

2 种羌活药用植物的叶绿体基因组密码子使用偏性分析

王琳璇, 王高峰, 谷威, 赵志敏, 马永贵* (青海师范大学生命科学学院, 青海西宁 810000)

摘要 对羌活(*Notopterygium incisum*)与宽叶羌活(*N. franchetii*)的叶绿体编码基因的密码子使用偏性(codon usage bias)及其影响因素进行了分析。结果表明:羌活和宽叶羌活的叶绿体基因组总 ENC 值分别为 49.79 和 49.70, 偏性较弱, 密码子第 3 位碱基倾向于使用 A/T。PR2-plot 分析表明, 选择压力和突变都是影响 2 种羌活叶绿体基因组密码子使用偏性的主要因素; 而中性绘图分析和 ENC-plot 分析表明选择压力是影响 2 种羌活叶绿体基因组密码子使用偏性的主要因素。相对同义密码子使用度(RSCU)分析确定了 2 种羌活中使用频率较高的 30 个高频密码子, 几乎都以 A 或 U 结尾, 只有 UUG 以 G 结尾。羌活的最优密码子 9 个, 宽叶羌活的最优密码子 7 个。2 种羌活有 5 个最优密码子相同, 分别是 UUG、CAU、CGU、AGU、GGA。该研究结果为后续药用羌活植物的基因表达、蛋白质翻译和遗传育种研究提供了研究依据和基础数据。

关键词 羌活; 宽叶羌活; 叶绿体基因组; 密码子使用偏性; 最优密码子

中图分类号 Q943.2 文献标识码 A

文章编号 0517-6611(2022)03-0095-07

doi:10.3969/j.issn.0517-6611.2022.03.025



开放科学(资源服务)标识码(OSID):

Codon Usage Bias in Chloroplast Genomes of Two Medicinal Species of *Notopterygium* H. de Boissieu

WANG Lin-xuan, WANG Gao-feng, GU Wei et al (College of Life Sciences, Qinghai Normal University, Xining, Qinghai 810000)

Abstract This study analyzed the codon usage bias and affecting factors of the protein coding genes (CDS) of the chloroplast genomes of two medicinal plants, *Notopterygium incisum* and *N. franchetii*. The results showed that the chloroplast genomes of *N. incisum* and *N. franchetii* had a total ENC value of 49.79 and 49.70, respectively, indicating weak codon usage bias. The third base of the codon of the two chloroplast genomes preferred to use A/T. The analysis of PR2-plot showed that selection pressure and mutation were the main factors affecting the bias of codon usage in *N. incisum* and *N. franchetii* CDS, while neutrality plot and ENC-plot analyses both illustrated selection pressure was the main factors affecting codon usage bias in the two species. Relative Synonymous Codon Usage (RSCU) analysis determined 30 high-frequency codons used in *N. incisum* and *N. franchetii*, almost all ending with A or U, with only UUG ends with G. There were nine optimal codons identified in *N. incisum*, and seven optimal codons found in *N. franchetii*. The five optimal codons shared by *N. incisum* and *N. franchetii* were namely UUG, CAU, CGU, AGU, GGA. The results provide important basal data for the future study of gene expression, protein translation and molecular breeding of medicinal *Notopterygium* plants.

Key words *Notopterygium incisum*; *Notopterygium franchetii*; Chloroplast genome; Codon usage bias; Optimal codons

密码子在蛋白质合成过程中起着传递遗传信息的关键作用, 在大多数情况下, 编码同一氨基酸的同一密码子的使用频率是不同的, 这一现象称为密码子使用偏性(codon usage bias)。某一物种或某一基因通常倾向于使用一种或几种特定的同义密码子, 这些密码子被称为最优密码子(optimal codon)。影响密码子使用偏性的因素很多, 最主要的因素是定向突变和中性选择^[1-3], 还包括 GC 含量、同义替换率^[4]、tRNA 丰度^[5]、有效翻译起始位点的选择^[6]、密码子亲水性和 DNA 复制起始位点^[7]、基因长度^[8]和表达水平^[9]等。在不同物种中, 突变和选择的影响强度不同。例如在蓝藻中, 影响密码子使用偏性较大的是 GC 组成与选择压力^[10], 但在果蝇中则是突变对的密码子使用偏性影响较大^[3]。

叶绿体是植物进行光合作用的器官, 也参与到植物的氨基酸合成等通路中。作为植物重要的细胞器, 叶绿体有自己一套独特的遗传信息^[11]。针对叶绿体基因组的密码子使用偏性研究, 可以帮助了解叶绿体蛋白质编码基因的使用偏好, 为后续遗传改造和基因表达研究提供基础数据。研究发现, 乳油木叶绿体基因组密码子使用偏性更多受突变影响^[12], 选择压力对其使用偏性影响较小。杜梨叶绿体密

子使用偏性则主要受选择的影响^[13]。胡晓艳等^[14]在对酸枣叶绿体基因组的研究中获得了 7 个最优密码子。

羌活属(*Notopterygium* Boissieu)植物隶属于伞形科(*Apiaceae*)美味芹族, 是我国特有的多年生草本药用植物, 主要分布在我国西南部高山地区^[15]。据《中国植物志(英文版)》记载, 该属包括宽叶羌活、羌活、羽苞羌活、澜沧羌活、卵叶羌活和细叶羌活 6 个种^[16]。《中国药典》规定的羌活基源植物为羌活与宽叶羌活, 入药部位为干燥根及根茎, 具有解表散寒、祛风除湿和止痛的功效。羌活主要产于四川、云南, 次产于青海及甘肃; 宽叶羌活主产于甘肃、青海, 次产于四川、山西和内蒙古等地。近年来, 由于过度采挖及生境受到严重破坏, 野生羌活的自然群体已经面临濒危, 被载入《中国生物多样性红色名录(高等植物卷)》^[17]。羌活作为一味传统的中药, 富含异欧芹素、羌活醇、阿魏酸等活性物质, 具有抗炎、抗氧化、抗心律失常、抗菌、抗癌细胞增殖等药效作用^[18]。目前有关羌活属叶绿体基因组密码子使用偏性研究鲜见报道。该研究拟对 2 种可以入药的羌活属植物, 即羌活和宽叶羌活进行叶绿体基因组的密码子使用模式分析, 研究结果不但可以阐明羌活光合作用相关基因的使用偏好, 也可为羌活基因表达研究及遗传育种研究提供理论基础。

1 材料与方法

1.1 数据获取 羌活样品采集于青海省果洛藏族自治州达日县, 99.46° E, 33.77° N, 海拔 3 943.24 m。宽叶羌活样品采集于青海省玛可河麦浪沟, 100.96° E, 32.84° N, 海拔 3 790.78 m。

基金项目 青海省基础研究计划项目(2018-ZJ-741)。

作者简介 王琳璇(1994—), 女, 河北廊坊人, 硕士研究生, 研究方向: 生物化学与分子生物学。* 通信作者, 教授, 博士, 从事动植物分子进化与适应研究。

收稿日期 2021-04-03

保存于干冰中,送于北京百迈客生物科技有限公司进行叶绿体基因组测序。利用 Illumina HiSeq X Ten 平台测序,使用 MIRA 序列组装软件的 MIRA v3.4.1.1 模块进行,最终得到羌活的叶绿体全基因组序列长度为 157 050 bp。宽叶羌活的叶绿体全基因组序列长度为 142 184 bp。测序结果提交至 NCBI (national center for biotechnology information), 登录号:羌活 MW820163、宽叶羌活 MW820162。使用 Geneious11.1.5 提取编码基因 (coding sequence, CDS) 并筛选,剔除小于 300 bp 的基因,并选择起始密码子为 ATG 的编码基因进行后续分析,最终羌活获得 51 条编码基因,宽叶羌活获得 50 条编码基因。

1.2 分析方法

1.2.1 密码子相关参数计算。使用软件 codonW1.4.2 分析获得的编码基因,计算有效密码子数 (effective number of codon, ENC)。ENC 范围一般在 20~61,越接近 20,表示密码子使用偏性越大。使用在线分析软件 CUSP (<http://embossgui.sourceforge.net/demo/>) 对叶绿体基因组密码子的 GC 含量等碱基组成信息进行研究,并对相关参数进行相关性分析^[19]。

1.2.2 中性绘图分析。分别以 C_{GC1} 、 C_{GC2} 、 C_{GC3} 表示密码子第 1、2、3 位的 GC 含量, C_{GC12} 表示 C_{GC1} 与 C_{GC2} 的平均数,以 C_{GC3} 和 C_{GC12} 为横、纵坐标绘制散点图,分析叶绿体基因组密码子第 3 位和前 2 位的相关性,从而分析影响密码子偏性的因素^[20]。以 C_{GC3} 、 C_{GC12} 为横、纵坐标绘制散点图,如果 C_{GC3} 和 C_{GC12} 呈显著相关,说明密码子第 3 位的碱基组成与前 2 位相似,密码子偏性主要受突变的影响;如果 C_{GC3} 和 C_{GC12} 无显著相关关系,趋势线斜率接近 0,则说明密码子第 3 位与前 2 位碱基组成差异较大,密码子碱基相对保守,这种情况密码子偏性更多受选择的影响^[21]。

1.2.3 PR2-plot 绘图分析。为进一步分析影响 2 种羌活叶绿体基因组密码子偏性的因素,以 C_{A3} 、 C_{T3} 、 C_{C3} 、 C_{G3} 分别表示密码子第 3 位 A、T、C、G 4 种碱基的含量,以 $C_{C3}/(C_{C3}+C_{G3})$ 为横坐标, $C_{A3}/(C_{A3}+C_{T3})$ 为纵坐标绘制散点图。如果大部分基因分布在图像上半部,说明密码子第 3 位碱基含量 A

>T,反之则 T>A,同理如果大部分基因分布在图像右半部,说明密码子第 3 位碱基含量 G>C,反之则 C>G。若点在图中均匀分布,则密码子偏性完全受突变影响^[20-21]。

1.2.4 ENC-plot 绘图分析。以 C_{GC3} 为横坐标,以 ENC 值为纵坐标绘制散点图,并以 $ENC_{期望}$ 绘制标准曲线。

$$ENC_{期望} = 2 + C_{GC3}^2 + \frac{29}{C_{GC3}^2 + (1 - C_{GC3})^2}$$

标准曲线表示在无选择压力的情况下 ENC 与 C_{GC3} 的关系,如果点主要分布于曲线附近,说明密码子偏性更多受突变的影响,如果点分布于距标准曲线较远的位置,则说明密码子偏性更多受选择的影响。根据公式 $(ENC_{期望} - ENC_{实际})/ENC_{期望}$ 制作 ENC 分布频数表,统计 $ENC_{实际}$ 与 $ENC_{期望}$ 的差距^[22]。

1.2.5 RSCU 分析。使用 CodonW 计算 2 种羌活叶绿体基因组的 RSCU (relative synonymous codon usage)。RSCU 是实际观察到的某密码子使用频率与预期 (该密码子编码的氨基酸的同义密码子使用频率相等时) 使用频率的比值。理论值与预期值相等时,密码子不存在偏性。所以当 $RSCU = 1.00$,表明密码子使用无偏性; $RSCU > 1.00$ 时,表明该密码子实际使用频率较高; $RSCU < 1.00$ 时,表明实际使用频率较低^[23]。

1.2.6 最优密码子分析。选出 2 种羌活叶绿体编码基因密码子 ENC 值最大的 10% 和最小的 10%,分别建立高低表达库,选出库中 $\Delta RSCU > 0.08$ 且整体 $RSCU > 1.00$ 的密码子^[24]。

2 结果与分析

2.1 密码子组成 2 种羌活叶绿体基因组密码子第 3 位的碱基组成结果显示,羌活的叶绿体基因组密码子第 3 位的 C_{A3} 、 C_{T3} 分别为 42.75% 与 46.97% (表 1),明显高于 C_{C3} 、 C_{G3} ,说明羌活叶绿体基因组密码子多以 AT 结尾;宽叶羌活的密码子第 3 位碱基组成与羌活类似,AT 含量较高。羌活与宽叶羌活的叶绿体密码子第 1 位 GC 含量从第 1 位到第 3 位逐渐下降,第 1 位分别为 46.16%、45.99%,第 2 位 38.17%、37.91%,第 3 位 29.97%、29.67%。2 种羌活的 ENC 均大于 49.00,说明 2 种羌活叶绿体基因组密码子偏性较弱。

表 1 2 种羌活叶绿体基因组密码子第 3 位碱基组成及 GC 含量

Table 1 GC contents and the third base of codon in chloroplast genes of *Notopterygium incisum* and *N. franchetii*

物种 Species	C_{T3} %	C_{C3} %	C_{A3} %	C_{G3} %	C_{GC1} %	C_{GC2} %	C_{GC3} %	C_{GC} %	ENC
羌活 <i>N. incisum</i>	46.97	17.05	42.75	18.26	46.16	38.17	29.97	38.10	49.79
宽叶羌活 <i>N. franchetii</i>	47.12	16.65	43.02	18.46	45.99	37.91	29.67	37.86	49.70

使用 CUSP 统计 2 种羌活叶绿体基因组编码基因的 C_{GC} 、 C_{GC1} 、 C_{GC2} 、 C_{GC3} 、ENC,结果表明,羌活和宽叶羌活的叶绿体基因组编码基因的 C_{GC} 平均值为 38.64% 和 38.44% (表 2),差异不大。羌活和宽叶羌活中 C_{GC} 最高的基因皆为 *rps11*, C_{GC} 最低的基因为 *yef1*。羌活和宽叶羌活编码基因的 ENC 值范围分别是 36.33~55.03、37.97~57.09,而且羌活约有 83% 编码基因的 ENC 值大于 45.00,宽叶羌活有 79% 编码基因的 ENC 值大于 45.00,说明多数编码基因密码子使用偏性较弱。2 种羌活的 *rps8* 的 ENC 值最小,分别为 36.33 和

37.97,表明 *rps8* 在 2 种羌活叶绿体编码基因中偏性最强。

对 2 种羌活的叶绿体编码基因的 C_{GC1} 、 C_{GC2} 、 C_{GC3} 、GC、ENC、叶绿体编码基因序列长度 (sequence length, SL) 等参数进行相关性分析,结果表明 2 种羌活的碱基组成相关性较接近, C_{GC1} 与 C_{GC2} 皆为极显著相关 (表 3),且 C_{GC1} 、 C_{GC2} 都与 C_{GC3} 无显著相关性;2 种羌活的 GC 含量与第 1、2、3 位组成均极显著相关。羌活与宽叶羌活的 SL 与 C_{GC1} 、 C_{GC2} 、 C_{GC3} 、GC、ENC 均无显著相关性。2 种羌活的相关性分析结果也存在一定差异,羌活的 ENC 与 C_{GC3} 极显著相关,而宽叶羌活的

ENC 除了与 C_{GC3} 极显著相关外,还与 GC 显著相关。

表 2 2 种无活叶绿体编码基因密码子 GC 含量

Table 2 GC contents of genes in *Notopterygium incisum* and *N. franchetii* chloroplast

基因 Gene	无活 <i>N. incisum</i>					宽叶无活 <i>N. franchetii</i>				
	$C_{GC1}/\%$	$C_{GC2}/\%$	$C_{GC3}/\%$	$C_{GC}/\%$	ENC	$C_{GC1}/\%$	$C_{GC2}/\%$	$C_{GC3}/\%$	$C_{GC}/\%$	ENC
<i>accD</i>	41.63	35.51	28.98	35.37	47.11	41.63	35.71	29.18	35.51	46.05
<i>atpA</i>	56.50	39.96	25.20	40.55	47.07	56.69	39.57	25.20	40.49	46.73
<i>atpB</i>	56.02	42.17	29.92	42.70	48.88	55.31	42.08	30.46	42.62	49.26
<i>atpE</i>	49.25	40.30	30.60	40.05	55.03	51.06	39.01	29.08	39.72	55.59
<i>atpF</i>	48.35	34.62	35.16	39.38	48.83	48.35	35.16	34.07	39.19	50.07
<i>atpI</i>	49.19	37.50	28.23	38.31	45.61	49.19	37.50	28.63	38.44	47.68
<i>ccsA</i>	34.26	37.04	27.47	32.92	50.90	34.77	36.00	26.15	32.31	46.33
<i>cemA</i>	40.43	28.26	30.43	33.04	45.54	38.82	27.06	28.24	31.37	50.72
<i>clpP</i>	57.58	37.37	30.81	41.92	49.83	57.58	36.87	32.32	42.26	53.14
<i>matK</i>	41.06	33.79	30.06	34.97	52.07	40.91	33.60	29.25	34.58	51.94
<i>ndhA</i>	43.68	39.01	22.80	35.16	42.78	43.41	39.56	22.25	35.07	42.74
<i>ndhB</i>	42.19	39.55	31.24	37.66	48.89	41.68	38.94	31.12	37.25	48.44
<i>ndhC</i>	44.63	32.23	25.62	34.16	46.79	47.93	33.06	25.62	35.54	47.85
<i>ndhE</i>	40.59	34.65	21.78	32.34	43.03	41.18	34.31	20.59	32.03	38.18
<i>ndhF</i>	36.39	35.44	24.53	32.12	43.77	36.33	35.12	23.99	31.81	43.43
<i>ndhG</i>	45.20	35.03	28.81	36.35	47.96	46.33	35.03	25.99	35.78	45.32
<i>ndhH</i>	51.02	36.80	27.41	38.41	47.75	51.02	36.55	27.16	38.24	47.39
<i>ndhI</i>	39.88	37.50	25.60	34.33	46.65	37.39	34.87	28.99	33.75	43.67
<i>ndhJ</i>	49.69	36.48	30.19	38.78	50.77	49.69	36.48	30.19	38.78	44.93
<i>ndhK</i>	44.69	44.25	26.99	38.64	50.97	44.25	44.69	26.99	38.64	49.85
<i>petA</i>	52.96	36.76	32.40	40.71	53.75	52.65	37.38	31.15	40.39	52.74
<i>petB</i>	48.61	41.67	33.33	41.20	46.04	49.07	41.67	29.17	39.97	42.71
<i>petD</i>	51.55	39.13	24.84	38.51	39.38	51.55	39.13	22.98	37.89	39.13
<i>psaA</i>	52.60	43.41	31.69	42.57	49.05	52.60	43.41	33.02	43.01	50.33
<i>psaB</i>	48.98	43.27	31.70	41.32	47.46	48.84	43.27	32.52	41.54	47.27
<i>psbA</i>	50.00	43.50	33.33	42.28	39.12	—	—	—	—	—
<i>psbB</i>	55.21	46.56	31.83	44.53	49.49	54.81	46.76	31.24	44.27	48.34
<i>psbC</i>	53.16	45.99	32.91	44.02	45.40	53.16	45.99	31.43	43.53	44.75
<i>psbD</i>	52.26	43.22	32.49	42.66	45.25	52.26	43.22	33.05	42.84	46.44
<i>rbcL</i>	58.19	43.07	32.14	44.47	49.37	58.40	43.07	32.14	44.54	49.49
<i>rpl14</i>	56.10	38.21	30.89	41.73	46.69	55.28	38.21	31.71	41.73	45.32
<i>rp116</i>	50.00	52.94	22.79	41.91	37.72	—	—	—	—	—
<i>rpl2</i>	—	—	—	—	—	50.55	49.09	33.09	44.24	54.12
<i>rpl20</i>	38.76	41.09	24.81	34.88	49.26	38.76	41.09	25.58	35.14	48.83
<i>rpl22</i>	40.00	35.76	29.09	34.95	49.12	41.21	35.76	28.48	35.15	47.75
<i>rpoA</i>	46.33	32.49	28.81	35.88	51.99	46.33	32.77	28.81	35.97	52.87
<i>rpoB</i>	50.89	38.10	29.04	39.34	49.10	50.23	37.63	28.01	38.62	47.91
<i>rpoC1</i>	50.29	37.61	27.70	38.53	49.77	49.56	37.76	29.01	38.78	50.47
<i>rpoC2</i>	45.94	36.76	29.86	37.52	50.52	45.72	36.38	29.48	37.19	50.35
<i>rps11</i>	54.68	55.40	27.34	45.80	54.05	53.24	56.12	28.78	46.04	52.97
<i>rps12</i>	52.42	48.39	27.42	42.74	46.99	52.42	48.39	27.42	42.74	46.99
<i>rps14</i>	44.55	47.52	37.62	43.23	50.28	45.54	48.51	35.64	43.23	49.95
<i>rps18</i>	38.24	43.14	29.41	36.93	49.25	37.25	43.14	30.39	36.93	49.40
<i>rps2</i>	42.62	41.77	29.54	37.97	49.32	43.46	40.93	29.11	37.83	49.41
<i>rps3</i>	45.41	33.49	26.15	35.02	49.86	46.58	33.79	24.20	34.86	50.55
<i>rps4</i>	50.99	38.12	31.19	40.10	54.15	51.49	38.12	30.20	39.93	53.72
<i>rps7</i>	53.21	45.51	22.44	40.38	44.46	51.92	44.23	23.08	39.74	47.25
<i>rps8</i>	41.48	40.00	24.44	35.31	36.33	40.00	37.78	24.44	34.07	37.97
<i>yef1</i>	34.43	30.32	27.59	30.78	48.08	34.43	29.82	26.30	30.18	47.84
<i>yef2</i>	41.84	35.18	36.83	37.95	52.96	41.95	35.09	36.46	37.83	52.76
<i>yef3</i>	48.52	38.46	28.99	38.66	49.49	47.93	38.46	31.95	39.45	54.88
<i>yef4</i>	44.32	42.16	37.84	41.44	54.29	44.32	41.62	36.76	40.90	57.09

表 3 2 种羌活编码基因密码子相关参数的相关性分析

Table 3 Correlation coefficient between the parameters of genes in *Notopterygium incisum* and *N. franchetii*

羌活 <i>N. incisum</i>	宽叶羌活 <i>N. franchetii</i>					
	C_{GC1}	C_{GC2}	C_{GC3}	GC	ENC	SL
C_{GC1}	1.000	0.459**	0.266	0.835**	0.192	-0.178
C_{GC2}	0.451**	1.000	0.276	0.795**	0.124	-0.242
C_{GC3}	0.186	0.055	1.000	0.584**	0.601**	0.179
GC	0.843**	0.756**	0.471**	1.000	0.349*	-0.152
ENC	0.035	-0.105	0.477**	0.131	1.000	0.106
SL	-0.173	-0.250	0.216	-0.145	0.158	1.000

注: * 表示在 0.05 水平显著相关; ** 表示在 0.01 水平显著相关

Note: * indicates significant correlation at 0.05 level; ** indicates significant correlation at 0.01 level

2.2 中性绘图分析 羌活的 C_{GC12} 取值范围为 0.324~0.550 (图 1), C_{GC3} 与 C_{GC12} 的回归系数为 0.218 7, 皮尔森系数为 0.147, 无显著相关性, 说明羌活的叶绿体密码子第 3 位碱基组成与第 1、2 位差异较大, 密码子偏性主要受选择的影响。宽叶羌活的 C_{GC12} 的取值范围 0.321~0.547 (图 1), C_{GC3} 与 C_{GC12} 的回归系数为 0.431 7, 皮尔森系数为 0.316, 相关性较低, 说明宽叶羌活的叶绿体密码子第 3 位碱基组成与第 1、2 位差异较大, 密码子偏性受选择影响较大。

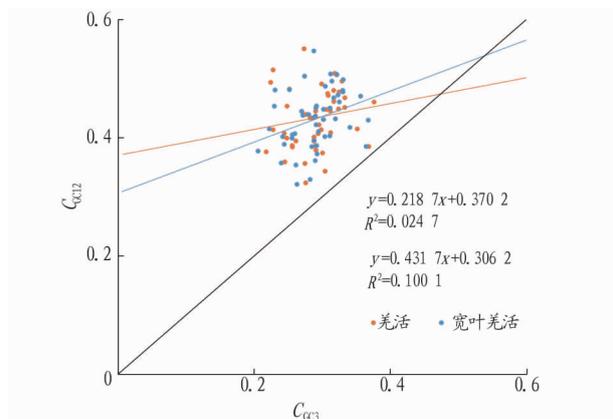


图 1 中性绘图分析

Fig. 1 Neutrality plot analysis

2.3 PR2-plot 分析 PR2-plot 分析以 $C_{G3}/(C_{G3}+C_{T3})$ 为横坐标, $C_{A3}/(C_{A3}+C_{T3})$ 为纵坐标绘制散点图, 结果表明, 2 种羌活大部分基因分布在右下方位置 (图 2), 说明羌活和宽叶羌活的叶绿体基因组密码子不仅受到突变的影响, 还受到选择的影响, 并且密码子第 3 位的碱基频率 $T>A, G>C$ 。

2.4 ENC-plot 分析 对 2 种羌活进行 ENC-plot 绘图分析, 结果见图 3。标准曲线表示在无选择压力的情况下, ENC 与 C_{GC3} 的关系。如果点主要分布于曲线附近, 说明密码子偏性更多受突变的影响, 如果点分布于距标准曲线较远的位置, 说明密码子偏性更多受选择的影响。2 种羌活的编码基因在标准曲线两侧均有基因分布, 大部分基因位于标准曲线下侧, 部分基因距标准曲线较远。

为了进一步分析 ENC 实际值与标准曲线的偏离情况, 制作 ENC 分布频数表, 结果见表 4, 羌活有 24 个基因的比值在 $-0.05 \sim <0.05$, 表明有 24 个基因的 ENC 实际值与期望值接近 (ENC 比值 $-0.05 \sim <0.05$), 其余 27 个基因的 ENC 实际值与期望值相差较大 (ENC 比值 >0.05 或 <-0.05), 其中 *ps-*

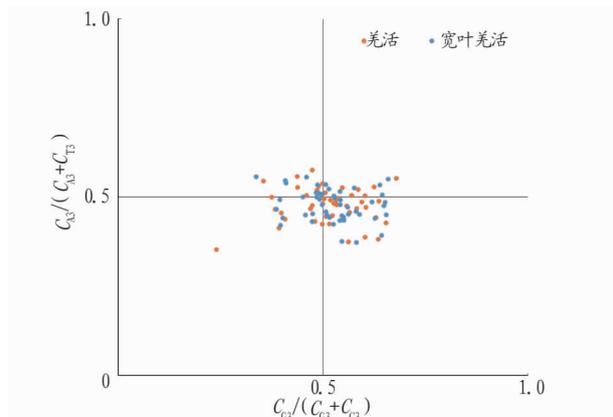


图 2 PR2-plot 分析

Fig. 2 Analysis of PR2-plot

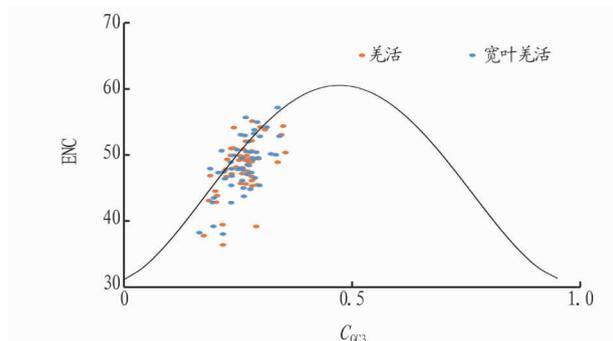


图 3 ENC-plot 分析

Fig. 3 Analysis of ENC-plot

ba, *rps8*, *rp116*, *petD*, *psbC*, *psbD*, *petB* 的 ENC 实际值距标准曲线较远 (ENC 比值 >0.15)。宽叶羌活有 23 个基因的比值在 $-0.05 \sim <0.05$, ENC 实际值与期望值接近, 27 个基因 ENC 实际值与期望值差距较大, 其中 *rps8*, *petB*, *petD*, *psbC*, *ndhE*, *ndhI* 的 ENC 实际值距标准曲线较远。说明 2 种羌活的叶绿体基因组密码子偏性受突变的影响较小, 更多受选择的影响。

2.5 RSCU 分析 由表 5 可知, 羌活和宽叶羌活各有 30 个密码子 $RSCU>1.00$, 说明各有 30 个密码子的使用有偏好性。实际使用频率偏高 ($RSCU>1.00$) 的密码子中结尾为 U 的有 16 个, 结尾为 A 的有 13 个, 只有 UUG 以 G 结尾。说明 2 种羌活叶绿体基因组密码子偏好以 A、U 结尾, 其中 U 的偏好性略强于 A。

2.6 最优密码子确定 2 种羌活的最优密码子均符合总

RSCU>1.00, Δ RSCU \geq 0.08。2 种羌活的最优密码子都多以 A/U 结尾,只有 UUG 以 G 结尾。确认了羌活的最优密码子 9 个(UUG、CUU、ACU、UAU、CAU、AAU、CGU、AGU、GGA),

宽叶羌活的最优密码子 7 个(UUG、CCA、CAU、CGU、AGU、AGA、GGA)。2 种羌活有 5 个最优密码子相同,分别是 UUG、CAU、CGU、AGU、GGA(表 6)。

表 4 ENC 比值频数分布
Table 4 Distribution of ENC ration

组限 Class boundary	组中值 Class mid-value	羌活 <i>N. incisum</i>		宽叶羌活 <i>N. francheti</i>	
		组数 Number	组频 Frequency	组数 Number	组频 Frequency
-0.15~<-0.05	-0.1	1	0.02	1	0.06
-0.05~<0.05	0	24	0.47	23	0.44
0.05~<0.15	0.1	19	0.45	20	0.46
0.15~<0.25	0.2	6	0.04	6	0.04
0.25~<0.35	0.3	1	0.02	0	0
总计 Total		51	1.00	50	1

表 5 2 种羌活的同义密码子使用度
Table 5 RSCU analysis of *Notopterygium incisum* and *N. francheti*

序号 No.	羌活 <i>N. incisum</i>				宽叶羌活 <i>N. francheti</i>			
	氨基酸 AA	密码子 Codon	数量 Number	RSCU	氨基酸 AA	密码子 Codon	数量 Number	RSCU
1	Phe	UUU	763	1.33	Phe	UUU	777	1.36
		UUC	384	0.67		UUC	368	0.64
2	Leu	UUA	693	1.94	Leu	UUA	697	1.96
		UUG	437	1.23		UUG	431	1.21
		CUU	456	1.28		CUU	450	1.27
		CUC	137	0.38		CUC	136	0.38
		CUA	284	0.80		CUA	279	0.79
3	Ile	CUG	131	0.37	Ile	CUG	138	0.39
		AUU	819	1.43		AUU	842	1.47
		AUC	347	0.61		AUC	316	0.55
		AUA	554	0.97		AUA	564	0.98
4	Met	AUG	462	1.00	Met	AUG	451	1.00
		5	Val	GUU		420	1.51	Val
GUC	127			0.46	GUC	132	0.47	
GUA	401			1.44	GUA	402	1.44	
GUG	168			0.60	GUG	170	0.61	
UCU	425			1.68	6	Ser	UCU	
UCC	245	0.97	UCC	238			0.95	
UCA	296	1.17	UCA	288			1.15	
UCG	149	0.59	UCG	164			0.66	
7	Pro	CCU	345	1.62	Pro	CCU	336	1.57
		CCC	153	0.72		CCC	160	0.75
		CCA	227	1.07		CCA	226	1.06
		CCG	127	0.60		CCG	133	0.62
8	Thr	ACU	431	1.63	Thr	ACU	435	1.66
		ACC	195	0.74		ACC	186	0.71
		ACA	311	1.18		ACA	311	1.19
		ACG	121	0.46		ACG	114	0.44
9	Ala	GCU	510	1.79	Ala	GCU	485	1.77
		GCC	187	0.66		GCC	175	0.64
		GCA	315	1.11		GCA	303	1.11
		GCG	126	0.44		GCG	130	0.48
10	Tyr	UAU	629	1.61	Tyr	UAU	629	1.61
		UAC	152	0.39		UAC	153	0.39
11	TER	UAA	27	1.59	TER	UAA	30	1.58
		UAG	11	0.65		UAG	15	0.79
12	His	CAU	367	1.49	His	CAU	374	1.52
		CAC	127	0.51		CAC	118	0.48
13	Gln	CAA	570	1.51	Gln	CAA	560	1.50
		CAG	186	0.49		CAG	185	0.50

接下表

续表 5

序号 No.	羌活 <i>N. incisum</i>				宽叶羌活 <i>N. franchetii</i>			
	氨基酸 AA	密码子 Codon	数量 Number	RSCU	氨基酸 AA	密码子 Codon	数量 Number	RSCU
14	Asn	AAU	732	1.50	Asn	AAU	744	1.52
		AAC	241	0.50		AAC	234	0.48
15	Lys	AAA	804	1.51	Lys	AAA	827	1.51
		AAG	263	0.49		AAG	268	0.49
16	Asp	GAU	669	1.62	Asp	GAU	669	1.60
		GAC	156	0.38		GAC	166	0.40
17	Glu	GAA	818	1.51	Glu	GAA	803	1.50
		GAG	269	0.49		GAG	269	0.50
18	Cys	UGU	165	1.54	Cys	UGU	164	1.54
		UGC	49	0.46		UGC	49	0.46
19	TER	UGA	13	0.76	TER	UGA	12	0.63
20	Trp	UGG	369	1.00	Trp	UGG	358	1.00
21	Arg	CGU	266	1.32	Arg	CGU	263	1.32
		CGC	75	0.37		CGC	76	0.38
22	Ser	CGA	277	1.38	Ser	CGA	275	1.38
		CGG	101	0.50		CGG	94	0.47
23	Arg	AGU	318	1.26	Arg	AGU	312	1.25
		AGC	83	0.33		AGC	82	0.33
24	Gly	AGA	366	1.82	Gly	AGA	366	1.83
		AGG	120	0.60		AGG	124	0.62
24	Gly	GGU	479	1.35	Gly	GGU	458	1.31
		GGC	171	0.48		GGC	164	0.47
24	Gly	GGA	525	1.48	Gly	GGA	529	1.51
		GGG	240	0.68		GGG	247	0.71

表 6 2 种羌活的最优密码子

Table 6 Optimal codons in chloroplast genome of *Notopterygium incisum* and *N. franchetii*

序号 No.	羌活 <i>N. incisum</i>		宽叶羌活 <i>N. franchetii</i>	
	密码子 Codon	氨基酸 Amino acid	密码子 Codon	氨基酸 Amino acid
1	UUG***	Leu	UUG***	Leu
2	CUU*	Leu	CCA*	Pro
3	ACU*	Thr	CAU*	His
4	UAU*	Tyr	CGU*	Arg
5	CAU***	His	AGU*	Ser
6	AAU**	Asn	AGA*	Arg
7	CGU*	Arg	GGA*	Gly
8	AGU*	Ser		
9	GGA**	Gly		

注: * 表示 $\Delta RSCU \geq 0.2$; ** 表示 $\Delta RSCU \geq 0.3$; *** 表示 $\Delta RSCU \geq 0.5$

Note: * indicates $\Delta RSCU \geq 0.2$; ** indicates $\Delta RSCU \geq 0.3$; *** indicates $\Delta RSCU \geq 0.5$

3 讨论

该研究对羌活和宽叶羌活基因组密码子使用偏好性及其主要影响因素进行了探讨和比较。结果表明, ENC 小于 45 的基因较少, 多数基因 ENC 较高, 说明 2 种羌活基因组密码子使用偏性均较弱。相关性分析结果表明, 2 种羌活的叶绿体基因组 ENC 都与 C_{GC3} 显著相关, 说明密码子第 3 位碱基组成对密码子偏性的影响较大, 这与原晓龙等^[25]对云南蓝果树的研究结论一致。而尹为治等^[26]研究表明, 坡垒的叶绿体基因组密码子第 1 位碱基组成对密码子偏性的影响更大。密码子组成分析显示, 2 种羌活的总 GC 和 C_{GC3} 较低, 且 $C_{GC1} > C_{GC2} > C_{GC3}$, 密码子的第 3 位倾向于使用 A/T, 第 3 位碱基使用频率 $T > A, G > C$ 。

突变和选择压力是 2 个影响密码子偏性的主要因素。PR2-plot 分析结果发现, 基因分布在图中发生了偏移, 可知 2 种羌活的叶绿体密码子使用偏性不仅受到了突变的影响, 还受到了选择压力的影响。中性绘图结果显示, 2 种羌活 C_{GC12}

与 C_{GC3} 相关性不显著, 说明密码子第 3 位的碱基组成与前两位差异大, 密码子相对保守, 表明选择压力对密码子的影响较大。ENC-plot 结果显示, 2 种羌活的多数叶绿体基因偏离标准曲线进一步证实选择压力是影响 2 种羌活叶绿体密码子使用偏性的主要因素。综上所述, 2 种羌活的叶绿体基因组密码子使用偏性受到突变和选择压力的影响, 最主要的影响因素是选择压力。影响竹亚科的巴山木竹、印度簕竹、*Olyra latifolia* L. 以及苦荞的叶绿体基因组密码子使用偏性的主要因素也是选择压力^[27-28]。而突变对榭蕨属的川滇榭蕨、栎叶榭蕨的叶绿体基因组密码子使用偏性的影响更大^[29]。

羌活和宽叶羌活的叶绿体基因组密码子使用模式较为相似。2 种羌活的总 GC 含量与密码子各位的 GC 含量差异不大; 叶绿体基因组密码子偏性都较弱; 密码子第 3 位都倾向于使用 A/T, 且碱基含量 $T > A, G > C$; 影响密码子使用偏性的主要因素都是选择压力。

通过 RSCU 分析, 确定了 2 种羌活中使用频率较高的 30 个高频密码子 ($RSCU > 1.00$), 几乎都以 A 或 U 结尾, 只有 UUG 以 G 结尾。确认了羌活的最优密码子 9 个 (UUG、CUU、ACU、UAU、CAU、AAU、CGU、AGU、GGA), 宽叶羌活的最优密码子 7 个 (UUG、CCA、CAU、CGU、AGU、AGA、GGA)。2 种羌活有 5 个最优密码子相同, 分别是 UUG、CAU、CGU、AGU、GGA。

该研究通过对 2 种羌活药用植物的叶绿体基因组编码基因的使用偏性进行比较分析, 得出主要结论: ① 2 种药用羌活叶绿体基因组的密码子使用偏性整体较为相似, 且偏性较弱; ② 密码子的第 3 位倾向于使用 A、T; ③ 突变和选择压力在 2 种羌活的密码子使用偏性中起主要作用, 其中选择压力对密码子使用偏性影响更大。该研究结果为后续药用羌活植物的基因表达、蛋白质翻译和遗传育种研究提供了研究依

据和基础数据,有潜在的经济价值和应用价值。

参考文献

- [1] MARAIS G, MOUCHIROUD D, DURET L. Neutral effect of recombination on base composition in *Drosophila* [J]. *Genetical research*, 2003, 81(2): 79-87.
- [2] PRABHA R, SINGH D P, SINHA S, et al. Genome-wide comparative analysis of codon usage bias and codon context patterns among cyanobacterial genomes [J]. *Marine genomics*, 2017, 32: 31-39.
- [3] VICARIO S, MORIYAMA E N, POWELL J R. Codon usage in twelve species of *Drosophila* [J]. *BMC evolutionary biology*, 2007, 7: 1-17.
- [4] SHARP P M, LI W H. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias [J]. *Molecular biology and evolution*, 1987, 4(3): 222-230.
- [5] OLEJNICZAK M, UHLENBECK O C. tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition [J]. *Biochimie*, 2006, 88(8): 943-950.
- [6] ZALUCKI Y M, POWER P M, JENNINGS M P. Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins [J]. *Nucleic acids research*, 2007, 35(17): 5748-5754.
- [7] HUANG Y, KOONIN E V, LIPMAN D J, et al. Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage [J]. *Nucleic acids research*, 2009, 37(20): 6799-6810.
- [8] 孙铮, 马亮, MURPHY R, 等. *Wolbachia pipientis wMel* 基因组水平上的密码子使用分析 [J]. *中国科学(C辑: 生命科学)*, 2009, 39(10): 948-953.
- [9] HIRAOKA Y, KAWAMATA K, HARAGUCHI T, et al. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe* [J]. *Genes to cells*, 2009, 14(4): 499-509.
- [10] SUBRAMANIAN S. Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes [J]. *Genetics*, 2008, 178(4): 2429-2432.
- [11] 黄琼林. 高良姜叶绿体基因组测序与特征分析 [J]. *热带作物学报*, 2021, 42(1): 1-6.
- [12] 原晓龙, 李云琴, 张劲峰, 等. 乳油木叶绿体基因组密码子偏好性分析 [J]. *分子植物育种*, 2020, 18(17): 5658-5664.
- [13] 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.
- [14] 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.
- [15] 刘文武, 蒋晓文, 张帅, 等. 羌活中香豆素类化学成分及其抗氧化活性研究 [J]. *中草药*, 2019, 50(6): 1310-1315.
- [16] 杨娇. 羌活属叶绿体基因组进化及物种界定研究 [D]. 西安: 西北大学, 2018: 5-6.
- [17] 阎梦颖. 珍稀濒危羌活属药用植物的谱系地理学研究 [D]. 西安: 西北大学, 2016: 5-6.
- [18] 郭培, 郎拥军, 张国桃. 羌活化学成分及药理活性研究进展 [J]. *中成药*, 2019, 41(10): 2445-2459.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] WRIGHT F. The 'effective number of codons' used in a gene [J]. *Gene*, 1990, 87(1): 23-29.
- [23] 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 Res*, 1987, 15(3): 1281-1295.
- [24] 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.
- [25] 原晓龙, 康洪梅, 王毅. 云南蓝果树叶叶绿体基因组密码子偏好性分析 [J]. *西北林学院学报*, 2020, 35(4): 26-31, 124.
- [26] 尹为治, 方正, 黄良鸿, 等. 海南 2 种龙脑香科植物叶绿体基因组密码子偏好性分析 [J]. *林业调查规划*, 2020, 45(6): 19-26.
- [27] 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.
- [28] 胡莎莎, 罗洪, 吴琦, 等. 苦荞叶绿体基因组密码子偏爱性分析 [J]. *分子植物育种*, 2016, 14(2): 309-317.
- [29] 沈宗芳, 陆添权, 张志荣, 等. 榭蕨属叶绿体基因组密码子偏好性分析 [J]. *广西植物*, 2021, 41(2): 266-273.
- [30] THOMMA B P H J, VAN ESSE H P, CROUS P W, et al. *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae [J]. *Mol Plant Pathol*, 2005, 6(4): 379-393.
- [31] 李铁强, 吴永英, 范洪臣, 等. 酵母双杂交系统研究进展 [J]. *生物信息学*, 2008(1): 46-48.
- [32] 赵光耀. 小麦全长 cDNA 文库构建、测序与分析 [D]. 北京: 中国农业科学院, 2006.
- [33] 谢卡斌. 水稻全长 cDNA 文库的构建和两个 microRNA 的功能研究 [D]. 武汉: 华中农业大学, 2008.
- [34] 赵怀印, 罗碧峰, 史洁玮, 等. 成熟期番茄果实 cDNA 文库的构建及 SIHSP17.7 互作蛋白的筛选 [J]. *分子植物育种*, 2020, 18(2): 374-380.
- [35] 韦淑亚. 甘蓝型油菜菌核病抗性相关基因 cDNA 文库构建及评价 [D]. 武汉: 华中农业大学, 2005.
- [36] 姜赞, 史自航, 付欣, 等. 番茄花柄脱落过程的酵母双杂交 cDNA 文库构建 [J]. *沈阳农业大学学报*, 2018, 49(2): 129-135.
- [37] 李颖波, 宗营杰, 刘成洪, 等. 盐胁迫下大麦酵母双杂交文库的构建与鉴定 [J]. *上海农业学报*, 2019, 35(6): 1-5.
- [38] 刘露露, 曲俊杰, 郭泽西, 等. 霜霉菌感染后葡萄叶片酵母双杂交 cDNA 文库构建 [J]. *南方农业学报*, 2020, 51(4): 829-835.
- [39] 苏玲, 李彬, 王青, 等. 金柑花蕾酵母双杂交 cDNA 文库构建及评价 [J]. *基因组学与应用生物学*, 2019, 38(7): 3169-3173.
- [40] 杨莎, 李燕, 郭峰, 等. 利用酵母双杂交系统筛选花生 AhCaM 相互作用蛋白 [J]. *作物学报*, 2015, 41(7): 1056-1063.
- [41] 钱瑜. 不结球白菜 BcRISPI 的功能分析 [D]. 南京: 南京农业大学, 2014.
- [42] TAO K E, DONG C H, MAO H, et al. Construction of a normalized full-length cDNA library of sesame developing seed by DSN and SMART™ [J]. *Agric Sci China*, 2011, 10(7): 1004-1009.
- [43] 朱佳慧, 徐秋芳, 袁平平, 等. 水稻幼苗酵母双杂交 cDNA 文库的构建及鉴定 [J]. *江苏农业科学*, 2018, 46(9): 47-50.
- [44] 胡斌林, 雷雷, 王效维, 等. 甘蓝花蕾酵母双杂交 cDNA 文库构建及评价 [J]. *西北农业学报*, 2015, 24(8): 145-152.

(上接第 94 页)

母双杂交 cDNA 文库时被广泛应用。因此,该试验以番茄不同组织为样本,提取总 RNA 后用 SMART 技术合成全长 cDNA 构建文库。

构建具有高质量的酵母双杂交 cDNA 文库为后续筛选互作蛋白奠定基础^[19]。通常从两个方面来评价构建的 cDNA 文库质量^[20]: ①用文库容量来衡量 cDNA 文库的代表性,文库滴度高于 1×10^6 CFU/mL 时满足要求; ②用插入片段大小和重组率描述重组序列完整性^[11]。在该试验中,文库容量为 2.31×10^7 CFU,文库滴度为 2.5×10^8 CFU/mL,表明文库中 cDNA 的种类丰富,具有较高的完整性,满足建库标准。文库插入片段大小 200~2 000 bp,重组率为 97%,证明 cDNA 序列较完整,可用于下一步筛选与转录因子 MYB102 相互作用的蛋白。

参考文献

- [1] 王孝宜, 李树德, 东惠茹, 等. 番茄品种耐寒性与 ABA 和可溶性糖含量的关系 [J]. *园艺学报*, 1998, 25(1): 56-60.
- [2] 郑东虎, 黄俊轩, 王丽娟. 番茄低温生态学的研究进展 [J]. *北方园艺*, 2002(3): 36-37.
- [3] 刘强, 张贵友, 陈受宜. 植物转录因子的结构与调控作用 [J]. *科学通报*, 2000, 45(14): 1465-1474.
- [4] OGATA K, KANEI-ISHII C, SASAKI M, et al. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation [J]. *Nat Struct Mol Biol*, 1996, 3: 178-187.
- [5] 陈修花. 番茄转录因子 MYB80 基因的克隆及功能分析 [D]. 北京: 北京农学院, 2018.