



20 种前胡属叶绿体基因组密码子偏好性分析

周梦丽, 李红, 郭蕾, 王霞, 王瑞红

(浙江理工大学生命科学与医药学院, 杭州 310018)

摘要: 为探讨 20 种前胡属植物叶绿体基因组密码子的使用偏好性及其影响因素, 使用 EMBOSS、CodonW、CUSP 软件计算了密码子组成, 并用 SPSS20.0 绘制相应的统计图。结果表明: 20 种前胡属叶绿体基因组的平均 GC 碱基数量占比为 39.89%, GC₃ 含量为 30.02%~30.22%, ENC 值为 49.73~49.90, 密码子偏好性较弱; 蛋白质编码区序列呈现 A/T 碱基富集现象, 第三位密码子大多以 A 或 U 结尾, 中性绘图等分析表明密码子偏好性主要影响因素是自然选择; 筛选出 29 个高频密码子, 其中 UUA、UGU、UCU 等 17 个为相同的最优密码子, 叶绿体基因组编码序列与叶绿体全基因组构建的系统进化树具有极高相似性, 密码子使用偏好性与进化关系有关。该研究结果为进一步开展前胡属分子系统学、遗传转化和叶绿体基因工程研究奠定了基础。

关键词: 前胡属; 叶绿体基因组; 密码子偏好性; 系统发育分析; 自然选择

中图分类号: Q943.2

文献标志码: A

文章编号: 1673-3851(2025)09-0664-12

引文格式: 周梦丽, 李红, 郭蕾, 等. 20 种前胡属叶绿体基因组密码子偏好性分析[J]. 浙江理工大学学报(自然科学), 2025, 53(5): 664-675.

Reference Format: ZHOU Mengli, LI Hong, GUO Lei, et al. Analysis of codon usage bias in 20 *Peucedanum* chloroplast genomes[J]. Journal of Zhejiang Sci-Tech University, 2025, 53(5): 664-675.

Analysis of codon usage bias in 20 *Peucedanum* chloroplast genomes

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

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

Abstract: To investigate codon usage bias in 20 *Peucedanum* chloroplast genomes, we employed bioinformatics tools EMBOSS, CodonW, and CUSP for codon composition analysis, visualized by SPSS20.0. The study showed that the average GC base content was 39.89%, GC₃ content ranged from 30.02% to 30.22% and ENC values ranged from 49.73 to 49.90, indicating that the codon usage bias was weak; the sequences of protein coding regions showed A/T base enrichment phenomenon, with a majority of the third codons ending with A/U, and neutrality plot and other analyses showed that the main driver of codon usage bias was natural selection. Twenty-nine high-frequency codons were screened, of which 17 were the same optimal codons such as UUA, UGU and UCU. Based on the extremely high similarity between the coding sequences of the chloroplast genome and the phylogenetic tree constructed from the whole chloroplast genome, it was shown that the codon usage bias was related to the evolutionary relationship. The findings of this study lay a foundation for further research on molecular phylogenetics, genetic transformation, and chloroplast genetic engineering in *Peucedanum* species.

Key words: *Peucedanum*; chloroplast genomes; codon usage bias; phylogenomics; natural selection

收稿日期: 2025-03-07 网络出版日期: 2025-06-17

基金项目: 浙江省“尖兵”“雄鹰”研发攻关计划(2023C02017); 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)

作者简介: 周梦丽(1999—), 女, 山东临沂人, 硕士研究生, 主要从事植物系统进化与转录组学方面的研究。

通信作者: 王瑞红, E-mail: ruihong8977@zstu.edu.cn

0 引言

前胡属(*Peucedanum* L.)是伞形科(Apiaceae)中一个广泛分布且种类繁多的属,有 100~120 种,分布于欧亚大陆、南非和澳大利亚^[1-2]。欧洲和东亚地区是前胡属多样性分布中心,中国有 40 种,其中 33 种为中国特有种^[3]。前胡属大多数物种在传统中药或民间偏方中得到了广泛使用。《中国药典》中记载的前胡(*Peucedani radix*)是白花前胡(*Peucedanum praeruptorum* Dunn)的干燥根,其药性凉,味甘苦,具有清热解表、祛痰止咳、散风清热的功效,常用于治疗咳嗽、痰多、风热感冒等^[4]。

叶绿体对光合作用至关重要,为植物提供能量并促进次生代谢产物的合成^[5]。叶绿体具有半自主性遗传特性,具有独立的 DNA 复制、转录和翻译系统^[6];叶绿体的遗传特性还包括核苷酸替换率低和母系遗传模式,这些特点使其成为植物系统发育和进化研究的重要资源^[7]。密码子是连接核酸和蛋白质的纽带,其简并性使得一种氨基酸可以由多个同义密码子编码^[8]。基因组编码过程中,同义密码子的使用偏好性(Codon usage bias, CUB)反映了遗传信息转录翻译过程中对特定核苷酸组合的适应性选择现象^[9]。多种因素会影响同义密码子的偏好性,包括基因表达水平和 GC 含量等^[10]。生物密码子偏好性的形成是由自然选择、突变压力和选择压力等多种因素共同塑造的^[11]。Liu 等^[12]对 7 个前胡属植物的叶绿体基因组比较显示高表达基因(如 *ndhF*)偏好使用与 tRNA 丰度匹配的密码子(UCU、GAA);48 种乌头属植物的叶绿体基因组密码子偏好性分析表明,高海拔植物的 GC3 含量显著低于低海拔植物,这种差异与高海拔环境的低温、强紫外线等因素有关;A/T 富集的密码子可能通过优化 tRNA 适配性提升光合基因的翻译效率,增强植物的生存适应性^[5]。水稻(*Oryza*)叶绿体基因组中,光合相关基因(如 *rbcL* 和 *psbA*)偏好使用 A/U 结尾的密码子(如 UUA、GUA),这些密码子与叶绿体高丰度 tRNA 匹配,显著提升光合蛋白的合成效率^[13]。因此,密码子使用偏好性不仅能反映物种或基因的起源和演化,而且有利于在叶绿体遗传工程中选择合适的基因,提高外源基因的表达效率,对基因功能和蛋白表达都有重要意义^[14]。

前胡属作为伞形科药用植物的重要类群,富含香豆素类等活性成分,但其叶绿体基因组密码子使

用模式及进化机制尚不明确。本文以 20 个前胡属植物为研究对象,解析其密码子偏好性特征及影响因素,并构建系统发育树以揭示物种进化轨迹,以期前胡属植物系统进化研究以及资源开发利用等提供理论依据。

1 材料与方法

1.1 前胡属植物叶绿体基因组序列获取

通过美国国家生物技术信息中心基因库(National center for biotechnology information, NCBI)获取 20 种前胡属植物的叶绿体基因组数据,将其作为研究对象,相关信息见表 1。为了避免因短序列长度引起采样偏差和估计误差^[15],筛选出的蛋白质编码序列(Coding sequence, CDS)长度应超过 300 bp,移除编码相同蛋白质的重复序列,具体方法参考文献^[16];每个选定的序列的起始密码子为 ATG,终止密码子为 TAA、TAG 或 TGA,且不应出现任何提前终止密码子。最终,每个物种筛选出 46 条 CDS 用于后续密码子使用偏好性分析。

表 1 前胡属叶绿体基因组的基因信息

编号	物种名称	登录号	长度/bp
1	天竺山前胡 <i>P. ampliatum</i>	OK336475	147403
2	<i>P. chujaense</i>	MT233391	147839
3	南川前胡 <i>P. dissolutum</i>	NC_084401	154686
4	刺尖前胡 <i>P. elegans</i>	OP379702	147790
5	广西前胡 <i>P. guangxiense</i>	NC_084402	153439
6	<i>P. hakuunense</i>	NC_060840	147426
7	华北前胡 <i>P. harry-smithii</i>	OK336476	147046
8	黄山前胡 <i>P. huangshanense</i>	NC_085568	147229
9	滨海前胡 <i>P. japonicum</i>	OK336477	147592
10	华山前胡 <i>P. ledebourielloides</i>	NC_087702	151015
11	南岭前胡 <i>P. longshengense</i>	OK336479	147967
12	马山前胡 <i>P. mshanense</i>	OK336478	154230
13	华中前胡 <i>P. medicum</i>	OK336473	152288
14	疆前胡 <i>P. morisonii</i>	NC_087703	147105
15	矮前胡 <i>P. nanum</i>	NC_084445	146900
16	前胡 <i>P. praeruptorum</i>	NC_060841	146930
17	红前胡 <i>P. shanianum</i>	NC_084403	148311
18	紫茎前胡 <i>P. violaceum</i>	NC_084444	148190
19	泰山前胡 <i>P. warwae</i>	NC_067882	147087
20	武隆前胡 <i>P. wulongense</i>	NC_084404	148112

注:*P. chujaense* 和 *P. hakuunense* 是韩国特有的珍稀物种,尚未有中文命名。

1.2 方法

1.2.1 密码子组成分析

使用 CodonW V1.4.2 软件^[17]对 20 种前胡属植物叶绿体基因组编码序列进行综合分析,包括相

对同义密码子使用度(Relative synonymous codon usage, RSCU)、有效密码子数(Effective number of codon, ENC)、同义密码子的 GC 含量以及第三位密码子上的 A、T、C、G 含量(A_3 、 T_3 、 C_3 、 G_3)。GC_{3s} 为同义密码子第三核苷酸位点的 GC 含量。基于 CUSP 计算密码子碱基构成,得出总 GC 含量(GC_{all})以及各位点的 GC 含量(GC₁、GC₂ 和 GC₃)。

1.2.2 ENC-plot 分析

ENC 值介于 20~61 之间,ENC \leq 35 表示基因有着显著的密码子偏差,ENC 值越高则代表密码子使用偏好性越弱。ENC-Plot 是以 GC_{3s} 为横坐标,ENC 为纵坐标绘制的图表,根据公式计算 ENC 期望值: $ENC_{exp} = 2 + GC_{3s} + 29/[GC_{3s}^2 + (1 - GC_{3s})^2]$,利用 R 语言中的 ggplot2 包绘制 ENC-Plot;通过比较 ENC 实际值与 ENC 期望值,可以判断密码子使用偏好性的主要影响因素。其中标准曲线代表无选择压力存在时,密码子偏好性由突变决定。

1.2.3 PR2-plot 分析

PR2-Plot 用于分析突变压力和选择对密码子偏好性的影响。PR2-plot 以 $A_3/(A_3 + T_3)$ 为纵坐标, $G_3/(G_3 + C_3)$ 为横坐标进行绘图,单个基因的密码子使用偏性通过其在图中的位置和向量方向来表示,其中图的中心点表示 A=T 和 C=G,处于无偏性使用时的密码子状态。

1.2.4 中性绘图分析

计算 CDS 中 GC₁ 和 GC₂ 含量,具体方法参考文献[18],取两者平均值 GC₁₂。以 GC₃ 为横坐标,GC₁₂ 为纵坐标绘制中性图。通过分析二者间的相关性,鉴别密码子偏好性的主要影响因素。二者显著相关,表明突变是影响密码子使用偏性的关键性因素;相关性不显著,则表明选择压力对密码子使用偏性的影响更大[18]。

1.2.5 同义密码子分析

RSCU 作为衡量密码子偏好性的指标,通过某一特定密码子的实际使用频率与在无偏好下预期使用频率的相对关系来揭示偏好程度[19]。在本文中,将数据整理于 Microsoft Excel 中,利用 Tbttools V1.113 生成同义密码子分析热图。

1.2.6 最优密码子分析

在最优密码子分析中,ENC 被用作衡量密码子偏好性的参考标准。根据基因 ENC 值排序,取前 10% 的基因(*ndhE*、*petD*、*rpl16* 和 *rps8*)建立高表达基因数据库,取后 10% 的基因(*atpE*、*rpl2*、*rps4*

和 *rps11*)建立低表达基因数据库,计算 2 个基因库中密码子的 RSCU 和 Δ RSCU(高低表达基因组 RSCU 值的差)值, Δ RSCU \geq 0.08 的为高表达密码子,满足高频密码子(RSCU $>$ 1)和高表达密码子(Δ RSCU \geq 0.08)的为最优密码子[20]。

1.2.7 系统发育分析

使用 MAFFT V7 软件[21]在默认参数下对 CDS 序列和叶绿体基因组序列进行比对,根据 Wen 等[22]的研究结果,选择毒芹(*Cicuta virosa* L.)和鸭儿芹(*Cryptotaenia japonica* Hassk.)作为外类群用于构建系统发育树的根。基于最大似然法(Maximum Likelihood, ML)和 GTRGAMMA 模型,进行 1000 次重复分析,构建 20 个前胡属植物的 46 个 CDS 序列和叶绿体基因组序列系统发育树。

2 结果与分析

2.1 前胡属叶绿体基因组密码子组成分析

根据已有文献以及查询 NCBI 数据库相关前胡属植物序列信息,选出 20 种前胡属植物。根据 CUSP 所计算出的密码子各类 GC 含量,结果如图 1 所示。图 1 显示,20 种前胡属植物的平均 GC_{all} 含量为 38.89%,马山前胡的 GC_{all} 含量最高,达到 39.02%;而 *P. chujaense* 的 GC_{all} 含量最低,为 38.81%。在 GC₁ 含量中,马山前胡具有最高值,为 47.97%;*P. hakuunense* 则最低,为 47.50%。对于密码子 GC₂ 含量,马山前胡以 39.06% 位居最高;*P. chujaense* 的 GC₂ 含量最低,为 38.89%。马山前胡的 GC₃ 值最高,为 30.22%;*P. chujaense* 的 GC₃ 值最低,为 30.02%。20 种前胡属植物的 GC_{all} 以及 GC₁、GC₂、GC₃ 含量均低于 50%,且由大到小顺序为:GC₁、GC₂、GC_{all}、GC₃ 和 GC_{3s},这表明 20 种前胡属植物的叶绿体基因组富含 A/T 碱基,密码子表现出对 A/T 碱基的偏好,并且 GC 含量在不同密码子位置上的分布不均匀,各位置之间存在显著差异。ENC 的理论值范围为 20~61,ENC \leq 35 的基因或基因组通常有较强的密码子偏好性。ENC 值越高则代表密码子偏好性越弱,而较高的 ENC 值表明存在稀有密码子富集现象,具有较低的基因表达水平和较弱的密码子偏好性。20 种前胡属植物的叶绿体基因组的 ENC 值范围为 49.73~49.90,均大于 45,表明其密码子使用偏好性相对较弱。

2.2 ENC-plot 分析

ENC-Plot 是一种有效的分析密码子使用模式

的方法,用于研究 GC_{3s} 对前胡属密码子使用偏好性的影响^[23]。本文中,20 种前胡属植物的 ENC 值和 GC_{3s} 值呈现出相似的分布模式(见图 2),表明这些物种的密码子偏好模式具有一定的相似

性;大部分基因位于标准曲线附近,与 ENC 期望值的差异较小,表明前胡属植物叶绿体基因组的密码子使用偏好性受到自然选择和突变压力的影响。

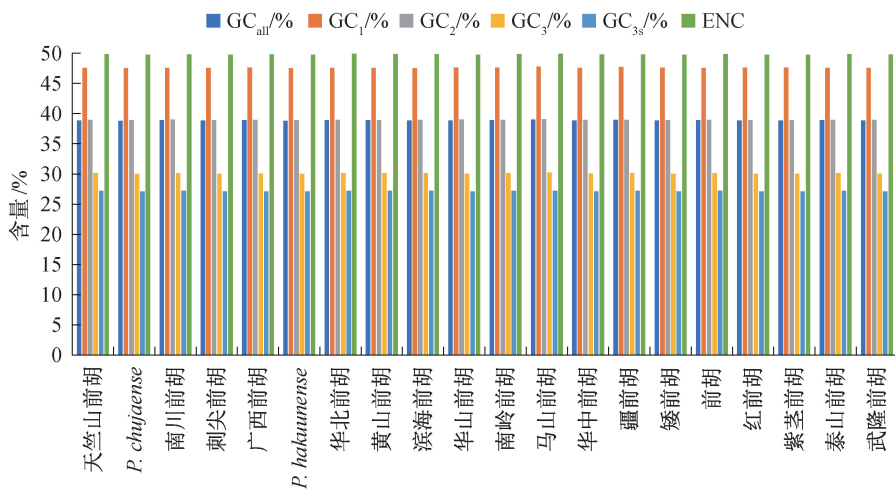
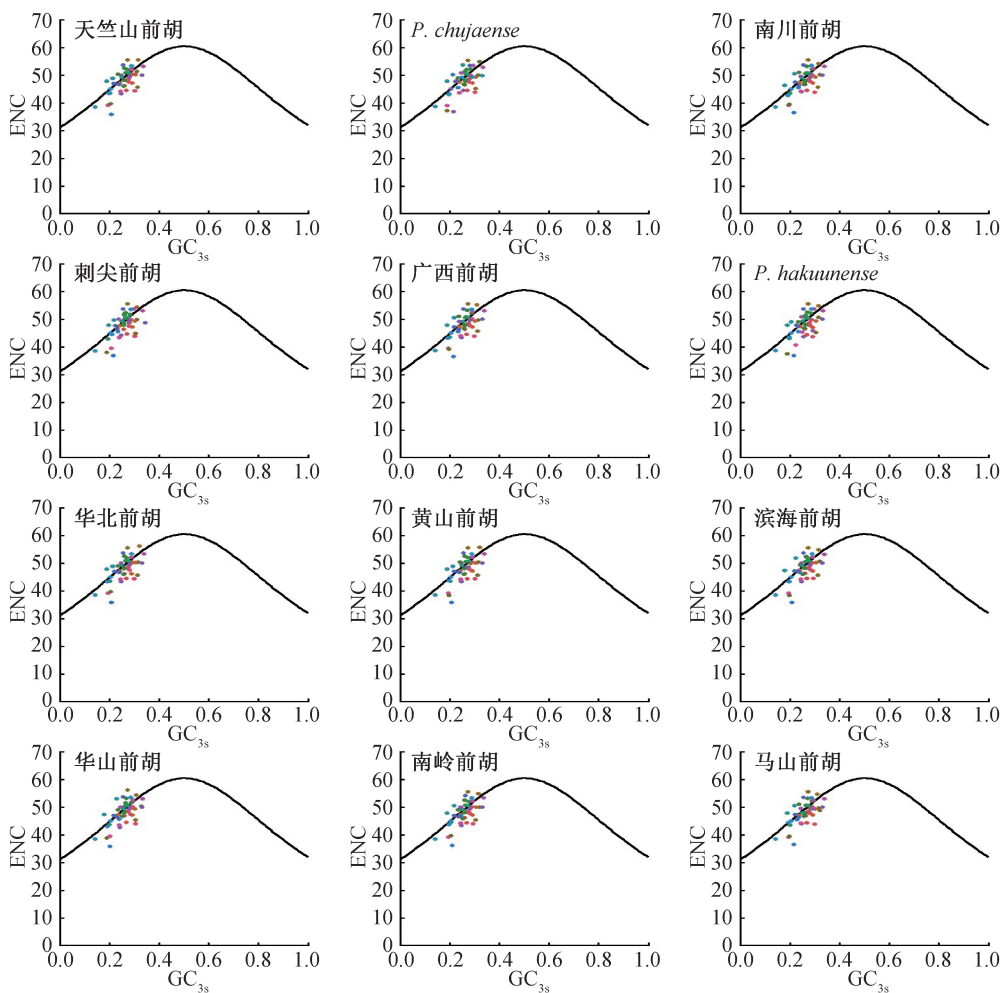


图 1 前胡属叶绿体基因组密码子使用偏好性基本参数各含量的柱状图

注:图中各物种密码子使用偏好性基本参数含量从左往右依次为 GC_{all} 、 GC_1 、 GC_2 、 GC_3 、 GC_{3s} 、ENC。



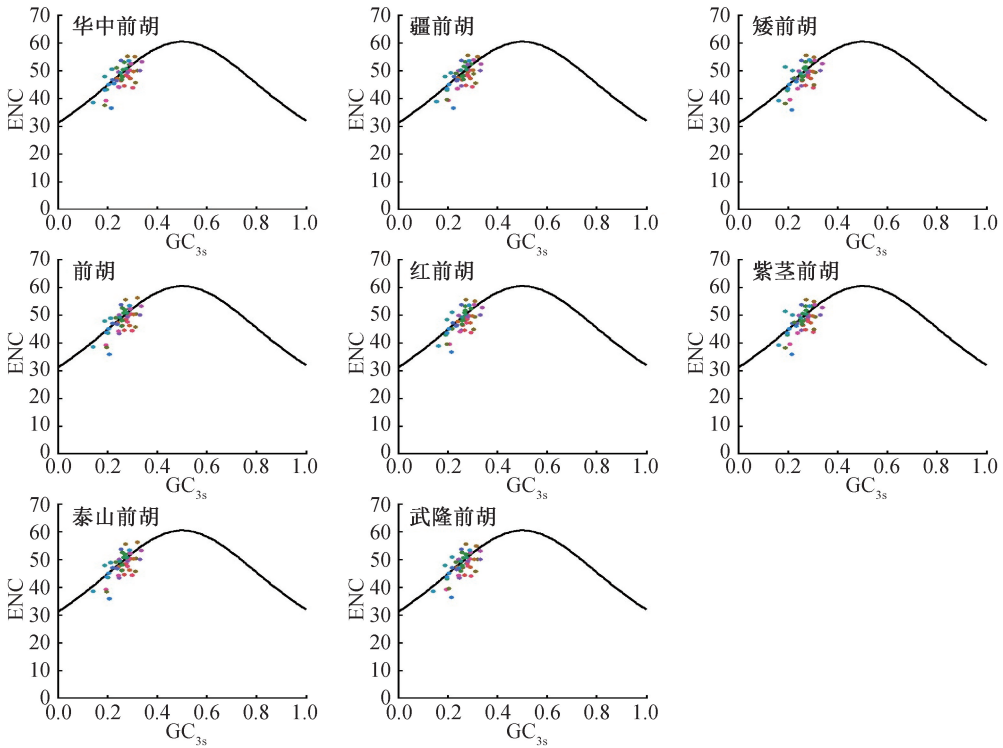


图2 前胡属叶绿体基因组 ENC-plot

注:图中各颜色的点表示不同的基因,曲线为标准曲线;利用R语言中的ggplot2包绘制ENC-Plot。

ENC 比值频率分布显示了 ENC 实际值与期望值之间的差异如图 3 所示,从图中可以看出:20 个前胡属植物的分布频率基本相似。大部分基因位于 $-0.05 \sim 0.05$, 占 57%, 表明这

些基因的 ENC 实际值与预期值相近,这些基因密码子偏性受到的突变压力大于选择压力,在其他区间分布的基因为 43%, 突变压力为主要影响因素。

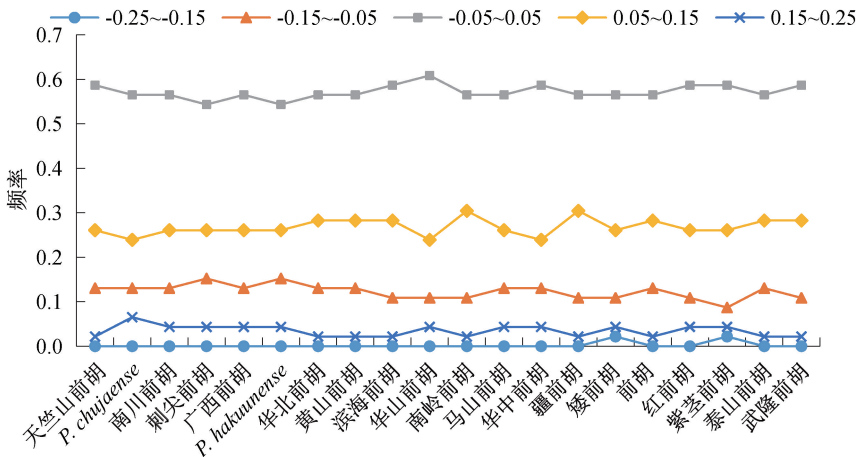


图3 前胡属叶绿体基因组 ENC 比值频率分布图

2.3 PR2-plot 分析

在无选择压力的条件下,密码子处于不同位置的突变是随机发生的,密码子碱基偏倚分析可以揭示 A、T、C、G 四种碱基的使用频率差异。从图 4 可以看出,20 种前胡属植物的基因分布相似,在 4 个象限中,基因分布不均衡,多数集中在下半区域,且右下象限的基因数量略少于左下象限。第三位碱基

中,C 的占比高于 G,T 的占比高于 A,C+T 的占比高于 G+A,第三位使用嘧啶的频率比嘌呤高。有极少数的基因偏离中心位置较远,*psbA* 基因的 $G_3/(G_3+C_3)$ 值显著高于其他基因,说明受到了更强的选择作用,与叶绿体内高丰度 tRNA 匹配,因而提高了翻译速率,说明了前胡属密码子使用偏好性主要受到自然选择的影响,并受到其他因素的综合影响。

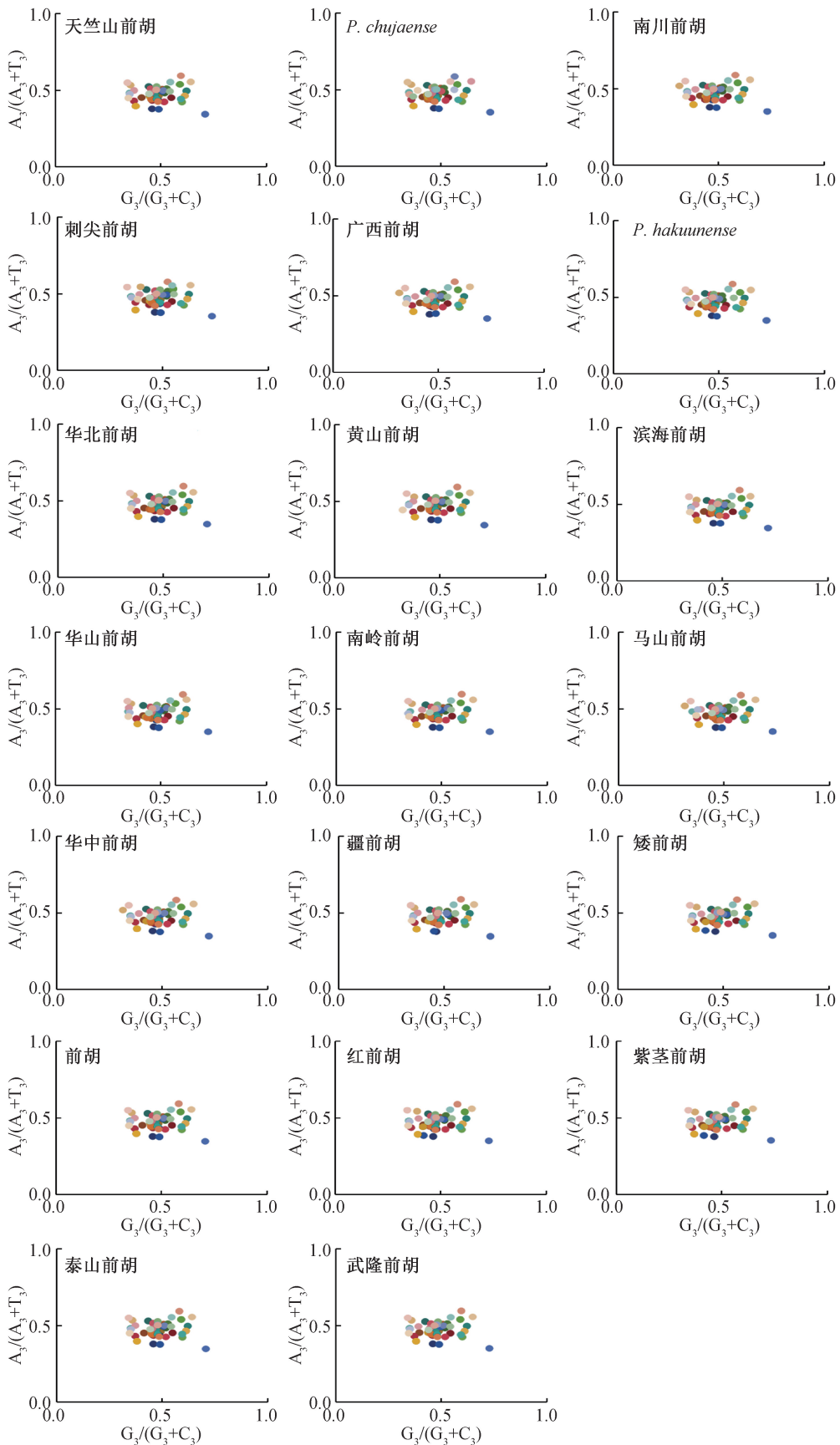


图 4 前胡属叶绿体基因组 PR2-plot

注:图中各颜色的点表示不同的基因。

2.4 中性绘图分析

中性图分析通过分析密码子不同位置的碱基组

成,以揭示突变和自然选择对密码子使用偏好性的影响^[24]。图5显示,20种前胡属植物叶绿体基因组

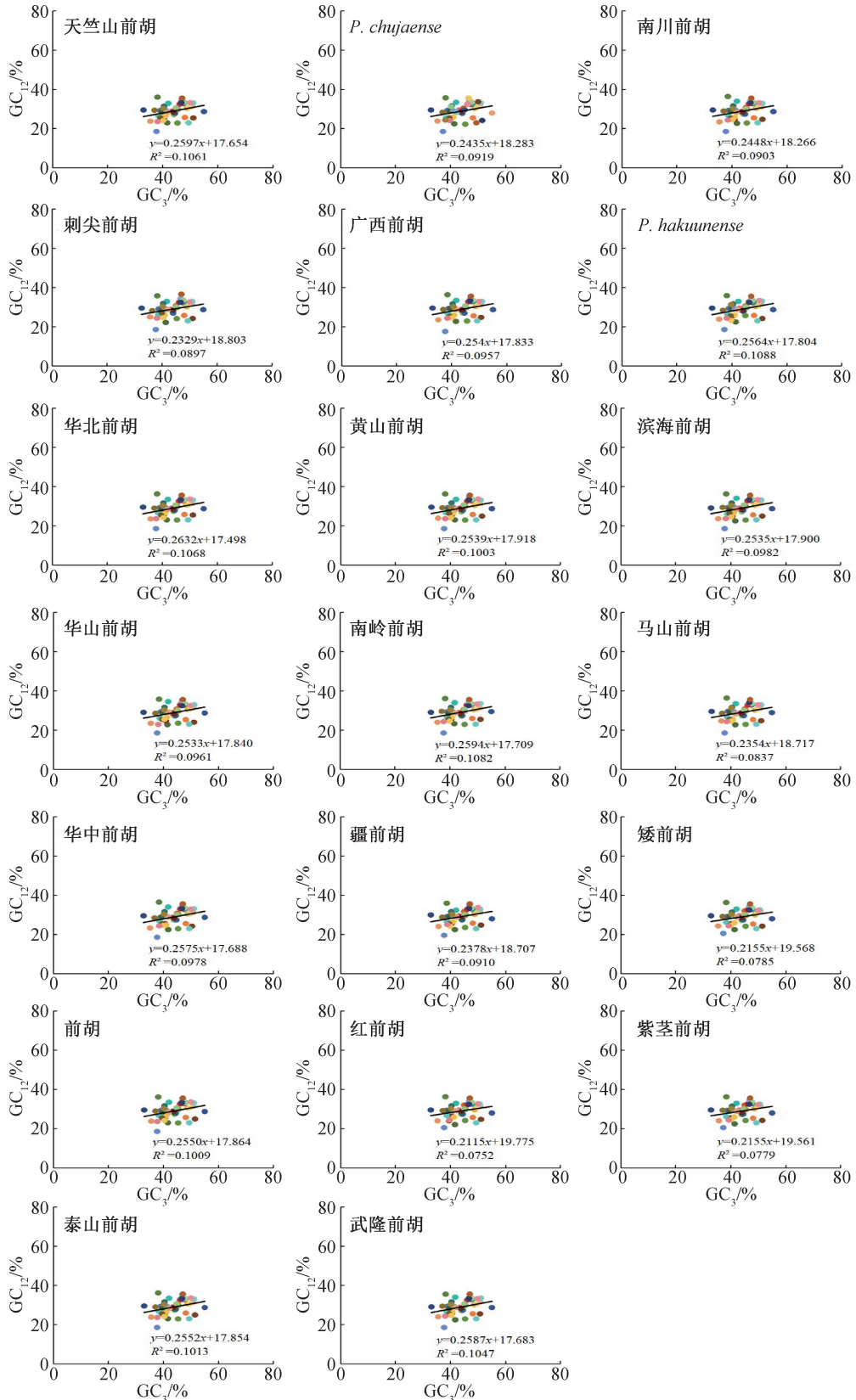


图5 前胡属叶绿体基因组中性绘图

注:图中各颜色的点表示不同的基因,R²为相关系数的平方。

的 GC₁₂ 值为 43.21% ~ 43.43%，而 GC₃ 值为 30.02% ~ 30.22%。GC₁₂ 和 GC₃ 的分布范围较窄,表明密码子不同位置的 GC 含量差异较小;GC₃ 值与 GC₁₂ 值之间的相关系数 R² 值较小,为 0.0720 ~ 0.1088,且二者之间无显著正相关性(P>0.05),说明密码子第三位的突变模式与前两位不同;回归线斜率范围为 0.2096~0.2632,表明突变压力的影响仅占 20.96% ~ 26.32%,而自然选择则占据了 73.68%~79.04%,判断在前胡属植物叶绿体基因组中,自然选择为主要驱动力,突变压力和其他因素的影响较小^[25]。

2.5 同义密码子分析

当 RSCU 为 1 时,表示不存在密码子偏好;当 RSCU>1 时,表示存在较强的密码子偏好;当 RSCU<1 时,则表示密码子偏好较弱^[26]。在 20 种前胡属植物的叶绿体基因组中,RSCU 表现出相似性,其变化范围不超过 0.03(见图 6),这说明前胡属植物叶绿体基因组在进化过程中相对保守。在前胡属植物的同义密码子中,UUA 的 RSCU 值最高,范围为 1.90~1.94。除去三个终止密码子、ATG 和

TGG 外,20 种前胡属植物的叶绿体基因组有 29 个高频密码子的 RSCU 值大于 1,其中 28 个以 A/U 结尾,如 UUU、UCU、UAU 等,这表明前胡属叶绿体基因组的高频密码子偏好以 A/U 结尾。

2.6 最优密码子分析

在最优密码子分析中,基于 ENC 建立了高低表达基因库。表 2 显示,20 种前胡属植物满足 RSCU>1 且 ΔRSCU≥0.08 的最优密码子为 18~22 个。其中,紫茎前胡和矮前胡有 18 个最优密码子、红前胡有 19 个、*P. hakuunense* 有 20 个。天竺山前胡、*P. chujaense*、广西前胡、华北前胡、黄山前胡、华山前胡、南岭前胡、疆前胡、前胡和泰山前胡各有 21 个最优密码子。南川前胡、刺尖前胡、滨海前胡、马山前胡、华中前胡和武隆前胡有 22 个最优密码子。20 种前胡属植物含有 17 个相同的最优密码子,分别是 UUU、UUA、UGU、UCU、UAU、GUU、GUA、GGU、GGA、GAA、CUU、CGA、CAU、CAA、AUU、ACA 和 AAA,它们均以 A/U 结尾,表明前胡属叶绿体基因组偏好 A/U 结尾的密码子。

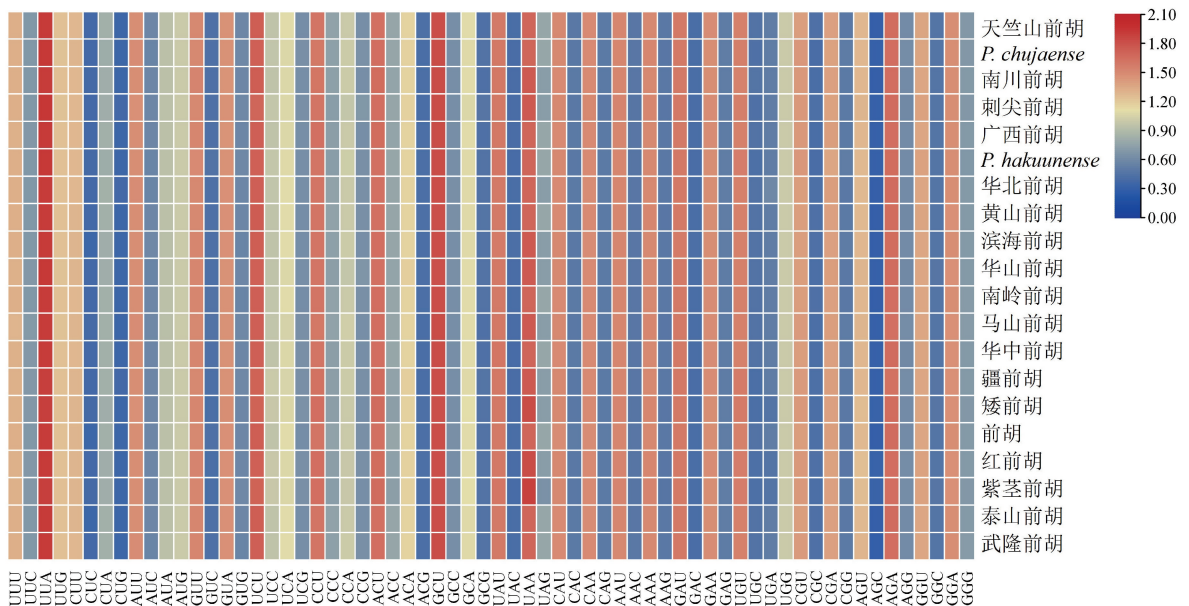


图 6 前胡属叶绿体基因组同义密码子

表 2 前胡属叶绿体基因组最优密码子

物种	同义密码子
天竺山前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
<i>P. chujaense</i>	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
南川前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CCA CAU CAA AUU ACU ACA AAA

表2(续)

物种	同义密码子
刺尖前胡	UUU UUA UGU UCU UCA UAU UAA GCU GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
广西前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
<i>P. hakuunense</i>	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GAA CUU CGA CAU CAA AUU ACU ACA AAA
华北前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
黄山前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
滨海前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CCA CAU CAA AUU ACU ACA AAA
华山前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
南岭前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
马山前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CCA CAU CAA AUU ACU ACA AAA
华中前胡	UUU UUA UGU UCU UAU UAA GCU GUU GUA GGU GGA GCA GAA CUU CGA CCA CAU CAA AUU ACU ACA AAA
疆前胡	UUU UUA UGU UCU UCA UAU GUU GUA GGU GGA GCA GAA CUU CGA CCA CAU CAA AUU ACU ACA AAA
矮前胡	UUU UUA UGU UCU UAU GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACA AAA
前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
红前胡	UUU UUA UGU UCU UCA UAU GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACA AAA
紫茎前胡	UUU UUA UGU UCU UAU GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACA AAA
泰山前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CAU CAA AUU ACU ACA AAA
武隆前胡	UUU UUA UGU UCU UCA UAU UAA GUU GUA GGU GGA GCA GAA CUU CGA CCA CAU CAA AUU ACU ACA AAA

注:灰色标注的密码子表示 20 种前胡属植物相同的最优密码子,而黑色标注密码子则表示在前胡属植物中存在差异的最优密码子。

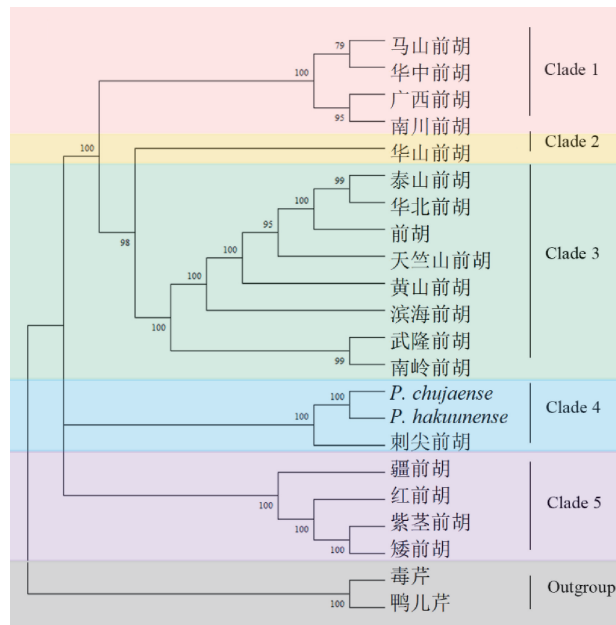
2.7 系统发育分析

基于 20 种前胡属植物的 46 个 CDS 序列和叶绿体全基因组序列,通过 MEGA7 软件采用最大似然法,进行 1000 次 bootstrap 重复分析,构建系统发育树,结果如图 7 所示。图 7 表明,ML 系统发育树中,大多数节点获得了较高的支持值,支持 20 种前胡属植物被划分为五个支系;基于 CDS 构建的系统发育树强有力地支持了支系 1 与(支系 2 + 支系 3)之间的姐妹群关系,同时支系 4 和支系 5 各自形成了独立的单系群;基于叶绿体全基因组的进化树支持以下拓扑结构:(支系 2 +(支系 3 +(支系 4 + 支系 5))),其中支系 1 被确定为一个独立的单系群。

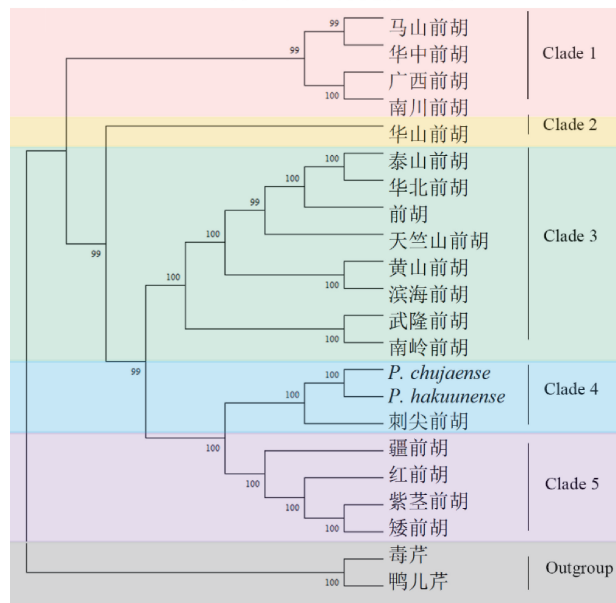
3 讨论

密码子各位置 GC 含量是受突变压力与选择共同作用的结果,氨基酸的第三位点主要反映突变压的大小,而第一和第二位的突变会改变氨基酸,并受到选择的限制。本文中,20 种前胡属植物的 GC 含量(GC_{all} 、 GC_1 、 GC_2 、 GC_3)均低于 50%,含量高低由大到小 GC_1 、 GC_2 、 GC_{all} 、 GC_3 和 GC_{3s} ,表明密码子第三位碱基对 A/T 的偏好更强,与多数双子叶植物^[17,23]的叶绿体基因组特征一致,说明高等植物在叶绿体基因 GC 含量及密码子使用上具有相似性,可能反映了一种保守的进化策略。通过 ENC-plot、

PR2-plot 及中性图分析发现,自然选择是前胡属植物密码子使用偏好性的主要驱动力,而突变压力的作用相对有限。自然选择通过翻译效率、mRNA 稳定性合蛋白功能来影响密码子的使用偏好性。自然选择可能倾向于高频密码子与优势 tRNA 的匹配,以提升光合相关基因(如 *psbA*、*rbcL*)的快速表达,适应多变环境。光合基因的密码子偏好性显著高于非光合基因,进一步支持了自然选择对其功能核心性的强化作用。本文鉴定的 29 个高频密码子,其中 18~22 个为最优密码子,且多数以 A/U 结尾,这与双子叶植物的叶绿体偏好模式高度重合,可能暗示了一种保守的翻译优化机制。



(a) 基于46个CDS序列的ML树



(b) 基于叶绿体全基因组序列的ML树

图 7 20 种前胡属植物的最大似然树

本文基于前胡属植物的 46 个 CDS 序列及全叶绿体基因组构建了高支持度的系统发育树,揭示了 5 个主要支系的划分,并结合形态特征与地理分布探讨了其进化驱动力。从地理分布方面来看,分布于中国西南地区的马山前胡、华中前胡、广西前胡和南川前胡被划分为支系 1。这些物种具有白色花瓣,且叶片下表面有叶脉隆起的形态特征,可能与该区域的弱光照、高湿度的生存环境相关。支系 5(疆前胡、红前胡、紫茎前胡和矮前胡),主要分布在青藏地区,海拔超过 1000 m,其中矮前胡能够在海拔 3500~3800 m 的高海拔地区繁衍生息,它们的椭圆形果实形态可能通过减少表面积来降低水分蒸发,因而能适应干旱寒冷环境。由于高海拔环境的低温、强紫外线等因素,A/U 富集的密码子可能通过优化 tRNA 适配性提升光合基因的翻译效率,增强了植物的生存适应性。支系 4 中的韩国特有种(*P. chujaense*、*P. hakuunense*)与中国东北地区的刺尖前胡聚为一类,暗示东亚地区的地理隔离(如长白山脉、朝鲜海峡)可能驱动了物种分化。从形态特征方面来看,支系 2 的华山前胡作为单型支系,其叶片为羽状复叶,果实为背腹压扁的分果,表面密布叉状粗糙刚毛,背棱稍微突出,每条棱槽内有 1 个油管,合生面有 2 个油管。羽状复叶可能通过增加叶片表面积提升光合效率,反映其对开阔山地生境的适应策略。相比之下,支系 3(如泰山前胡、华北前胡等)根茎粗壮,且花序梗有极短绒毛,可能与其分布的中国中部石灰岩地貌相关。粗壮根系有助于植株在石缝中生长,绒毛则减少蒸腾作用。这些形态特征的差异,提示了前胡属植物可能通过叶片结构、果实形态等特征变化分布于多样化生态位。未来研究可通过比较不同生态位前胡属植物的叶绿体基因组,结合转录组分析,揭示密码子偏好性与生态适应性的具体关联机制。

4 结 论

本文以 20 种前胡属植物的叶绿体基因组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,得出以下结论:

a) 20 种前胡属植物的 GC_{all} 含量以及 GC_1 、 GC_2 和 GC_3 含量均低于 50%,且含量从高到低依次为 GC_1 、 GC_2 、 GC_{all} 、 GC_3 、 GC_{3s} 。最优密码子数量为 18~22 个,且这些密码子倾向于以 A/U 结尾。前胡属植物有 17 个相同的最优密码子,分别是 UUU、UUA、UGU、UCU、UAU、GUU、GUA、

GGU、GGA、GAA、CUU、CGA、CAU、CAA、AUU、ACA 和 AAA。

b)与突变压力等因素作用相比较,自然选择是影响前胡属叶绿体基因组密码子使用偏好性的主要因素。

c)基于前胡属植物的 46 个 CDS 序列和完整叶绿体基因组构建的进化树高度相似,揭示了前胡属蛋白质编码片段在遗传和保守表型中的重要性。

关于密码子偏好性的研究有待进一步深入:

a)可以将密码子偏好性与 DNA 相关技术结合,用于药用植物的鉴定和质量控制;b)密码子偏好性与植物次生代谢产物的合成密切相关,须进一步研究植物在不同环境条件下的次生代谢产物积累;c)结合转录组学数据,可以研究密码子偏好性与基因表达水平之间的关系,揭示植物在不同发育阶段或环境胁迫下的基因表达调控机制。这些方面的研究可以进一步促进植物科学和农业的发展。

参考文献:

[1] 单人焯, 余孟兰, 刘守炉. 中国植物志[M]. 北京: 科学出版社, 1992: 123-175.

[2] Song B, Liu C, Xie D, et al. Plastid phylogenomic analyses reveal the taxonomic position of *Peucedanum franchetii* [J]. *Plants*, 2023, 12(1): 97.

[3] Uy N P, Kim H, Ku J, et al. Regional variations in *Peucedanum japonicum* antioxidants and phytochemicals [J]. *Plants*, 2024, 13(3): 377.

[4] 张雪梅. 前胡属植物的系统学研究进展[J]. *黑龙江农业科学*, 2012(8): 156-158.

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

[6] Lin Y, Cheng X, Yang D, et al. Advances in chloroplast expression of recombinant proteins in higher plants [J]. *Chinese Journal of Biotechnology*, 2018, 34(5): 631-643.

[7] Chen M M, Zhang M, Liang Z S, et al. Characterization and comparative analysis of chloroplast genomes in five *Uncaria* species endemic to China [J]. *International Journal of Molecular Sciences*, 2022, 23: 11617.

[8] 王平, 赵琴, 严少普, 等. 柴胡属 5 种植物叶绿体基因组密码子偏好性分析[J]. *分子植物育种*, 2023, 1-13.

[9] 黄祥, 楚光明, 郑新开, 等. 睡莲属叶绿体基因组密码子偏好性及系统发育分析[J]. *中国农业科技导报*, 2022, 24(4): 75-84.

[10] 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*, 2021, 27(12): 2727-2737.

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

[12] Liu C K, Lei J Q, Jiang Q P, et al. The complete plastomes of seven *Peucedanum* plants: Comparative and phylogenetic analyses for the *Peucedanum* genus [J]. *BMC Plant Biology*, 2022, 22(1): 101.

[13] 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.

[14] Wang Z, Xu B, Li B, et al. Comparative analysis of codon usage patterns in chloroplast genomes of six *Euphorbiaceae* species [J]. *The Journal of Life and Environmental Sciences*, 2020, 8: e8251.

[15] Chen X, Chang Q, Xia P G, et al. The complete chloroplast genome of *Clematis henryi* var. *ternata* (Ranunculaceae) [J]. *Mitochondrial DNA Part B*, 2021, 6(4): 1319-1320.

[16] Song H, Liu J, Chen T, et al. Synonymous codon usage pattern in model legume *Medicago truncatula* [J]. *Journal of Integrative Agriculture*, 2018, 17(9): 2074-2081.

[17] 蒋瑞平, 赵辰晖, 李文杰, 等. 豆科植物 IPI 基因密码子偏好性 [J]. *浙江农业学报*, 2022, 34(6): 1114-1123.

[18] 石乃星, 谢平选, 李立, 等. 滇黄精叶绿体全基因组序列及其密码子使用偏性分析 [J]. *热带亚热带植物学报*, 2022, 30(3): 336-348.

[19] 黄松, 张辉, 邱林, 等. 白蜡树叶绿体基因组结构特征及系统发育分析 [J]. *江苏农业科学*, 2024, 52(16): 36-44.

[20] 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.

[21] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability [J]. *Molecular Biology and Evolution*, 2013, 30(4): 772-780.

[22] Wen J, Xie D F, Price M, et al. Backbone phylogeny and evolution of *Apioideae* (Apiaceae): New insights from phylogenomic analyses of plastome data [J]. *Molecular Phylogenetics and Evolution*, 2021, 161: 107183.

[23] 丁祥青, 陈丝雨, 陈佳婷, 等. 11 种金花茶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.

[24] Shen Z, Gan Z, Zhang F, et al. Analysis of codon usage patterns in citrus based on coding sequence data [J]. *BMC Genomics*, 2020, 21(Suppl 5): 234.

[25] Yengkhom S, Uddin A, Chakraborty S. Deciphering codon usage patterns and evolutionary forces in chloroplast genes of *Camellia sinensis* var. *assamica* and *Camellia sinensis* var. *sinensis* in comparison to *Camellia pubicosta* [J]. *Journal of Integrative Agriculture*, 2019, 18(12): 2771-2785.

[26] 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.

团队介绍

药用植物次生代谢团队专注于药用植物质量评价及其次生代谢调控分子机制方面的研究工作,对药用植物规范化栽培、质量标准、代谢工程和次生代谢调控的信号转导机制进行了系统研究。王瑞红,浙江植物学会会员,药用植物次生代谢团队成员;2010年9月—2015年7月在浙江大学生命科学学院硕博连读,2015年10月至今在浙江理工大学生命科学与医药学院担任副教授。团队主持包括国家自然科学基金在内的重大科研项目30余项,发表相关论文400余篇,企业委托项目30余项,指导建立中药材基地20余个,取得了显著的社会、环境和经济效益。团队近年来对石斛、丹参、黄精、三叶青、前胡、覆盆子等中药材栽培、种质资源评价及次生代谢产物合成的分子机制等进行了深入的研究,成功地将组学技术分析方法应用于多个物种的研究中,实验方法及体系较为成熟。

(责任编辑:张会巍)