 9
	CUU *	317	1.249 7	Trp	UGG	268	1.000 0
	UUA *	538	2.120 9	Tyr	UAC	93	0.357 7
	UUG *	293	1.155 1		UAU *	427	1.642 3

注: * 表示 RSCU > 1 。

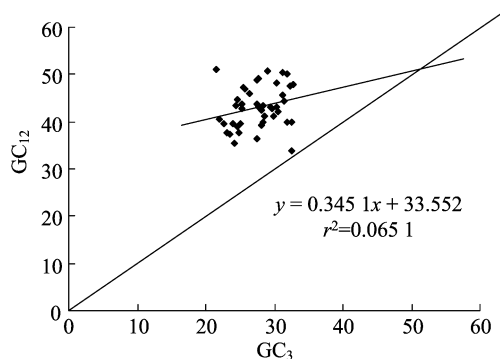


图2 中性绘图分析

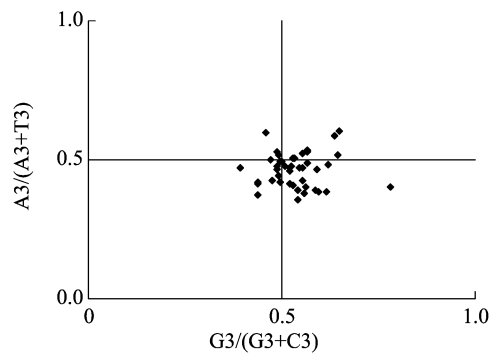


图3 PR2-plot 绘图分析

表 5 最优密码子分析

氨基酸	密码子	高表达基因		低表达基因		Δ RSCU
		数量	RSCU	数量	RSCU	
Ter	UGA	1	0.600 0	0	0.000 0	0.600 0
Ala	<u>GCA</u>	51	1.658 5	12	1.021 3	0.637 2
	GCG	12	0.390 2	3	0.255 3	0.134 9
Cys	UGC	2	0.235 3	0	0.000 0	0.235 3
Asp	<u>GAU</u>	48	1.714 3	16	1.600 0	0.114 3
Glu	GAG	12	0.585 4	7	0.400 0	0.185 4
Phe	UUC	25	0.641 0	7	0.368 4	0.272 6
Gly	GGC	23	0.691 7	4	0.290 9	0.400 8
Asn	AAC	21	0.763 6	9	0.409 1	0.354 5
His	CAC	13	0.426 2	1	0.285 7	0.140 5
Lys	<u>AAG</u>	16	0.405 1	9	0.310 3	0.094 8
Leu	CUC	9	0.364 9	2	0.114 3	0.250 6
	CUG	13	0.527 0	4	0.228 6	0.298 4
	<u>UUG</u>	30	1.216 2	16	0.914 3	0.301 9
Pro	<u>CCA</u>	15	0.937 5	6	0.827 6	0.109 9
	CCC	12	0.750 0	3	0.413 8	0.336 2
Gln	CAG	12	0.600 0	0	0.000 0	0.600 0
Arg	CGC	6	0.346 2	1	0.088 2	0.258 0
	CGG	5	0.288 5	1	0.088 2	0.200 3
	<u>CGU</u>	33	1.903 8	13	1.147 1	0.756 7
Ser	AGC	10	0.681 8	3	0.233 8	0.448 0
	<u>UCA</u>	18	1.227 3	12	0.935 1	0.292 2
Thr	ACC	26	1.118 3	8	0.761 9	0.356 4
	ACG	8	0.344 1	2	0.190 5	0.153 6
Val	<u>GUA</u>	36	1.618 0	15	1.333 3	0.284 7
	GUG	15	0.674 2	5	0.444 4	0.229 8
	<u>UAU</u>	37	1.720 9	23	1.586 2	0.134 7

注:有下划线的表示最优密码子。

有 4 个,U 碱基结尾有 3 个,碱基 G 结尾仅有 1 个。

3 讨论与结论

在杜鹃叶绿体基因组中,密码子第 3 位的 GC

含量远远少于前 2 位碱基,说明其更偏好使用 A/U,也表明了杜鹃叶绿体基因组密码子使用模式在进化上相对保守,遵循了双子叶植物叶绿体基因组密码子的第 3 位高 A/U 偏好性的规律。杜鹃叶绿体基因组中存在 8 个最优密码子,分别为 GCA、GAU、UUG、CCA、CGU、UCA、GUA、UAU,其中 3 个以 U 结尾,4 个以 A 结尾,1 个以 G 结尾。一般认为富含 A/U 碱基的序列中,最优密码子也同样富含 A/U,而富含 G/C 碱基的序列中,最优密码子也同样富含 G/C^[12],本研究证实了这一结论。同时,如果密码子的偏好性与物种亲缘性高度相关,或者亲缘关系更接近,那么 2 种生物之间的密码子会有更相近的使用模式。贾守宁等研究表明,杜鹃属 9 个物种中密码子的偏好性均保持高度一致^[13],也同样论证了这个观点。

在不同物种中密码子偏好性形成的主要影响因素有较大差异。倒提壶(*Cynoglossum amabile*)叶绿体基因组密码子的偏好性主要受自然选择压力的影响^[14],这可能因为倒提壶是生长在陆地上的高等植物,在进化上比较保守。蒺藜苜蓿(*Medicago truncatula*)主要是受到突变压力的影响^[15],这可能是由于其植株的再生时间比较短,具有大量的突变体,生物多样性较高。芒果(*Mangifera indicate*)同时受自然选择和突变压力的共同作用^[16],这可能是由于芒果喜高温干燥的气候,但是不耐严寒低温。镰翅羊耳蒜(*Liparis bootanensis*)除了受自然选择、突变压力的作用,可能还受碱基组成等其他因素的影响^[17],因为它是附生草本,生长环境变动较大,受到多种因素的影响。本研究中性绘图结果表明,杜鹃叶绿体基因组密码子的偏好性主要是受自然选择的作用。PR2-plot 分析表明,叶绿体基因组的密码子偏好也可能受到突变压力和自然选择的共同作用;表明杜鹃叶绿体基因组密码子的偏好性不是单个因素作用的结果,可能受到突变压力、自然选择共同因子的作用。

因杜鹃属植物具有典型频繁的自然杂交现象,当 2 个遗传背景差异较大的个体产生可育杂交后代时,基因组发生了重组与整合,这可能会导致该属植物的叶绿体基因组结构发生变异。前人研究表明,杜鹃属植物大多地处严寒低氧的高原地带,受环境胁迫的影响,可能发生基因的部分缺失和复制^[13]。杜鹃属的不同物种间基因缺失的情况差别很大,说明了杜鹃属植物叶绿体基因组具有较高的

突变性。Shen 等比较了 8 个不同杜鹃花科的叶绿体基因组,发现其基因组大小、结构和 GC 含量存在显著差异,这种差异被认为主要是由维持生活的营养方式不同导致的(自养和异养)^[18]。这表明,由于存在自然杂交和不同生活方式的缘故,可能会导致杜鹃和其他杜鹃属植物的叶绿体基因组密码子的偏好性存在相对较大偏差。刘仁林的研究表明,映山红亚属很可能由耳叶杜鹃(*R. auriculatum*)亚组最近的祖先类群演化而来,这些祖先类群大多分布于中国广西壮族自治区、广东省和湖南省南部;但映山红亚属的分布区十分广阔,除了新疆、宁夏 2 地至今没有记载之外,全国其他各省、市均有分布^[19]。杜鹃在演化过程中,要适应各种自然环境变化导致的自然选择的巨大压力,这可能是杜鹃叶绿体基因组密码子使用偏好性主要受自然选择影响的原因所在;具有广阔的分布范围,是叶绿体基因组密码子使用偏好性主要受自然选择影响的植物的一种共性。

探究密码子偏好性的使用模式,获得最优密码子,具有十分重要的意义。以最优密码子为依据,设计叶绿体基因的表达载体,能够更有效地提升基因在叶绿体基因组中的表达效率和水平^[20-21]。另外,已知密码子的使用模式有助于推断和预测未知基因的表达和功能^[22],从而对基因工程和育种提供参考。本研究揭示了杜鹃叶绿体基因组密码子的使用模式和最优密码子,旨在为杜鹃及其近缘物种的密码子优化提供理论依据,为后续的育种研究以及品种优化提供帮助。

参考文献:

- [1] 胡计红,陈桂信,杨惠婷,等. 屏南龙源四季杜鹃古树组培快繁技术研究[J]. 热带作物学报,2020,41(4):755-763.
- [2] 苏学鹏. 四季杜鹃红山茶与山茶种间杂交技术要点浅析[J]. 南方农业,2020,14(32):219-220.
- [3] 杨丽,严露露,李慧娥,等. 映山红杜鹃 MADS 基因家族的鉴定与分析[J]. 分子植物育种,2021,19(19):6290-6301.
- [4] 罗雪,严贤春,侯哲,等. 珙桐叶绿体基因组密码子使用偏好性分析[J/OL]. 分子植物育种. [2022-09-06]. <https://kns.cnki.net/kcms/detail/46.1068.S.20220830.1837.009.html>.
- [5] 朱灵芝,朱沛煌,李荣,等. 马尾松 *PmDXR* 基因密码子偏好性分析[J]. 林业科学研究,2021,34(2):102-113.
- [6] 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.
- [7] Wang Z J,Xu B B,Li B,et al. Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species[J]. PeerJ,2020,8:e8251.
 - [8] 江淑珍,连辉,熊远芳,等. 米槠叶绿体基因组密码子偏好性分析[J/OL]. 分子植物育. [2022-06-30]. <http://kns.cnki.net/kcms/detail/46.1068.S.20210722.1131.006.html>.
 - [9] 龙凌云,黄秋伟,李慧敏,等. 金柑叶绿体基因组密码子使用偏好性分析[J/OL]. 分子植物育种. (2022-05-20)[2022-07-01]. <https://kns.cnki.net/kcms/detail/46.1068.S.20220520.1055.002.html>.
 - [10] 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.
 - [11] Li H E,Guo Q,Li Q,et al. Long-reads reveal that *Rhododendron delavayi* plastid genome contains extensive repeat sequences, and recombination exists among plastid genomes of photosynthetic Ericaceae[J]. PeerJ,2020,8:e9048.
 - [12] 郝丙青,夏莹莹,叶航,等. 香花油茶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(9):178-186.
 - [13] 贾守宁,张颖敏,赵国福,等. 头花杜鹃、陇蜀杜鹃及杜鹃属植物叶绿体基因组比较分析[J/OL]. 广西植物. [2022-07-05]. <https://kns.cnki.net/kcms/detail/45.1134.Q.20220629.1301.002.html>
 - [14] 赵振宁,余潇,寸孟人,等. 倒提壶(*Cynoglossum amabile* Stapf et Drumm.) 叶绿体基因组密码子偏好性分析[J/OL]. 分子植物育种. [2022-07-06]. <https://kns.cnki.net/kcms/detail/46.1068.S.20220630.1511.008.html>
 - [15] 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.
 - [16] 唐玉娟,赵英,黄国弟,等. 芒果叶绿体基因组密码子使用偏好性分析[J]. 热带作物学报,2021,42(8):2143-2150.
 - [17] 刘江枫. 镰翅羊耳蒜叶绿体基因组密码子偏好性分析[J]. 福建农业学报,2021,36(6):629-635.
 - [18] Shen J S,Li X Q,Zhu X T,et al. The complete plastid genome of *Rhododendron pulchrum* and comparative genetic analysis of Ericaceae species[J]. Forests,2020,11(2):158.
 - [19] 刘仁林. 中国杜鹃属映山红亚属植物系统分类研究[D]. 南昌:江西农业大学:139-140.
 - [20] 李显煌,杨生超,辛雅萱,等. 灯盏花叶绿体基因组密码子偏好性分析[J]. 云南农业大学学报(自然科学版),2021(3):384-392.
 - [21] 张海霞,王玉道,许雪妮. 苜蓿质膜内在蛋白编码基因 MsPIPs 家族的密码子偏好性分析[J]. 江苏农业学报,2021,37(6):1393-1399.
 - [22] Tang D F,Wei F,Cai Z Q,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):1-9.