

黄芩属药用植物叶绿体基因组密码子使用偏好性分析

荣正普, 王杰敏, 裴林, 石卫红, 张叶, 王晓媛, 孙会改*, 韩晓伟*

1. 河北省中药炮制技术创新中心, 河北 石家庄 050090

2. 河北中医药大学, 河北 石家庄 050200

摘要: **目的** 为探究黄芩属药用植物叶绿体基因组密码子的使用偏好性及其影响因素, 为黄芩属药用植物遗传育种、资源保护等研究提供理论依据。**方法** 从美国国家生物技术信息中心 (NCBI) 数据库中下载了 16 种黄芩属药用植物的叶绿体基因组, 使用 Codon W 1.4.2 及 CUSP 和 SPSS 27 等软件对 16 种黄芩属药用植物叶绿体基因的编码区 (CDS) 序列密码子偏好性进行分析。**结果** 16 种黄芩属药用植物的叶绿体基因组编码区的同义密码子和最优密码子使用偏好性一致, 都是以 A/U 结尾, 高频密码子有 UUA、AGA。PR2-plot、中性图和 ENC-plot 的结果显示 16 种黄芩属药用植物叶绿体基因组的密码子使用偏好性同时受到自然选择和内部碱基突变的影响, 但自然选择是叶绿体基因组密码子偏好性形成的主导因素。**结论** 16 种黄芩属药用植物的叶绿体基因组密码子偏好 A/U 结尾, 与其他因素相比自然选择是其密码子偏好形成的主导因素。

关键词: 黄芩属; 药用植物; 叶绿体基因组; 自然选择; 密码子使用偏好性

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2025)01-0269-13

DOI: 10.7501/j.issn.0253-2670.2025.01.026

Analysis on codon usage bias of chloroplast genomes in medicinal plants from genus *Scutellaria*

RONG Zhengpu, WANG Jiemin, PEI Lin, SHI Weihong, ZHANG Ye, WANG Xiaoyuan, SUN Huigai, HAN Xiaowei

1. Innovation Center of Hebei Province Traditional Chinese Medicine Processing Technology, Shijiazhuang 050090, China

2. Hebei University of Chinese Medicine, Shijiazhuang 050200, China

Abstract: Objective To explore the codon usage bias of chloroplast genomes of medicinal plants in the genus *Scutellaria* and its affecting factors, provide a theoretical basis for subsequent genetic breeding and resources protection studies of *Scutellaria* medicinal plants. **Methods** The chloroplast genomes of 16 *Scutellaria* medicinal plants were downloaded from the NCBI database, and the coding sequences (CDS) codon preferences of the chloroplast genes of the 16 *Scutellaria* medicinal plants were analyzed using Codon W 1.4.2 and the software CUSP and SPSS 27. **Results** The synonymous codons and optimal codons of chloroplast genome coding region of 16 *Scutellaria* medicinal plants in usage bias were consistent, all ending in A/U, and the high-frequency codons were UUA and AGA. The results of PR2-plot, neutral map, and ENC-plot showed that the codon usage bias in the chloroplast genomes of the 16 *Scutellaria* medicinal plants were affected by both natural selection and internal base mutations, but natural selection was the dominating factor in the formation of codon bias in the chloroplast genomes. **Conclusion** The chloroplast genomes of 16 *Scutellaria* medicinal plants have codon preferences for A/U endings, and natural selection is the dominant factor in the formation of their codon preferences compared with other factors.

Key words: *Scutellaria* Linn.; medicinal plants; chloroplast genome; natural selection; codon usage bias

黄芩属 *Scutellaria* Linn. 是被子植物门 唇形科 (Labiatae) 下的一个属, 该属物种有 300 多种, 大多为多年生或 1 年生草本, 少数物种为半灌木, 极少数物种为灌木。黄芩属植

收稿日期: 2024-07-08

基金项目: 国家重点研发计划项目 (2017YFC1701701); 国家中医药管理局科技司“中药质量保障与创新项目”(2024002845); 河北省省级科技计划资助(H2024423018, 22326418D); 河北省二期现代农业产业技术体系创新团队(HBCT2023080201, HBCT2023080205); 河北中医药大学大学生创新项目 (s202414432021, s202414432050); 河北省中医药管理局科研计划项目 (2023351)

作者简介: 荣正普 (2000—), 硕士研究生, 研究方向为中药资源与开发。E-mail: 18573787295@163.com

*通信作者: 韩晓伟 (1978—), 博士, 副教授, 硕士生导师, 研究方向为中药资源与开发。

Tel: (0311)89926415 E-mail: hanxiaowei2015@126.com

物, 广范分布, 中国有黄芩属植物 102 种 50 变种^[1]。黄芩属植物多可入药, 普遍含有黄酮、二萜、苯乙醇苷类等多种化学成分, 具有抗肿瘤、抗菌、抗氧化和保护心血管系统等药理作用^[2], 且部分黄芩属药用植物极具观赏性, 可供观赏, 有很好的药用和经济价值。

叶绿体有独立于细胞核的基因组是绿色植物能量转换与光合作用的重要场所^[3]。叶绿体基因组有着母体遗传的特征, 具有自身的基因组结构和蛋白质组成体系, 其遗传性特征独立于核基因组。与线粒体基因组的分子内重组和核基因组的复杂性相比, 其基因组大小适中, 结构整齐, 可以稳定遗传^[4]。这些性质有利于对叶绿体基因组的分析。目前叶绿体基因组已在分子进化、生物多样性分析、DNA 条形码、系统发育和遗传表达等研究领域被广泛使用。

密码子使用偏好性 (codon usage bias, CUB) 是指编码相同氨基酸的同义密码子不均衡使用的现象^[5]。密码子的使用偏好性广泛存在于自然界生物中, 且受自然选择和突变压力等多种因素的影响^[6-7]。分析密码子的使用偏好性不仅可以探讨基因的进化规律和基因表达过程中的调控机制^[8], 还可以预测外源基因的最佳宿主, 提高外源基因在宿主中的表达。同时基因的表达水平可以通过密码子优化来评估^[9], 为基因工程提供理论依据。随着现代测序技术的发展药用植物的基因领域研究也逐步展开, 其叶绿体基因组密码子偏好性逐渐引起科学界关注, 如百合属^[10]、川贝母^[11]等的叶绿体密码子偏好分析研究已见文献报道。黄芩属药用植物的生物活性成份的药理作用已被人广泛研究, 但目前有关黄芩属药用植物的叶绿体基因组密码子偏好性分析尚未见研究结果。本研究通过对中国 16 种黄芩属药用植物的叶绿体基因组的密码子使用偏好性和影响密码子使用偏好性的各种因素进行分析, 为黄芩属药用植物的分子育种、外源高效表达、种质资源保护以及性状改良等方面提供科学参考。

1 材料

从美国国家生物技术信息中心 (NCBI) 数据库中下载 16 种黄芩属药用植物的叶绿体基因组, 其基因组信息见表 1, 选择叶绿体基因组的蛋白质编码序列 (CDS)。每个编码序列不重复并且长度不小于 300 bp 用于后续分析, 并确保在起始处有正确

表 1 16 种黄芩属药用植物叶绿体基因组信息

Table 1 Information on chloroplast genomes of 16 *Scutellaria medicinal plants*

物种名	长度/bp	登录号
四裂花黄芩 <i>S. quadrilobulata</i> Y. Z. Sun ex C. H. Hu	152 066	MN128381.1
深裂叶黄芩 <i>S. przewalskii</i> Juz	151 675	MN128382.1
直萼黄芩 <i>S. orthocalyx</i> Hand.-Mazz	152 071	MN128383.1
毛叶黄芩 <i>S. mollifolia</i> C. Y. Wu & H. W. Li	152 417	MN128384.1
囊距黄芩 <i>S. calcarata</i> C. Y. Wu & H. W. Li	152 033	MN128385.1
滇黄芩 <i>S. amoena</i> C. H. Wright	151 833	MN128386.1
阿尔泰黄芩 <i>S. altaica</i> Fisch. ex Sweet	151 779	MN128387.1
藏黄芩 <i>S. kingiana</i> Prain	152 395	MN128389.1
并头黄芩 <i>S. scordifolia</i> Fisch. ex Schrank	152 366	MT712016.1
假活血草 <i>S. tuberifera</i> C. Y. Wu et C. Chen	152 332	MW376477.1
岩藿香 <i>S. franchetiana</i> H. Lév	151 852	MW376478.1
半枝莲 <i>S. barbata</i> D. Don	152 050	MW376479.1
黄芩 <i>S. baicalensis</i> Georgi	152 731	NC_027262.1
黄芩 <i>S. insignis</i> Nakai (<i>insignis</i> 黄芩)	151 908	NC_028533.1
缙云黄芩 <i>S. tsinyunensis</i> C. Y. Wu & S. Chow	152 089	NC_050161.1
龙头黄芩 <i>S. meehanioides</i> C. Y. Wu	152 484	NC_057189.1

的起始密码子, 在序列中间无终止密码子且末端处有正确的终止密码子。

2 方法

2.1 密码子相关参数的计算

使用 Codon W 1.4.2 和 CUSP 程序计算以下参数: 相对同义密码子使用度 (relative synonymous codons usage, RSCU)、有效密码子数 (effective number of codons, ENC)、密码子数 (N)、总 GC 含量 (GC_{all}) 和密码子第 1 位、第 2 位、第 3 位的 GC 含量 (GC₁、GC₂、GC₃), 使用 SPSS 27 软件对各参数进行相关性分析。所有的 CUB 参数均采用内部计算机程序进行估计。

2.2 RSCU 分析

RSCU 是指编码相应氨基酸的多个同义密码子中, 某个密码子的相对使用概率^[12]。RSCU > 1 说明此密码子的使用频率较为频繁且该物种对其具有较强的偏好性, 而 RSCU < 1 说明这个密码子喜好性较弱且使用频次较少。RSCU = 1 则说明对这个密码子的使用并无偏好^[13]。

2.3 密码子使用情况的对应性分析 (COA)

COA 是一种理解变量和样本之间关系的多变量统计方法。根据功能对基因进行分类, 并利用 Codon W 程序对每个基因的 RSCU 值进行对应分析^[14]。利用单个密码子的 RSCU 值, 分析 16 种

黄芩药用植物叶绿体基因密码子使用模式的主要趋势。

2.4 PR2-plot 分析

PR2 绘图方法是通过对密码子中第 3 个碱基组成的分类, 研究自然选择和突变压力对密码子使用偏好性的影响^[15]。分别以 $G_3/(G_3+C_3)$ 为横坐标, $A_3/(A_3+T_3)$ 为纵坐标制作散点图, 通过各点偏离中心点 ($A=T, G=C$) 确定碱基对序列所偏离的范围与方位。当密码子的第 3 位碱基表现为 $A=T$ 和 $C=G$ 时, 就说明在 DNA 的 2 条互补链上并不具有选择效应, 突变是影响密码子偏好性的唯一因素^[16]。

2.5 ENC-plot 分析

ENC 反映了基因中同义密码子密码子偏离随机选择的程度。ENC 被广泛用于衡量密码子偏好水平。以 ENC 为纵坐标, GC_3 为横坐标, 构建二维散点图, 并依据公式 $ENC=2+GC_3+29/[GC_3+(1-GC_3)^2]$ 在图中绘制 ENC 期望值标准曲线^[17]。各个基因点都分布在标准曲线周围, 这就说明密码子使用偏好性只受突变压力的影响。如各个基因点都集中在预期曲线下方, 则表明密码子使用偏好性受自然选择影响更多^[18]。

2.6 中性绘图分析

中性绘图分析以 GC_1 、 GC_2 的平均值 GC_{12} 为纵坐标, GC_3 为横坐标绘制散点图并进行回归分析, 通常用回归系数定量分析突变压力和自然选择对密码子使用偏好性的影响。回归系数越接近 1, 则密码子偏好性更多受内部碱基突变

的影响, 回归系数越接近 0, 则受自然选择的影响更多^[19]。

2.7 测量独立于长度和组成 (measurement-is independent of the length and composition, MILC)

MILC 即不依赖于基因长度和核苷酸组成的一种基因表达水平的测量方法。高 MILC 值意味着基因表达水平高, 反之亦然^[20]。

2.8 系统进化分析

在 NCBI 数据库中下载 16 种黄芩属药用植物叶绿体基因组上编码同一蛋白 psbA 的氨基酸序列, 以同属植物薄荷 *Mentha canadensis* L. 作为外类群, 利用 MAGA11 软件采用邻接法 (neighbor-joining, NJ), 设置 bootstrap replications 为 1000, 线程数量为 7 构建进化树来分析它们之间的进化关系。

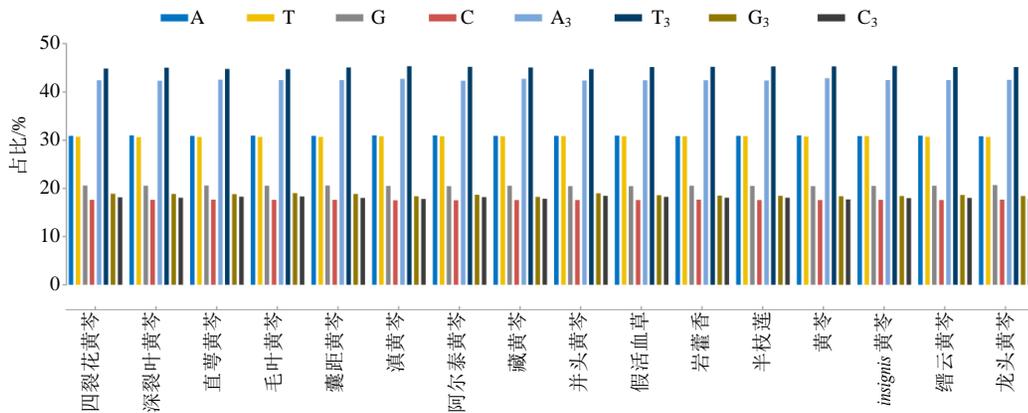
3 结果与分析

3.1 叶绿体基因组碱基组成分析

16 种黄芩属植物叶绿体基因碱基情况分布如图 1 所示, 观察到腺嘌呤 (A) 的使用率最高, 其次是胸腺嘧啶 (T)、鸟嘌呤 (G)、胞嘧啶 (C), 这表明碱基 A、T、G 和 C 的分布不均匀和对 A、T 的偏置性, 且各物种间的碱基分布规律相似, 也表明该 16 种黄芩属药用植物在进化程度上接近。

3.2 相对同义密码子使用偏好性分析

同义密码子的使用度结果如图 2 示, 阿尔泰黄芩、藏黄芩对 UGA 密码子无偏好性外对其余密码子都有偏好性。16 种黄芩属药用植物对其他 63 个



A、T、C、G 与 A_3 、 T_3 、 C_3 、 G_3 分别代表叶绿体基因组中各碱基的占比与各碱基在密码子第 3 位的占比。

A、T、C、G and A_3 、 T_3 、 C_3 、 G_3 represent proportion of each base in chloroplast genome and proportion of each base in the third position of codon.

图 1 16 种黄芩属植物叶绿体基因组碱基的分布

Fig. 1 Distribution of bases in chloroplast genomes of 16 *Scutellaria* plants

表2 密码子偏好性与碱基组成的相互关系

Table 2 Interrelationship between codon usage bias and base composition

物种名称	指数	相关系数									
		A	T	G	C	GC _{all}	A ₃	T ₃	G ₃	C ₃	GC ₃
四裂花黄芩	SCUO	0.155	0.113	-0.149	-0.147	-0.236	0.340	0.471	-0.653	-0.666	-0.784
	P	0.251	0.402	0.270	0.275	0.077	0.010	<0.001	<0.001	<0.001	<0.001
深裂叶黄芩	SCUO	0.176	0.065	-0.120	-0.107	-0.189	0.367	0.401	-0.625	-0.677	-0.783
	P	0.191	0.632	0.375	0.428	0.160	0.005	0.002	<0.001	<0.001	<0.001
直萼黄芩	SCUO	0.140	0.085	-0.103	-0.116	-0.202	0.307	0.462	-0.647	-0.676	-0.773
	P	0.298	0.528	0.446	0.390	0.131	0.020	<0.001	<0.001	<0.001	<0.001
毛叶黄芩	SCUO	0.135	0.081	-0.105	-0.104	-0.172	0.310	0.451	-0.653	-0.671	-0.774
	P	0.317	0.548	0.438	0.441	0.202	0.019	<0.001	<0.001	<0.001	<0.001
囊距黄芩	SCUO	0.141	0.086	-0.101	-0.114	-0.186	0.303	0.416	-0.636	-0.650	-0.768
	P	0.296	0.523	0.454	0.397	0.167	0.022	0.001	<0.001	<0.001	<0.001
滇黄芩	SCUO	0.167	0.078	-0.134	-0.100	-0.188	0.305	0.395	-0.639	-0.671	-0.787
	P	0.209	0.563	0.315	0.455	0.158	0.020	0.002	<0.001	<0.001	<0.001
阿尔泰黄芩	SCUO	0.168	0.085	-0.127	-0.104	-0.194	0.352	0.423	-0.632	-0.688	-0.791
	P	0.207	0.525	0.344	0.439	0.145	0.007	0.001	<0.001	<0.001	<0.001
藏黄芩	SCUO	0.208	0.040	-0.138	-0.142	-0.185	0.274	0.389	-0.552	-0.637	-0.754
	P	0.121	0.767	0.307	0.293	0.167	0.039	0.003	<0.001	<0.001	<0.001
并头黄芩	SCUO	0.125	-0.017	-0.046	0.029	-0.064	0.311	0.442	-0.670	-0.675	-0.777
	P	0.350	0.900	0.731	0.830	0.635	0.017	0.001	<0.001	<0.001	<0.001
假活血草	SCUO	0.142	0.061	-0.106	-0.075	-0.162	0.327	0.420	-0.648	-0.665	-0.772
	P	0.286	0.651	0.429	0.578	0.224	0.012	0.001	<0.001	<0.001	<0.001
岩藿香	SCUO	0.159	0.081	-0.125	-0.101	-0.186	0.327	0.396	-0.629	-0.648	-0.762
	P	0.236	0.548	0.354	0.454	0.166	0.013	0.002	<0.001	<0.001	<0.001
半枝莲	SCUO	0.152	0.063	-0.104	-0.074	-0.168	0.335	0.403	-0.640	-0.657	-0.764
	P	0.255	0.636	0.437	0.581	0.207	0.010	0.002	<0.001	<0.001	<0.001
黄芩	SCUO	0.145	0.087	-0.126	-0.119	-0.180	0.225	0.372	-0.589	-0.645	-0.771
	P	0.283	0.521	0.349	0.377	0.179	0.092	0.004	<0.001	<0.001	<0.001
insignis 黄芩	SCUO	0.155	0.030	-0.106	-0.033	-0.142	0.327	0.374	-0.634	-0.655	-0.762
	P	0.244	0.823	0.429	0.806	0.287	0.012	0.004	<0.001	<0.001	<0.001
缙云黄芩	SCUO	0.122	0.125	-0.159	-0.135	-0.230	0.317	0.425	-0.645	-0.653	-0.773
	P	0.366	0.355	0.237	0.316	0.085	0.016	0.001	<0.001	<0.001	<0.001
龙头黄芩	SCUO	0.122	0.125	-0.159	-0.135	-0.230	0.317	0.425	-0.645	-0.653	-0.773
	P	0.308	0.303	0.117	0.219	0.052	0.011	0.002	<0.001	<0.001	<0.001

表3 16种黄芩属植物叶绿体基因 GC 含量及 SCUO、MILC 汇总

Table 3 GC content of chloroplast genes and summarization of SCUO and MILC in 16 *Scutellaria* plants

物种名称	同义密码子使用顺序	MILC	GC _{all} /%	GC ₁ /%	GC ₂ /%	GC ₃ /%	回归系数
四裂花黄芩	0.20	0.56	38.36	44.45	39.42	31.21	0.000 67
深裂叶黄芩	0.20	0.56	38.33	44.44	39.47	31.08	0.061 70
直萼黄芩	0.20	0.56	38.37	44.36	39.47	31.26	0.059 94
毛叶黄芩	0.20	0.56	38.35	44.61	39.07	31.36	0.076 44
囊距黄芩	0.20	0.55	38.36	44.69	39.28	31.11	0.046 14
滇黄芩	0.20	0.56	38.19	44.54	39.34	30.68	0.019 43
阿尔泰黄芩	0.20	0.56	38.15	44.04	39.43	30.99	0.053 06
藏黄芩	0.20	0.55	38.25	44.71	39.30	30.76	0.032 39
并头黄芩	0.20	0.56	38.22	44.29	39.01	31.35	0.122 80
假活血草	0.20	0.56	38.19	44.36	39.30	30.93	0.053 41
岩藿香	0.20	0.56	38.33	44.42	39.71	30.87	0.033 49
半枝莲	0.20	0.56	38.24	44.36	39.54	30.83	0.046 17
黄芩	0.20	0.56	38.21	44.86	39.21	30.55	0.010 25
insignis 黄芩	0.20	0.56	38.26	44.41	39.64	30.72	0.036 86
缙云黄芩	0.20	0.56	38.29	44.53	39.45	30.88	0.023 74
龙头黄芩	0.20	0.56	38.47	45.03	39.67	30.72	0.015 43

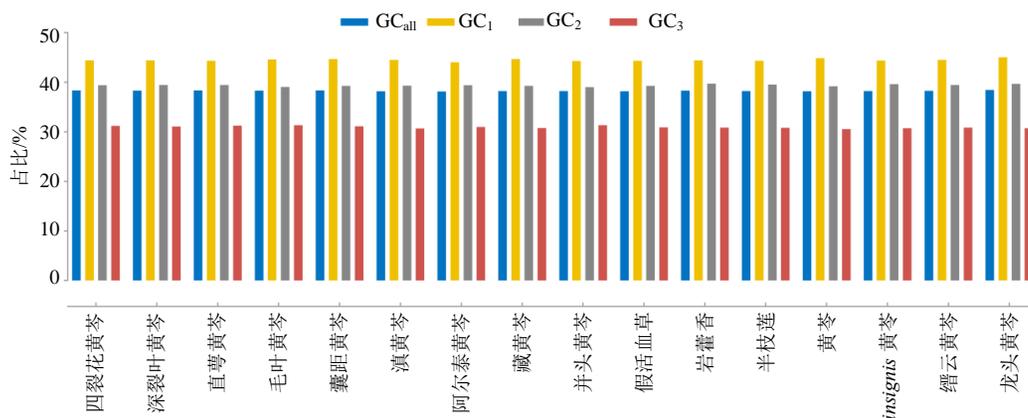


图3 16种黄芩属叶体基因组总GC含量及GC₁~GC₃分布

Fig. 3 Distribution of overall GC content, GC₁—GC₃ of cp genes in 16 medicinal plants of *Scutellaria* spp.

表4 16种黄芩属药用植物碱基组成性状的t检验

Table 4 t test for base composition traits of chloroplast genomes in 16 *Scutellaria* medicinal plants

物种名称	A+T	A+G	A+C	T+G	T+C	G+C	A ₃ +T ₃	A ₃ +G ₃	A ₃ +C ₃	T ₃ +G ₃	T ₃ +C ₃	G ₃ +C ₃	GC _{all} +GC ₁	GC _{all} +GC ₂	GC _{all} +GC ₃
四裂花黄芩	<i>t</i> 0.263	16.594	22.824	15.581	21.396	7.211	-2.505	23.200	26.523	25.701	65.117	0.780	-5.851	-1.321	8.188
	<i>P</i> 0.793	<0.001	<0.001	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001	<0.001	0.437	<0.001	0.189	<0.001
深裂叶黄芩	<i>t</i> 0.478	16.706	22.954	15.666	21.608	7.260	-2.845	23.034	26.968	27.494	72.415	0.864	-5.677	-1.429	8.881
	<i>P</i> 0.634	<0.001	<0.001	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001	0.389	<0.001	0.156	<0.001
直萼黄芩	<i>t</i> 0.294	16.618	22.785	15.722	21.564	7.236	-2.289	23.853	26.405	26.528	66.202	0.584	-5.754	-1.409	8.350
	<i>P</i> 0.769	<0.001	<0.001	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001	<0.001	0.560	<0.001	0.162	<0.001
毛叶黄芩	<i>t</i> 0.352	16.620	22.778	15.667	21.501	7.214	-2.285	22.582	26.114	25.083	64.560	0.767	-6.183	-0.879	8.104
	<i>P</i> 0.726	<0.001	<0.001	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001	<0.001	0.445	<0.001	0.381	<0.001
囊距黄芩	<i>t</i> 0.276	16.595	22.826	15.657	21.525	7.306	-2.774	24.239	27.074	27.542	68.781	0.885	-6.226	-1.134	8.718
	<i>P</i> 0.783	<0.001	<0.001	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001	0.378	<0.001	0.259	<0.001
滇黄芩	<i>t</i> 0.311	16.922	23.179	15.832	21.662	7.113	-2.622	25.457	29.070	29.107	71.027	0.666	-6.064	-1.397	9.407
	<i>P</i> 0.757	<0.001	<0.001	<0.001	<0.001	<0.001	0.010	<0.001	<0.001	<0.001	<0.001	0.507	<0.001	0.165	<0.001
阿尔泰黄芩	<i>t</i> 0.349	17.008	23.347	15.875	21.766	7.027	-2.839	24.154	27.421	27.292	67.339	0.554	-5.256	-1.578	8.631
	<i>P</i> 0.728	<0.001	<0.001	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001	<0.001	0.581	<0.001	0.117	<0.001
藏黄芩	<i>t</i> 0.111	16.427	22.560	15.421	21.069	7.324	-2.553	25.289	28.984	29.262	72.165	0.486	-6.304	-1.308	9.561
	<i>P</i> 0.911	<0.001	<0.001	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001	<0.001	0.628	<0.001	0.193	<0.001
并头黄芩	<i>t</i> 0.191	16.746	22.788	15.592	21.119	6.760	-2.361	22.857	26.049	25.448	65.711	0.450	-5.643	-0.993	7.967
	<i>P</i> 0.849	<0.001	<0.001	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001	<0.001	0.654	<0.001	0.323	<0.001
假活血草	<i>t</i> 0.263	16.900	23.085	15.690	21.359	6.982	-2.578	23.293	26.094	26.268	64.589	0.377	-5.713	-1.295	8.437
	<i>P</i> 0.793	<0.001	<0.001	<0.001	<0.001	<0.001	0.011	<0.001	<0.001	<0.001	<0.001	0.707	<0.001	0.198	<0.001
岩藿香	<i>t</i> 0.037	16.558	22.748	15.414	20.983	6.853	-2.914	24.433	27.153	27.232	65.702	0.486	-5.671	-1.644	8.856
	<i>P</i> 0.971	<0.001	<0.001	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001	0.628	<0.001	0.103	<0.001
半枝莲	<i>t</i> 0.185	16.847	23.066	15.676	21.351	6.983	-2.863	24.474	27.004	27.643	67.159	0.433	-5.691	-1.553	8.781
	<i>P</i> 0.853	<0.001	<0.001	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001	0.666	<0.001	0.123	<0.001
黄芩	<i>t</i> 0.294	16.694	22.973	15.566	21.362	6.903	-2.524	25.666	29.402	28.417	66.206	0.797	-6.575	-1.203	9.926
	<i>P</i> 0.770	<0.001	<0.001	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001	<0.001	0.427	<0.001	0.232	<0.001
insignis 黄芩	<i>t</i> 0.155	16.665	22.714	15.517	21.034	6.771	-2.970	24.647	27.346	28.327	68.525	0.401	-5.811	-1.721	9.215
	<i>P</i> 0.877	<0.001	<0.001	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001	0.689	<0.001	0.088	<0.001
缙云黄芩	<i>t</i> 0.284	16.616	22.893	15.539	21.370	7.164	-2.705	24.062	27.089	26.446	63.573	0.674	-5.913	-1.408	8.774
	<i>P</i> 0.777	<0.001	<0.001	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001	0.502	<0.001	0.162	<0.001
龙头黄芩	<i>t</i> 0.143	15.915	21.940	15.030	20.658	7.275	-2.645	23.815	25.734	27.227	65.748	0.581	-6.319	-1.471	8.887
	<i>P</i> 0.886	<0.001	<0.001	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001	0.563	<0.001	0.144	<0.001

表 5 16种黄芩属叶绿体密码子 SCUO 和 MILC 之间的 Karl Pearson 分析

Table 5 Karl Pearson analysis between chloroplast codons SCUO and MILC in 16 *Scutellaria* plants

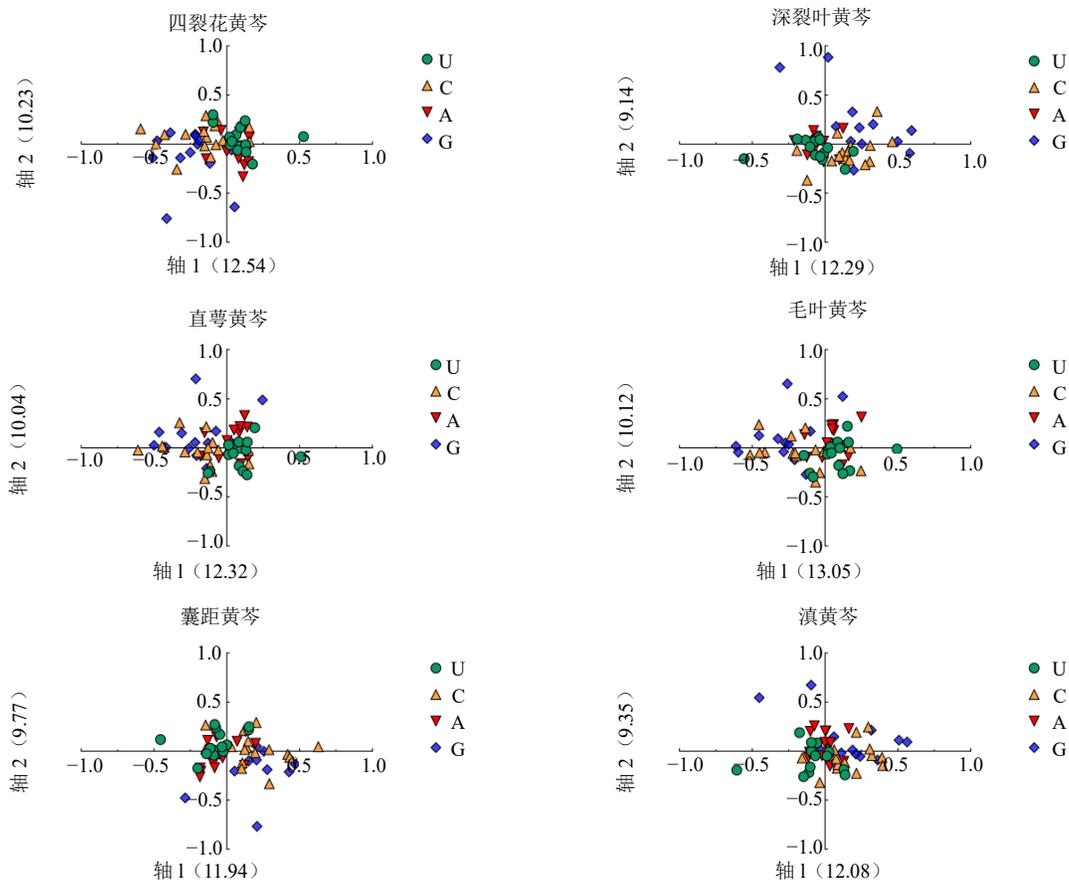
物种名称	<i>r</i>	<i>P</i>
四裂花黄芩	0.481	0.000
深裂叶黄芩	0.500	0.000
直萼黄芩	0.533	0.000
毛叶黄芩	0.501	0.000
囊距黄芩	0.523	0.000
滇黄芩	0.481	0.000
阿尔泰黄芩	0.484	0.000
藏黄芩	0.528	0.000
并头黄芩	0.408	0.001
假活血草	0.486	0.000
岩藿香	0.497	0.000
半枝莲	0.515	0.000
黄芩	0.484	0.000
<i>insignis</i> 黄芩	0.388	0.003
缙云黄芩	0.481	0.000
龙头黄芩	0.481	0.000

3.5 16种黄芩属植物叶绿体基因密码子 COA

通过 COA 分析,了解不同黄芩叶绿体基因组之间密码子使用情况的差异。COA 的轴 1 和轴 2 如图 4 所示。在 16 种黄芩属药用植物中,A\U 末端密码子和 G\U 末端密码子都占据了小区的 4 个正方形,因此有些点可以很好地分离出来。不同小区中点的分布模式不同,表明密码子的使用情况不同,不同黄芩品种的模式不同。部分密码子离基因轴非常接近,说明突变压力下的组成特征可能影响叶绿体基因的密码子使用偏好性。

3.6 PR2-plot 分析

结果如图 5 所示,可以看到基因在平面上零散分布在 4 个领域,少数基因在中心附近,表明自然选择和突变压力可能影响黄芩属物种叶绿体基因密码子使用偏好性。16 种黄芩属药用植物叶绿体基因主要分布在右下方区域,表明密码子第 3 位上碱基使用频率 U 高于 A, G 高于 C^[25]。若密码子使用偏好性完全是由突变引起的,则 G 和 C 以及 A 和 U 的使用频率必须相等^[26]。A 和 U 以及 G 和 C 的使用不均衡(图 1),这说明 16 种黄芩属药用植物的密码子使用偏好性不仅存在自然选择因素的影



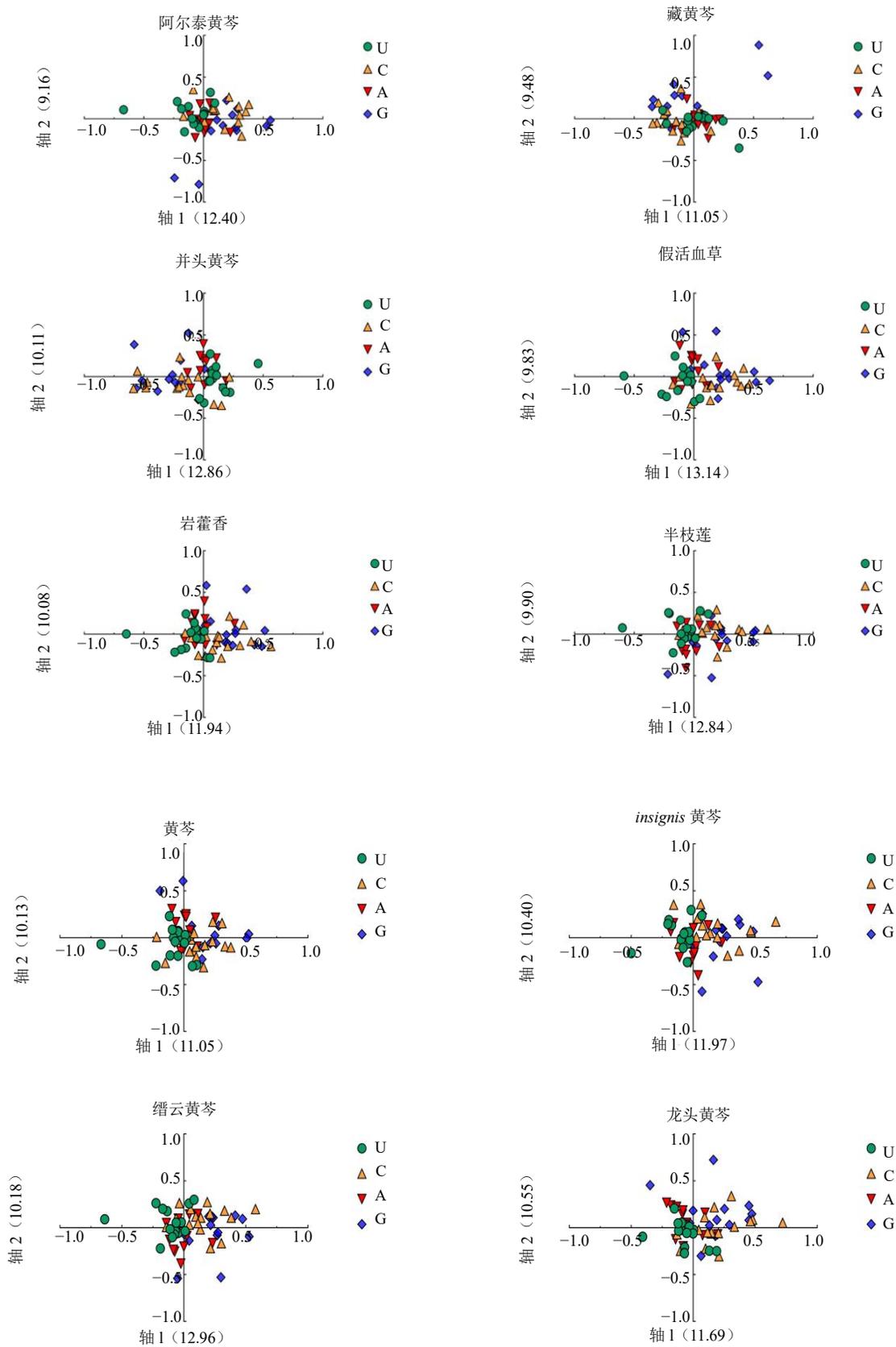


图 4 16种黄芩属植物叶绿体基因密码子对应性分析

Fig. 4 Correspondence analysis of chloroplast genes in 16 *Scutellaria* plants

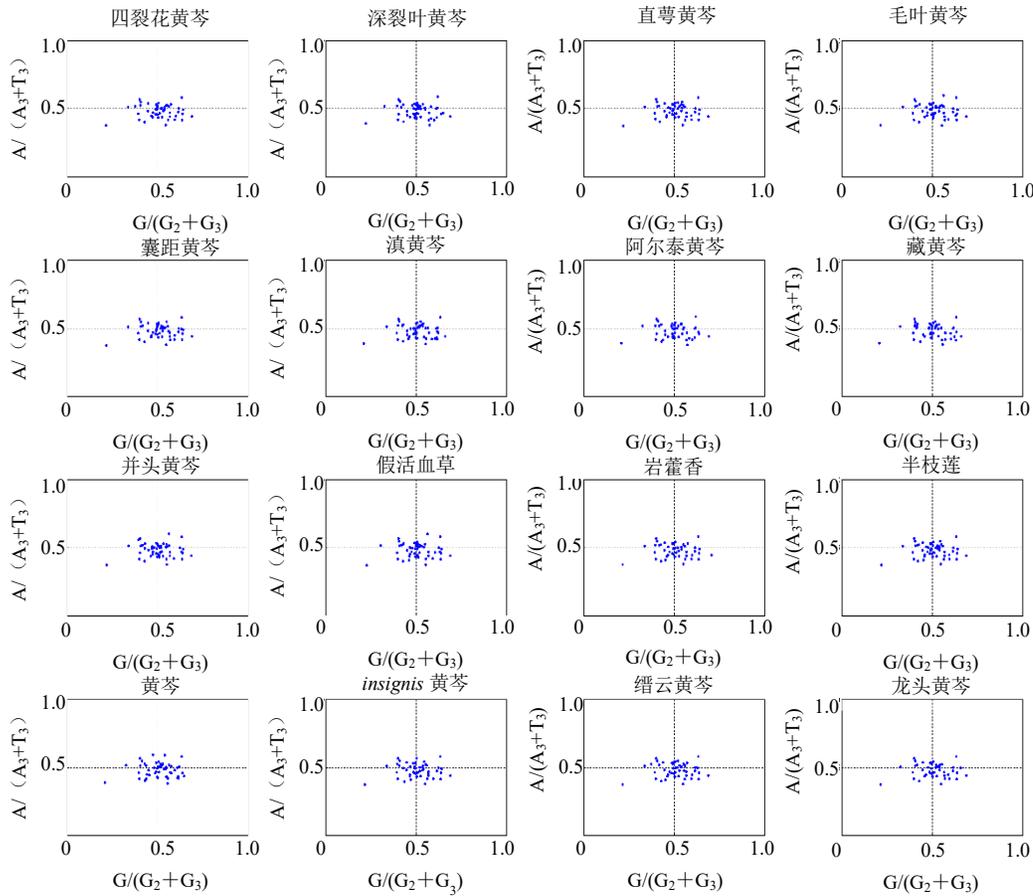


图5 16种黄芩属药用植物叶绿体基因的PR2偏倚图谱

Fig. 5 PR2 bias mapping of chloroplast genes of 16 *Scutellaria* medicinal plants

响，而且也受到自身内部碱基突变的影响。

3.7 中性绘图分析

中性绘图结果如图6所示，大部分叶绿体基因分布相对集中，GC₃分布范围较窄，部分点均有对角分布，表明自然选择和突变压力可能对其密码子偏好性都有影响。其中四裂花黄芩、深裂叶黄芩、直萼黄芩、毛叶黄芩、囊距黄芩、滇黄芩、阿尔泰黄芩、藏黄芩、并头黄芩、假活血草、岩藿香、半枝莲、黄芩、*insiginis* 黄芩、缙云黄芩、龙头黄芩回归线的斜率分别为0.02948、0.2543、0.2287、0.2601、0.2076、0.1490、0.2361、0.1949、0.3607、0.2273、0.1825、0.2136、0.1144、0.1974、0.1482、0.1130，无显著相关性，因此判断影响16种黄芩属药用植物叶绿体基因组的密码子使用偏好性的因素中，自然选择占主导地位。

3.8 ENC-plot 分析

结果如图7所示，ENC和GC₃在16种黄芩属药用植物叶绿体基因组中的分布相似。大部分基因点集中分布在曲线两侧，只有少数点靠近曲线的位置，表现出的偏性较弱；大多数ENC值位于曲线的下方，表现出的偏性较强，说明突变压力对叶绿体基因组密码子使用偏好性影响不大，更多受自然选择压力影响^[27]。这也符合中性图分析的结果^[28]。

置，表现出的偏性较弱；大多数ENC值位于曲线的下方，表现出的偏性较强，说明突变压力对叶绿体基因组密码子使用偏好性影响不大，更多受自然选择压力影响^[27]。这也符合中性图分析的结果^[28]。

3.9 系统发育分析

通过进化树分析可了解黄芩在该16种黄芩属药用植物中的系统位置及它们之间的进化和亲缘关系，结果如图8所示，薄荷独为一支，16种黄芩属药用植物聚为一大支，其进一步又可分为2个分支，其中缙云黄芩独为一支，其余15种聚为一大支，说明缙云黄芩与其他15种黄芩的亲缘关系更远。四裂花黄芩与深裂叶黄芩聚为一支说明二者进化程度相似且亲缘关系相近（可信度为100），二者与黄芩的亲缘关系较远与缙云黄芩亲缘关系最远。

4 讨论

碱基组成是基因组DNA的一个重要特征，对密码子的使用也有重要影响^[29]。不同碱基核苷酸的频率取决于突变和反向突变之间的平衡。密码子使用偏好性是生物基因组进化的一个重要特征，最优

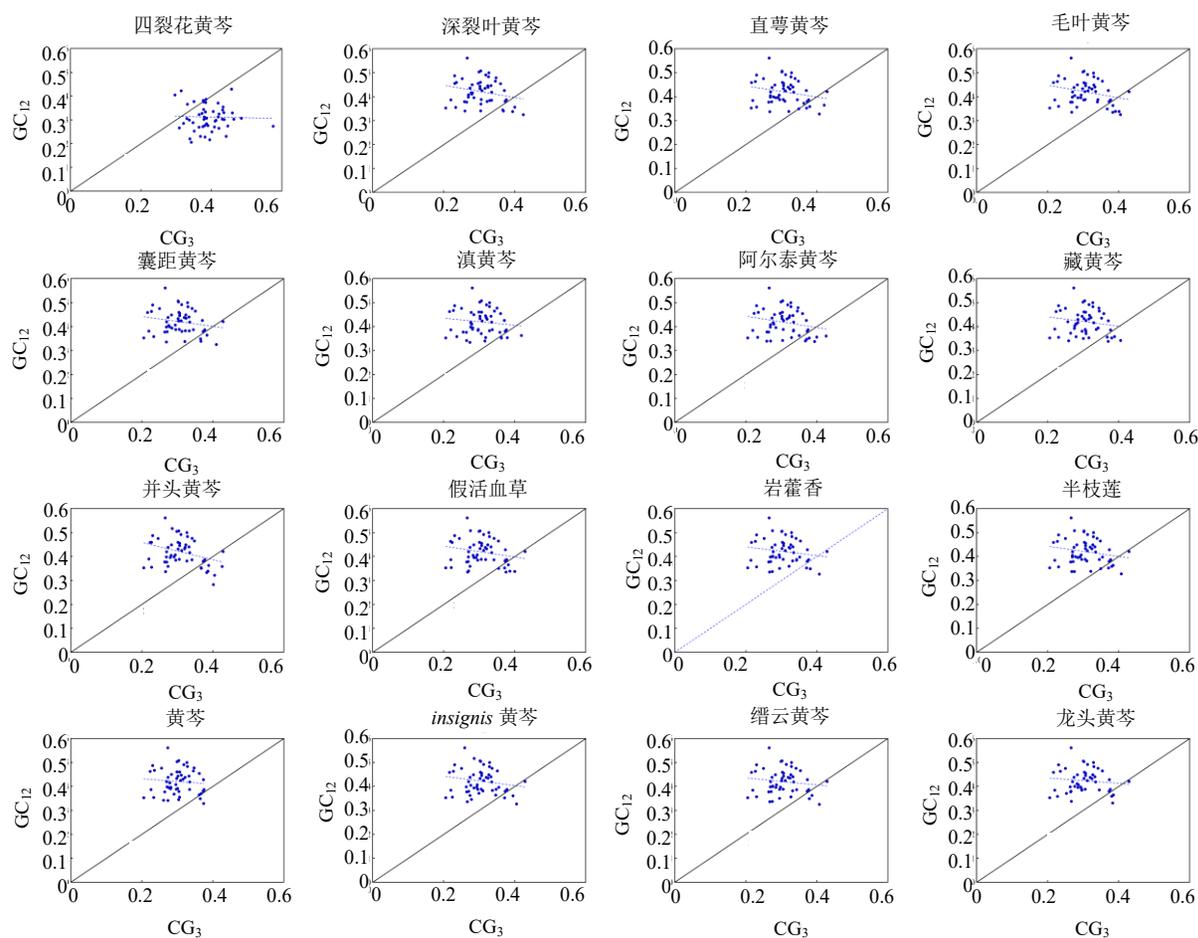


图 6 16种黄芩属药用植物叶绿体基因中性图

Fig. 6 Neutral map of chloroplast genes in 16 *Scutellaria* medicinal plants

密码子的形成一般是由基因突变和自然选择共同作用的结果^[27]。密码子偏好性分析对研究分子进化和异源蛋白表达等有重要意义^[30]。为了探讨影响 16 种黄芩属药用植物叶绿体基因组密码子使用偏好性及其因素，为后续黄芩属药用植物遗传育种、资源保护等研究提供参考。本研究通过对四裂花黄芩、深裂叶黄芩、直萼黄芩、毛叶黄芩、囊距黄芩、滇黄芩、藏黄芩、并头黄芩、假活血草、岩藿香、半枝莲、黄芩等黄芩属植物的叶绿体基因组序列进行了 Karl Pearson 相关性分析。研究发现，16 种黄芩属药用植物的叶绿体基因组大小最大相差 1 056 bp (表 1)，各物种间差异较小。密码子的使用模式与 GC 的含量有紧密的联系。本研究中 16 种黄芩属药用植物叶绿体基因组的 GC₁(44.03%~44.86%)、GC₂(39.01%~39.71%)、GC₃(30.55%~31.36%)和总 GC(38.15%~38.47%)含量都小于 50%，并且 GC₃含量最低，说明该 16 种药用黄芩的叶绿体基因组更偏向于采用 A/T 碱基和以 A/U

结尾的密码子，黄芩属植物中不同物种的叶绿体基因组中的总体密码子使用偏好性相似，与已报道的秋茄^[31]、五指毛桃^[32]、籽粒苋^[33]、油茶^[34]等多种植物叶绿体基因组使用偏好性一致，也就说明高等植物在叶绿体基因 GC 含量及密码子使用上是具有相似性的^[35]。密码子第 3 个碱基与第 1、2 位碱基相比有较高的变异性，但它的突变不能改变氨基酸的类型，但它仍被认为是决定氨基酸类型的一个重要特征，因此 GC₃经常被用作密码子偏好性的一个重要指标。由于碱基突变偏好 G\C 或 A\U 造成了基因组内同义密码子的使用偏好性^[36]。虽然碱基组成被认为是引起黄芩属植物叶绿体密码子使用偏好性的主要因素，但嘧啶和嘌呤之间的不对称性表明除此之外，还有一些其他因素也影响黄芩属植物叶绿体基因组的密码子使用偏好性^[29]。在 ENC 图(图 7)中，大多数基因的分布没有遵循标准曲线，这为这一观点提供了额外的支持。在这些因素中，突变压力和自然选择是导致嘧啶和嘌呤在

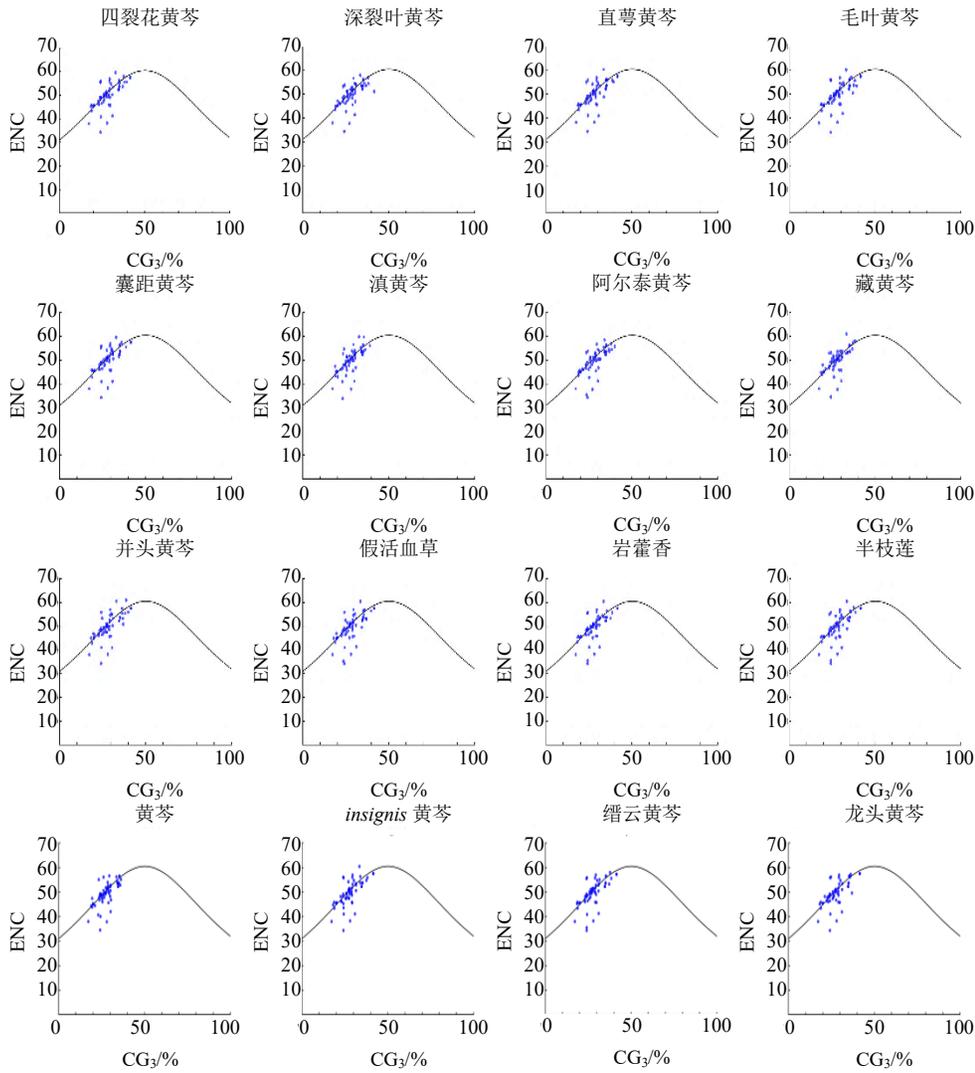


图 7 16 种黄芩属药用植物叶绿体基因组 ENC-plot 分析

Fig. 7 ENC-plot analysis of chloroplast genes in 16 *Scutellaria* medicinal plants

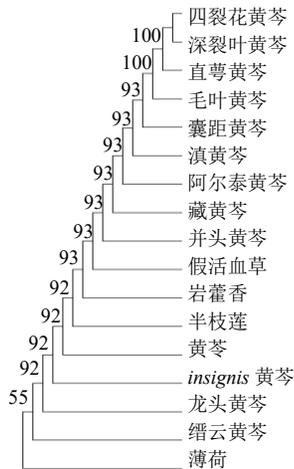


图 8 16 种黄芩属药用植物系统发育图

Fig. 8 Phylogenetic diagram of 16 *Scutellaria* medicinal plants

DNA 序列中不对称使用的最合理的原因。

PR2-plot 结果表明,影响 16 种黄芩叶绿体的密码子使用偏好性的还有除内部碱基突变外的其他因素,如自然选择。与酸枣^[37]、使君子科植物^[38]、珙桐^[39]、蒙古韭^[40]等的结论相似。ENC 的值大于 35 的占多数表明多数密码子的偏性较弱,作为密码子偏好性图像可视化的分析手段,ENC-plot 分析从 16 种黄芩属药用植物 ENC 和 GC₃ 的分布关系中发现,自然选择压力对密码子使用偏好性的影响较大,中性图分析的结果佐证了这一结论。

黄芩属植物的密码子使用模式以及叶绿体基因的遗传多样性将促进黄芩属植物物种的系统发育和遗传学研究。本研究分析了四裂花黄芩、深裂叶黄芩、直萼黄芩、毛叶黄芩、囊距黄芩、滇黄

芩、藏黄芩、并头黄芩、假活血草、岩藿香、半枝莲、黄芩等黄芩属药用植物的叶绿体基因组编码基因的密码子偏好性形成的影响因素和特点,对黄芩属药用植物叶绿体基因组学、物种进化特征及遗传多样性分析等研究具有重要意义,但只选取了16种该属药用植物有一定的局限性,后续研究考虑增加黄芩属药用植物叶绿体基因组的数量,为研究其基因组提供更丰富的参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国科学院. 中国植物志编委会. 中国植物志 [M]. 北京: 科学出版社, 1991: 212-230.
- [2] 李影, 周小青, 陈杰, 等. 黄芩属植物化学成分及药理作用的研究进展 [J]. 江西中医药大学报, 2023, 35(2): 111-120.
- [3] 原明星, 尹康权, 李俄仁措, 等. 叶绿体基因编辑技术: 进展与展望 [J]. 中国草地学报, 2024, 46(6): 126-134.
- [4] 林丹, 李冰冰, 翟晓巧. 泡桐属植物亲缘关系研究 [J]. 河南林业科技, 2019, 39(2): 15-17.
- [5] Begum N S, Chakraborty S. Influencing elements of codon usage bias in Birnaviridae and its evolutionary analysis [J]. *Virus Res*, 2022, 310: 198672.
- [6] Chakraborty S, Uddin A, Choudhury M N. Factors affecting the codon usage bias of *SRY* gene across mammals [J]. *Gene*, 2017, 630: 13-20.
- [7] 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.
- [8] 胡振民, 万青, 李欢, 等. 茶树 CsNRT1.1 基因密码子使用特性分析 [J]. 江苏农业学报, 2019, 35(4): 896-903.
- [9] 王艳, 赵懿琛, 赵德刚. 杜仲基因密码子使用模式分析 [J]. 广西植物, 2021, 41(2): 274-282.
- [10] 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.
- [11] 李滢, 匡雪君, 孙超, 等. 川贝母转录组密码子使用偏好性分析 [J]. 中国中药杂志, 2016, 41(11): 2055-2060.
- [12] 洪园淑, 周玉梅, 李文娟, 等. 苦豆子 SaLDC 密码子偏好性分析、优化及原核表达 [J]. 中成药, 2021, 43(2): 429-435.
- [13] 张俊焱, 曾阳, 李锦萍, 等. 2 种药用獐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.
- [14] 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.
- [15] 黄蔚虹, 陈永杰, 孙彦阔, 等. H9N2 禽流感病毒全基因组密码子使用偏好性及影响因素分析 [J]. 华南农业大学学报, 2020, 41(3): 15-22.
- [16] 原晓龙, 康洪梅, 王毅. 云南蓝果树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.
- [17] 代江鹏, 蔡一鸣, 刘巧珍, 等. 甘草属 7 种植物叶绿体基因组密码子使用偏好性分析 [J]. 中草药, 2023, 54(9): 2907-2916.
- [18] 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, 2021, 36(3): 384-392.
- [19] 季凯凯, 宋希, 陈春国, 等. 木兰科叶绿体基因组的密码子使用特征分析 [J]. 中国农业科技导报, 2020, 22(11): 52-62.
- [20] Wang Z J, Cai Q W, Wang Y, et al. Comparative analysis of Codon bias in the chloroplast genomes of *Theaceae* species [J]. *Front Genet*, 2022, 13: 824610.
- [21] 杨亚蒙. 基于叶绿体基因组的葡萄属分子系统学研究 [D]. 北京: 中国农业科学院, 2019.
- [22] 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.
- [23] Chakraborty S, Basumatary P, Nath D, et al. Compositional features and pattern of codon usage for mitochondrial CO genes among reptiles [J]. *Mitochondrion*, 2022, 62: 111-121.
- [24] 高梦琦, 邹建珍, 霍小位, 等. 基于转录组数据分析药用大黄的密码子使用偏好性 [J]. 中草药, 2021, 52(20): 6344-6349.
- [25] 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.
- [26] 喻凤, 韩明. 紫花苜蓿叶绿体基因组密码子偏好性分析 [J]. 广西植物, 2021, 41(12): 2069-2076.
- [27] 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.
- [28] 陈芬, 郑玉斯, 梁霞, 等. 山麦冬叶绿体基因组密码子使用的偏好性 [J]. 北部湾大学学报, 2021, 36(2): 59-67.
- [29] Zhang R Z, Zhang L, Wang W, et al. Differences in Codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild *Solanum* species [J]. *Int J Mol Sci*, 2018, 19(10): 3142.

- [30] 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.
- [31] 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.
- [32] 黄琼林, 谭靖怡, 叶晓霞, 等. 五指毛桃叶绿体基因组结构与序列特征分析 [J]. 福建农林大学学报: 自然科学版, 2024, 53(3): 323-331.
- [33] 冯瑞云, 梅超, 王慧杰, 等. 籽粒苋叶绿体基因组密码子偏好性分析 [J]. 中国草地学报, 2019, 41(4): 8-15.
- [34] 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.
- [35] 马孟莉, 张薇, 孟衡玲, 等. 豆蔻属药用植物叶绿体基因组密码子偏性分析 [J]. 中草药, 2021, 52(12): 3661-3670.
- [36] 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.
- [37] 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.
- [38] 钟才荣, 何斯敏, 方赞山, 等. 使君子科 3 种红树植物叶绿体基因组密码子使用偏好性分析 [J]. 植物资源与环境学报, 2024, 33(2): 1-12.
- [39] 罗永坚, 王茹, 赵仁菲, 等. 珙桐叶绿体基因组同义密码子使用偏好性分析 [J]. 北京林业大学学报, 2024, 46(3): 8-16.
- [40] 王媛媛, 杨美青. 蒙古韭叶绿体基因组密码子使用偏好性分析 [J]. 分子植物育种, 2021, 19(4): 1084-1092.

[责任编辑 时圣明]