江苏农业学报(*Jiangsu J.of Agr.Sci.*),2025,41(9):1704-1713
http://jsnyxb.jaas.ac.cn

郑梦迪,陈斯琪,尹念念,等. 黄精属植物叶绿体基因组系统发育与密码子使用偏好分析[J]. 江苏农业学报,2025,41(9):1704-1713.
doi:10.3969/j.issn.1000-4440.2025.09.005

黄精属植物叶绿体基因组系统发育与密码子使用偏好分析

郑梦迪, 陈斯琪, 尹念念, 冯言熙
(西安医学院,陕西 西安 710021)

摘要: 为探索黄精属植物叶绿体基因组的进化、密码子选择倾向和相关影响要素。本研究利用 MISA、IRscope、mVISTA 等在线工具分析 8 种黄精属植物叶绿体基因组特征;利用 CodonW 1.4.2、CUSP、Excel、MEGA11 等软件进行密码子偏好性和系统发育分析。结果表明,碱基 A 和碱基 T 组成的单核苷酸重复序列被频繁使用;叶绿体基因组界限稳定,区域组成差异小。密码子中 GC 含量从第 1 碱基位到第 3 碱基位递减,末位偏好碱基 A 或碱基 U;叶绿体基因组中密码子的选择倾向受多因素共同作用,其中自然选择是主要原因。最终筛选出 8 种黄精属植物共有 9 个相同的最优密码子。系统发育研究结果揭示独花黄精和距药黄精与黄精属其他 6 种植物的关系较远。本研究结果为黄精属植物的系统发育分析提供了支持,并为黄精药材的鉴定、品种优化及遗传育种工作奠定了理论基础。

关键词: 黄精属; 叶绿体基因组; 密码子使用偏好; 系统发育分析
中图分类号: S567 **文献标识码:** A **文章编号:** 1000-4440(2025)09-1704-10

Phylogenetic analysis and codon bias in the chloroplast genomes of *Polygonatum* species

ZHENG Mengdi, CHEN Siqi, YIN Niannian, FENG Yanxi
(*Xi'an Medical University*, *Xi'an* 710021, *China*)

Abstract: To investigate the chloroplast genome phylogeny, codon usage bias patterns and related influencing factors of plants in the genus *Polygonatum*, the MISA, IRscope, and mVISTA were used to analyze chloroplast genome characteristics of eight plants in the genus *Polygonatum*. CodonW 1.4.2, CUSP, Excel, and MEGA11 were utilized to perform codon bias and phylogenetic analysis. The results indicated that mononucleotide repeat sequences composed of bases A and T were frequently utilized. The boundaries of the chloroplast genomes were highly conserved, exhibiting minimal variation in genomic composition. The GC content in codons progressively declined from the first to the third codon position, with a pronounced bias toward base A or U at the third position. Codon usage bias in chloroplast genomes was shaped by multiple factors, with natural selection serving as the primary driving force. Ultimately, nine identical optimal codons were identified as common to all eight studied *Polygonatum* species. Phylogenetic analysis revealed that *Polygonatum hookeri* and *Polygonatum franchetii* exhibited a distant relationship to the other six studied *Polygonatum* species. This study provides crucial support for phylogenetic analyses within the *Polygonatum* genus, while also laying a theoretical foundation for the authentication of *Polygonatum* medicinal materials, cultivar improvement, and genetic breeding efforts.

Key words: *Polygonatum*; chloroplast genome; codon bias; phylogenetic analysis

收稿日期: 2024-11-06
基金项目: 陕西省自然科学基础研究计划项目(2021JQ-782);西安医学院科研能力提升计划项目(2022NLTS084、2024NLTS1-22)
作者简介: 郑梦迪(1988-),女,陕西咸阳人,博士,副教授,主要从事药用植物研究。(E-mail)670677492@qq.com

传统药材黄精素有“仙人余粮”和“百岁草”的美誉。《名医别录》首次记载黄精“味甘平无毒,能补中益气、祛风湿、安五脏”[1],而《抱朴子内篇》指出“花效优于实,实效优于根,但花不易得”[2],可见

在当时黄精的花、果实和根茎皆可入药,而现在仅取干燥根茎入药[3]。经考证古籍中记载的可以当药材使用的“兖州黄精”、“永康军黄精”实为现代命名的多花黄精(*Polygonatum cyrtonema*),“荆门军黄精”则为现代的独花黄精(*Polygonatum hookeri*)。“滁州黄精”、“解州黄精”和“相州黄精”等均为轮生叶形态[4],表明古代有多种黄精属植物被用作药材,而中华人民共和国药典将黄精的正宗来源限定为滇黄精(*Polygonatum kingianum* Collett & Hemsl.)、黄精(*Polygonatum sibiricum* Redouté)和多花黄精(*Polygonatum cyrtonema*)3 种[3]。

密码子负责识别和传递重要的遗传信息。除了蛋氨酸和色氨酸各有一个密码子,其他氨基酸由多个同义密码子编码[5]。在物种的长期进化过程中,由于自然选择和适应性演化,某些同义密码子被频繁选择和使用,从而形成了所谓的密码子偏好性。生物遗传信息传递中的密码子偏好性可通过分析确定,有助于优化基因效率和产物准确性,预测基因功能和表达,并提供科学依据探索物种间关系和遗传进化[5-6]。研究发现,物种间遗传距离与密码子偏好性相关,遗传距离近的物种在密码子偏好上更相似[7],且密码子偏好性与基因表达水平密切相关[8]。

叶绿体的遗传信息明显表现为较低的核苷酸替代率和单亲遗传特性[9]。尽管已有部分黄精属植物的叶绿体基因组序列被公开,但关于其密码子偏好性的研究较少[10-11]。为此,本研究针对黄精属 8 种药用植物,深入研究其叶绿体基因组密码子偏好,并分析影响因素。通过比较基因组、系统发育和结构分析,旨在为黄精属植物的分子育种和进化研究提供理论支持。

1 材料与方法

1.1 数据材料的获取

自美国国家生物技术信息中心[GenBank(https://www.ncbi.nlm.nih.gov/)]获取黄精[*Polygonatum sibiricum* Redouté(OQ532972.1)]、棒丝黄精[*Polygonatum cathcartii* Baker(OP428726.1)]、滇黄精[*Polygonatum kingianum* Collett & Hemsl.(OR992389.1)]、点花黄精[*Polygonatum punctatum* Royle ex Kunth(OQ532974.1)]、距药黄精[*Polygonatum franchetii* Hua(OP615253.1)]、独花黄精[*Polygonatum hookeri* Baker(OR656539.1)]、长梗黄精[*Polygonatum filipes* Merr. ex C. Jeffrey & McEwan(OR656540.1)]、五叶黄精[*Polygonatum acuminatifolium* Kom.(OP428729.1)]的完整叶绿体基因组。利用 CPGAVAS 2 在线软件对核苷酸序列进行注释,删除重复的以及长度不足 300 bp 的序列,人工筛选出起始密码子为 ATG,终止密码子为 TAA、TGA 或 TAG 的序列,得到 8 种黄精植物叶绿体的完整编码区序列(CDS)416 条。

1.2 叶绿体基因组特征分析

采用 MISA(https://webblast.ipk-gatersleben.de/misa)[12]工具鉴定 8 种黄精属植物叶绿体基因组中的简单重复序列(SSR)。使用 IRscope[13]可视化工具绘制 IR 边界对比图,以便对比叶绿体基因组结构的边界差异。使用 mVISTA[14]工具进行叶绿体基因组的全序列比对,分析变异情况。

1.3 密码子使用偏好分析

使用 CodonW 1.4.2 和 EMBOSS 的 CUSP 程序对 416 条编码序列(CDS)进行密码子使用指数分析,衡量密码子使用偏好和基因表达效率。以密码子第 3 位上的 GC 含量(GC_3)为横坐标、GC_{12}(GC_1 和 GC_2 的平均值,GC_1:密码子第 1 位上的 GC 含量;GC_2:密码子第 2 位上的 GC 含量)为纵坐标绘制散点图进行相关性分析[15]。以 $G_3/(G_3+C_3)$ 为横轴,$A_3/(A_3+T_3)$ 为纵轴绘制散点图,中心点代表 $A=T$ 和 $C=G$,分析第 3 位碱基的含量和关系[16]。依据有效密码子数(ENC)$= 2+GC_{3s}+29/[GC_{3s}^2+(1-GC_{3s})^2]$($GC_{3s}$:同义密码子第 3 位 GC 含量)[17]计算出有效密码子数的期望值,绘制以 GC_3 和理论 ENC 值为坐标的标准曲线图,并用表格列出 ENC 比值[ENC 比值=(ENC 预期−ENC 实际)/ENC 预期]。ENC 比值反映了密码子使用的偏差程度,ENC 比值越低,表示密码子偏好性越明显。同义密码子使用分析(COA)是一种评估相对同义密码子使用率($RSCU$ 值)变化和基因在多维空间分布的方法。通过 CodonW 1.4.2软件可以将基因映射到 58 个同义密码子的 58 维向量空间,并利用散点图(以 Axis1 和 Axis2 为坐标轴)进行可视化分析。分析密码子偏好性时,需关注基因在图像上的分布离散度和各轴的贡献率,以识别影响因素[18]。按照 ENC 比值选出最高 10%和最低 10%的基因,即本研究中 ENC 比值最高和最低各 5 条 CDS,构建出高、低表达库,并利用 $RSCU$ 值,计算出相应高低表达库的 $RSCU$ 差值($\triangle RSCU$)。$RSCU$ 值超过 1 的密码子定义为常用密码子,$\triangle RSCU$ 值≥0.08 的为高表达密

码子。密码子需同时满足这两个条件才被视为最佳密码子[19]。

1.4 系统发育分析

利用 MEGA 11 软件[20]基于精选的黄精属植物叶绿体基因编码区序列(CDS),通过邻接法构建系统发育树,并用1 000次 Bootstrap 测试其稳定性。

2 结果与分析

2.1 SSR 分析

从表 1 可知,在黄精属植物叶绿体基因组中发现了44~49 个 SSR 位点,涵盖单核苷酸、双核苷酸和三核苷酸 3 种重复序列。黄精有 2 种核苷酸重复序列,SSR 位点 44 个。棒丝黄精、滇黄精、距药黄精、五叶黄精有 3 种核苷酸重复序列,SSR 位点分别为 49 个、48 个、45 个、47 个。点花黄精、独花黄精、长梗黄精有 4 种核苷酸重复序列,SSR 位点分别为 46 个、47 个、44 个。且 8 种黄精属植物叶绿体基因组均为以碱基 A 和碱基 T 组成的单核苷酸重复序列次数最多,占比最大,这表明在碱基序列形成过程中碱基 A 和碱基 T 使用最频繁。

表 1 黄精属植物叶绿体基因组的 SSR 分析

Table 1 Types and amounts of SSRs in the chloroplast genomes of the *Polygonatum* species

SSR 类型	黄精	棒丝黄精	滇黄精	点花黄精	独花黄精	距药黄精	长梗黄精	五叶黄精
A/T	40	45	43	40	42	38	38	41
C/G	0	0	0	1	1	2	1	4
AT/AT	4	3	4	4	3	5	4	5
AAT/ATT	0	1	1	1	1	0	1	0
总计	44	49	48	46	47	45	44	47
A/T 占比(%)	90.91	91.84	89.58	86.96	89.36	84.44	86.36	87.23

2.2 LSC 区、SSC 区、IRa 区、IRb 区的收缩和扩张分析

8 种黄精属植物叶绿体基因组均呈现典型的四分体环状结构,具有 SSC 区、LSC 区、IRa 区、IRb 区 4 个区域,IRa 区、IRb 区的边界区域并非完全固定,这些边界区域时常经历着细微的变异[21]。本研究揭示了黄精属植物叶绿体基因组的 4 个界定区域。黄精的 JLB(LSC/IRb)位于 *rps19* 基因内,并扩展至 IRb 区 219 bp;其他 7 种黄精属植物的 JLB(LSC/IRb)则界定在 *rpl22* 基因与 *rps19* 基因之间。棒丝黄精、滇黄精、长梗黄精、五叶黄精的 JSB(IRb/SSC)缺失 *ycf1* 基因,黄精、点花黄精、独花黄精和距药黄精的 JSB 边界在 *ycf1* 基因和 *ndhF* 基因之间,8 种黄精属植物的 JSB 边界在 *ndhF* 基因内并向 SSC 区有 2 176~2 189 bp 的扩张。8 种黄精属植物的 JSA(SSC/IRa)边界均跨越 *ycf1* 基因,并向 IRa 区域有 883~896 bp 的扩张。在 JLA(IRa/LSC)边界处,除黄精的边界在 *rps19* 基因中间且向 LSC 区扩张了 26 bp,其余 7 种黄精属植物的 JLA 边界均在 *rps19* 基因与 *psbA* 基因之间。综合以上分析结果,进化过程中黄精属 8 种植物叶绿体基因组中 IR 区的变化幅度较小,叶绿体基因组较为保守。

2.3 叶绿体基因组序列变异分析

基于注释后的黄精叶绿体基因组对比其余 7 种黄精属植物叶绿体基因组比较分析。黄精属植物叶绿体基因组的 4 个区间构成相似;结合边界分析,8 种黄精属植物的 LSC 区和 SSC 区差异较大,其中 LSC 区变异程度最大,IR 区间最保守。非基因编码区变异程度较高,基因编码区较为保守,但 *ycf1*、*psbN*、*rpl16*、*ycf2* 等基因编码区变异程度较大,存在显著差异。

2.4 密码子的碱基组成

8 种黄精属植物叶绿体基因组的 CDS 数量分别为:黄精 52 条、棒丝黄精 51 条、滇黄精 51 条、点花黄精 51 条、独花黄精 53 条、距药黄精 53 条、长梗黄精 53 条、五叶黄精 52 条。GC 含量(GC_{all})的平均值分别为 38.28%、38.30%、38.33%、38.34%、38.28%、38.30%、38.30%、38.35%,含量基本一致。GC_3、GC_2、GC_1 含量分别在39.34%~46.88%、39.39%~39.62%、28.47%~28.89%,在 8 种黄精属植物的叶绿体基因组中,第一密码子位置的 GC 含量最高,第 3 位置 GC 含量最低,表明 G、C 在 3 个碱基位置上的含量不均

匀,更多分布在第 1 个碱基位置上,且密码子以碱基 A 或碱基 U 结尾的频率偏高。*ENC* 比值衡量密码子使用频率与同义密码子平均频率的偏差,超过 35 视为偏好性低。本研究中 8 种黄精属植物叶绿体基因组 *ENC* 比值平均为48.50%~48.02%,表明它们密码子使用偏好性较低。密码子适应指数(*CAI*)接近 1 时,表示基因密码子偏好性强,8 种黄精属植物叶绿体基因组 *CAI* 均为 0.17,表明叶绿体基因密码子偏好性较弱。密码子偏好指数(*CBI*)反映高表达密码子比例,0~1 范围内数值越大表示偏好性越强,小于 0 则表示偏好性更弱,本研究中 8 种黄精属植物叶绿体基因组 *CBI* 值均为-0.09,表明偏好性低于平均水平。最佳密码子频率(*Fop*)衡量最优密码子使用频率,0~1 范围内数值越大表示偏好性越强,0 表示未使用最优密码子,1 表示仅使用最优密码子,本研究中 8 种黄精属植物叶绿体基因组 *Fop* 均为 0.36,显示密码子偏好性较弱。

相关性分析结果显示(图 1),8 种黄精属植物的叶绿体基因组 GC_1 与 GC_2 均呈极显著相关($P<0.01$),GC_1、GC_2与 GC_3之间无显著相关性,表明密码子的第 1 位和第 2 位碱基组成相似,而第 1 位、第 2 位碱基组成与第 3 位碱基组成不同。8 种黄精属植物叶绿体基因组的 GC_3 与 GC_{3s}之间极显著相关($P<0.01$),显示同义突变主要发生在密码子第 3 位碱基。黄精叶绿体基因组的 *GC* 与 GC_3之间相关性不显著($P>0.05$),且其他 7 种黄精属植物叶绿体基因组的 *GC* 与 GC_3均呈极显著相关($P<0.01$),第 3 位碱基对 GC 含量有关键影响,且在 8 种黄精属植物叶绿体基因组中,GC_3 与 *ENC* 比值极显著相关($P<0.01$),显示第 3 位碱基显著影响密码子偏好性。*CAI*、*CBI* 和 *Fop* 指代了基因的表达水平,三者与 GC_1、GC_2、GC_3、GC_{3s}、GC_{12}两两之间呈现正相关关系,有的呈现出显著($P<0.05$)或极显著($P<0.01$)相关性,这显示了构成密码子的 G、C 碱基对基因表达具有影响。

2.5 中性绘图分析

黄精属 8 种植物叶绿体基因组密码子 GC_{12}为 32.64%~55.76%,GC_3 为21.50%~39.61%,说明碱基 G、C 多出现在密码子碱基位的前 2 位,第 3 位碱基 A、T 的使用频率高于碱基 G、C。基因位点几乎完全分布在对角线上半部分(图 2),且 8 种黄精属植物叶绿体基因组的 GC_3 与 GC_{12}回归系数均较小,范围为0.078 1~0.158 0,表示密码子使用偏好性受自然选择影响较大,受突变影响较小。

2.6 PR2-plot 分析

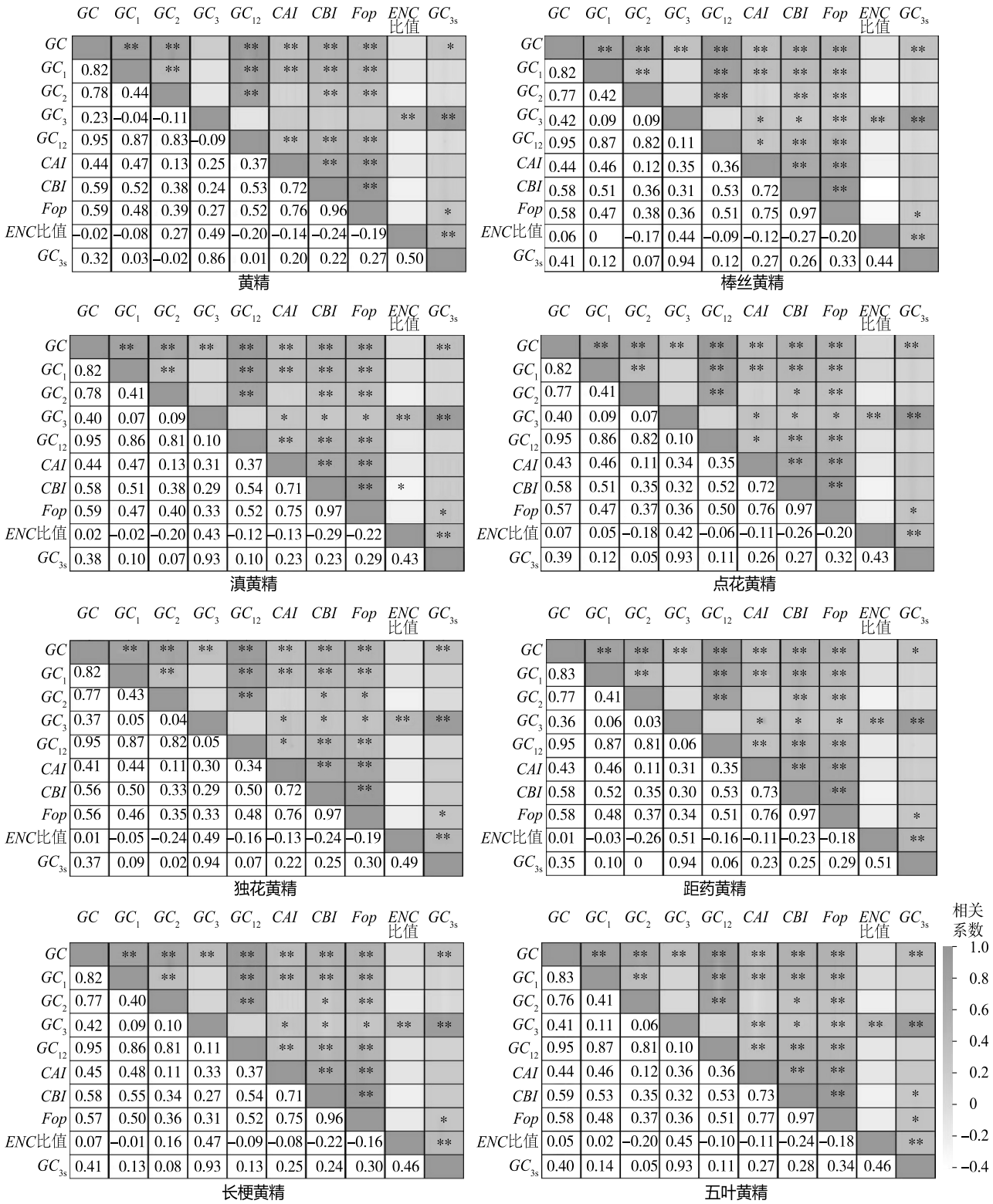
如图 3 所示,8 种黄精属植物叶绿体基因位点在 4 个区域的分布不均,较多基因位点沿中线以下分布,主要位于 $G_3/(G_3+C_3)<0.5$ 和 $A_3/(A_3+T_3)<0.5$ 的区域,$G_3/(G_3+C_3)$ 平均值为0.490 8~0.495 5,$A_3/(A_3+T_3)$为0.481 5~0.483 1,这说明在密码子的第 3 位上碱基 C 的使用频率高于碱基 G,碱基 T 的使用频率高于碱基 A。长梗黄精叶绿体的基因位点多在 $A_3/(A_3+T_3)<0.5$ 区域,在 $G_3/(G_3+C_3)=0.5$ 线两侧均匀分布,显示密码子第 3 位上碱基 T 多于 A,C 与 G 相当。这暗示长梗黄精叶绿体基因组的密码子偏好受突变压力、自然选择等因素的影响,突变压力影响较大。

2.7 *ENC*-plot 分析

从图 4、表 2 和表 3 可以看出,8 种黄精属植物的大多数叶绿体基因偏离 *ENC* 期望值标准曲线,大部分位于标准曲线的上方或附近;位于-0.050~0.050 区间的基因数量最多,占比为 58.82%~62.26%,即实际 *ENC* 与期望 *ENC* 接近;其中距药黄精叶绿体基因组中位于-0.05~0.05 区间的基因有 33 个(占比为 62.26%),五叶黄精、独花黄精和长梗黄精叶绿体基因组中位于-0.050~0.050 区间的基因有 32 个(占比分别为 61.54%、60.38%、60.38%),黄精、棒丝黄精和滇黄精叶绿体基因组中位于-0.050~0.050 区间的基因有 31 个(占比分别为 59.62%、60.78%、60.78%),点花黄精叶绿体基因组中位于-0.050~0.050 区间的基因有 30 个(占比为 58.82%)。位于曲线下部分的基因密码子偏好不仅受自然选择影响,更受突变压力影响。

2.8 对应分析

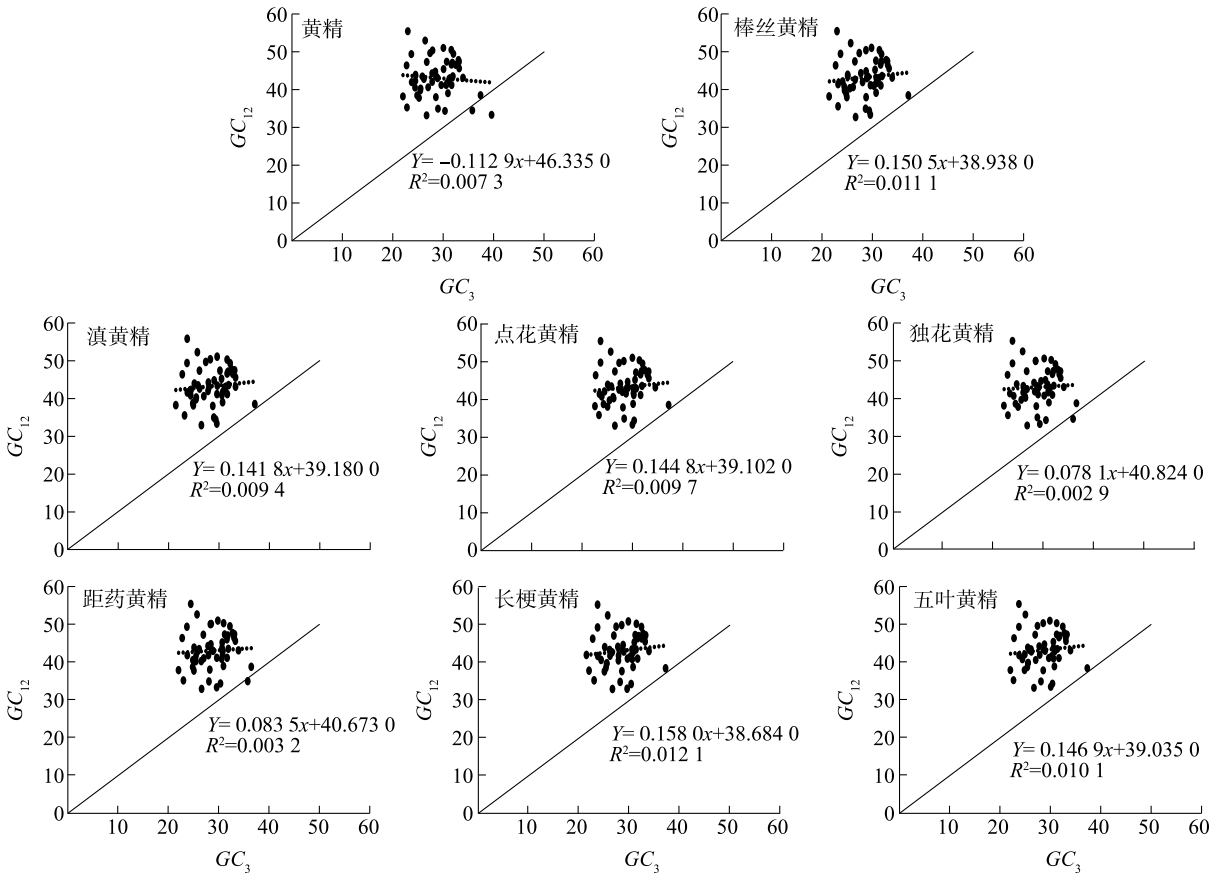
从图 5 可知,基因位点在轴 1、轴 2 上零散分布,8 种黄精属植物叶绿体基因密码子使用并未展现出明显的偏好性。8 种黄精属植物叶绿体基因前 4 主成分轴的累积变异贡献率分别为:35.15%、37.91%、35.44%、35.03%、35.70%、34.93%、34.48%、35.00%。其中第 1 轴可解释变异占比最高,分别为 10.29%、11.38%、10.51%、10.16%、10.17%、10.34%、9.99%、10.25%,表明第 1 轴对密码子偏好性的贡献率较大,但仅代表部分变异的总和。说明黄精属植物叶绿体基因的密码子偏好性受多因素影响。



GC:鸟嘌呤和胞嘧啶的含量;GC_1:密码子第 1 位上 GC 含量;GC_2:密码子第 2 位上 GC 含量;GC_3:密码子第 3 位上 GC 含量;GC_{12}:第 1、第 2 位上碱基 GC 平均含量;CAI:密码子适应指数;CBI:密码子偏好指数;Fop:最佳密码子频率;ENC 比值:有效密码子数;GC_{3s}:同义密码子第 3 位上 GC 含量;* 表示二者之间显著相关($P<0.05$),** 表示二者之间极显著相关($P<0.01$)。

图 1 黄精属植物叶绿体基因组中密码子使用特性的相关性分析

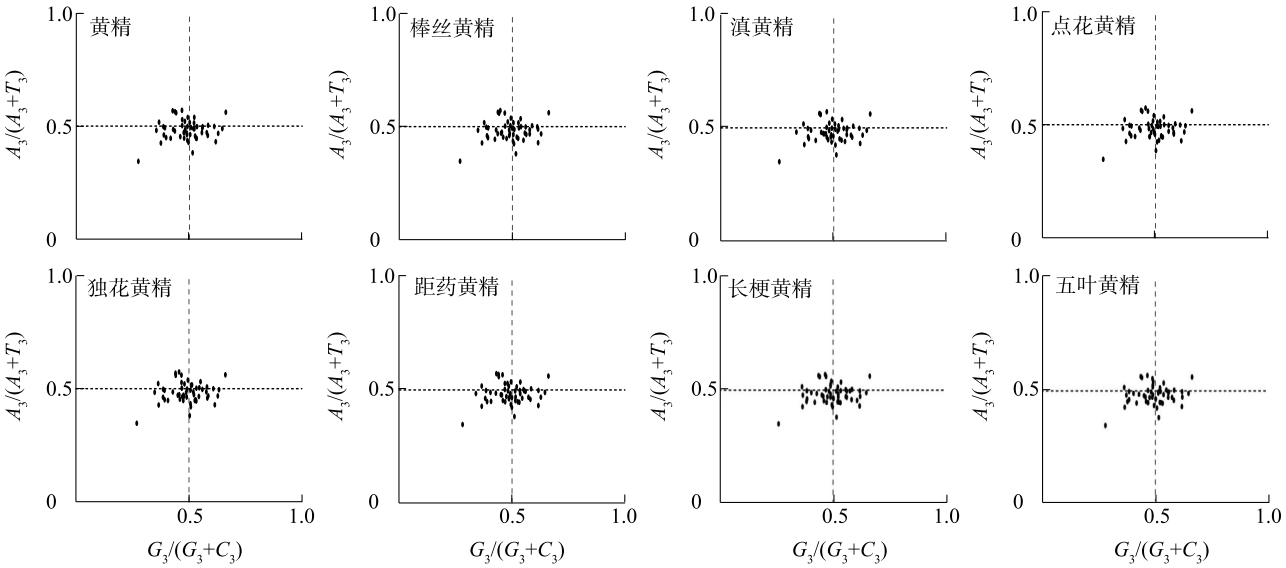
Fig.1 Correlation analysis heatmap of codon usage indicators of *Polygonatum* chloroplast genomes



GC_{12}:第 1、第 2 位上碱基 GC 平均含量;GC_3:密码子第 3 位上 GC 含量。

图 2 黄精属植物叶绿体基因组中性绘图分析结果

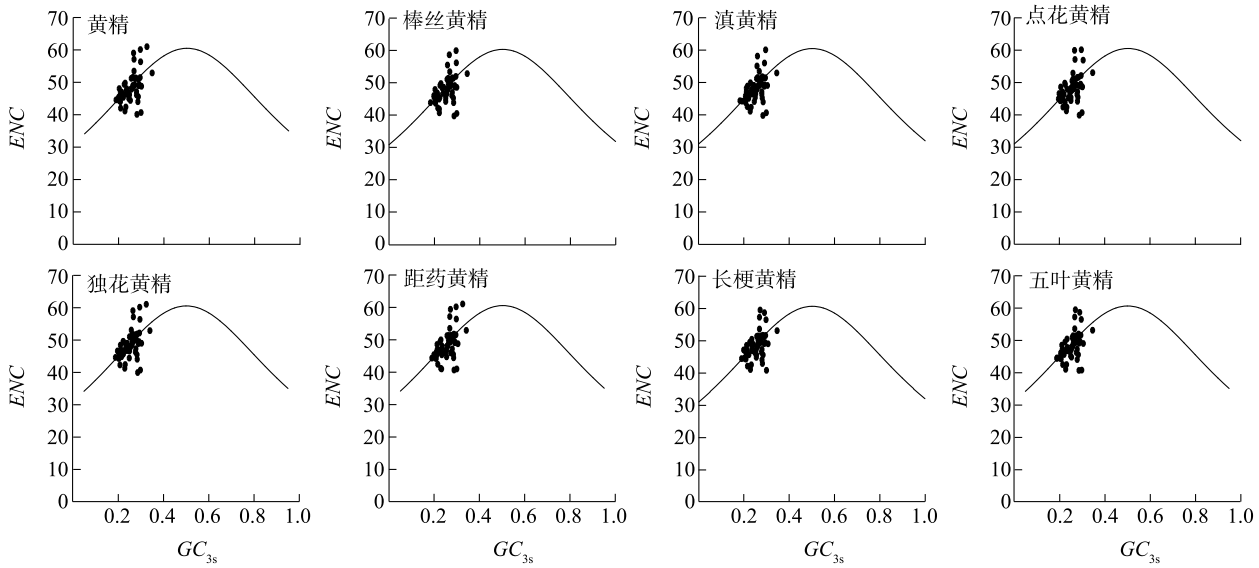
Fig.2 Neutrality plot analysis of chloroplast genomes in eight *Polygonatum* species



A_3:密码子第 3 位上 A 碱基含量;T_3:密码子第 3 位上 T 碱基含量;G_3:密码子第 3 位上 C 碱基含量;C_3:密码子第 3 位上 G 碱基含量。

图 3 黄精属植物叶绿体基因组 PR2-plot 分析结果

Fig.3 PR2-plot analysis of chloroplast genomes in eight *Polygonatum* species



ENC:有效密码子数;GC_{3s}:同义密码子第 3 位上 GC 含量。

图 4　黄精属 8 种植物叶绿体基因组的 *ENC*-plot 分析结果

Fig.4　*ENC*-plot analysis of chloroplast genomes in eight *Polygonatum* species

表 2　黄精、棒丝黄精、滇黄精和点花黄精叶绿体基因组的 *ENC* 比值频数分布

Table 2　Frequency distribution of effective number of codons (*ENC* values) in chloroplast genomes of *Polygonatum sibiricum*, *Polygonatum cathcartii*, *Polygonatum kingianum* and *Polygonatum punctatum*

黄精				棒丝黄精				滇黄精				点花黄精			
ENC 比值范围	*ENC* 比值中间值	基因数（个）	频率	*ENC* 比值范围	*ENC* 比值中间值	基因数（个）	频率	*ENC* 比值范围	*ENC* 比值中间值	基因数（个）	频率	*ENC* 比值范围	*ENC* 比值中间值	基因数（个）	频率
−0.250~−0.151	−0.2	2	0.038 5	−0.250~−0.151	−0.2	2	0.039 2	−0.250~−0.151	−0.2	2	0.039 2	−0.250~−0.151	−0.2	2	0.039 2
−0.150~−0.051	−0.1	7	0.134 6	−0.150~−0.051	−0.1	6	0.117 6	−0.150~−0.051	−0.1	6	0.117 6	−0.150~−0.051	−0.1	6	0.117 6
−0.050~0.050	0	31	0.596 2	−0.050~0.050	0	31	0.607 8	−0.050~0.050	0	31	0.607 8	−0.050~0.050	0	30	0.588 2
0.051~0.150	0.1	10	0.192 3	0.051~0.150	0.1	10	0.196 1	0.051~0.150	0.1	10	0.196 1	0.051~0.150	0.1	11	0.215 7
0.151~0.250	0.2	2	0.038 5	0.151~0.250	0.2	2	0.039 2	0.151~0.250	0.2	2	0.039 2	0.151~0.250	0.2	2	0.039 2

ENC:有效密码子数;*ENC* 比值:有效密码子数的预期值与实际值的差值,除以有效密码子数的预期值。

表 3　独花黄精、距药黄精、长梗黄精和五叶黄精叶绿体基因组的 *ENC* 比值频数分布

Table 3　Frequency distribution of effective number of codons (*ENC* values) in chloroplast genomes of *Polygonatum hookeri*, *Polygonatum franchetii*, *Polygonatum filipes* and *Polygonatum acuminatifolium*

独花黄精				距药黄精				长梗黄精				五叶黄精			
ENC 比值范围	*ENC* 比值中间值	基因数（个）	频率	*ENC* 比值范围	*ENC* 比值中间值	基因数（个）	频率	*ENC* 比值范围	*ENC* 比值中间值	基因数（个）	频率	*ENC* 比值范围	*ENC* 比值中间值	基因数（个）	频率
−0.250~−0.151	−0.2	2	0.037 7	−0.250~−0.151	−0.2	2	0.037 7	−0.250~−0.151	−0.2	1	0.018 9	−0.250~−0.151	−0.2	1	0.019 2
−0.150~−0.051	−0.1	8	0.150 9	−0.150~−0.051	−0.1	7	0.132 1	−0.150~−0.051	−0.1	8	0.150 9	−0.150~−0.051	−0.1	7	0.134 6
−0.050~0.050	0	32	0.603 8	−0.050~0.050	0	33	0.622 6	−0.050~0.050	0	32	0.603 8	−0.050~0.050	0	32	0.615 4
0.051~0.150	0.1	9	0.169 8	0.051~0.150	0.1	9	0.169 8	0.051~0.150	0.1	10	0.188 7	0.051~0.150	0.1	10	0.192 3
0.151~0.250	0.2	2	0.037 7	0.151~0.250	0.2	2	0.037 7	0.151~0.250	0.2	2	0.037 7	0.151~0.250	0.2	2	0.035 8

ENC:有效密码子数;*ENC* 比值:有效密码子数的预期值与实际值的差值,除以有效密码子数的预期值。

2.9　最优密码子筛选

根据 *RSCU*>1 为标准筛选高频率密码子(图 6),其中点花黄精叶绿体基因具有 31 个高频密码子,其他 7 种黄精属植物叶绿体基因各有 30 个高频率密码子。根据 *ENC* 建立高低表达库,以同时满足 *RSCU*>1 且△*RSCU*≥0.08 筛选出叶绿体基因最优

密码子,点花黄精叶绿体基因有 10 个最优密码子,长梗黄精叶绿体基因有 12 个最优密码子,其他 6 种黄精属植物叶绿体基因均有 11 个最优密码子(图 7)。这 8 种黄精属植物叶绿体基因共有 9 个相同的最佳密码子,分别为 CUU、UAU、CAU、AAU、GAU、UCA、CCA、ACA、GGA。

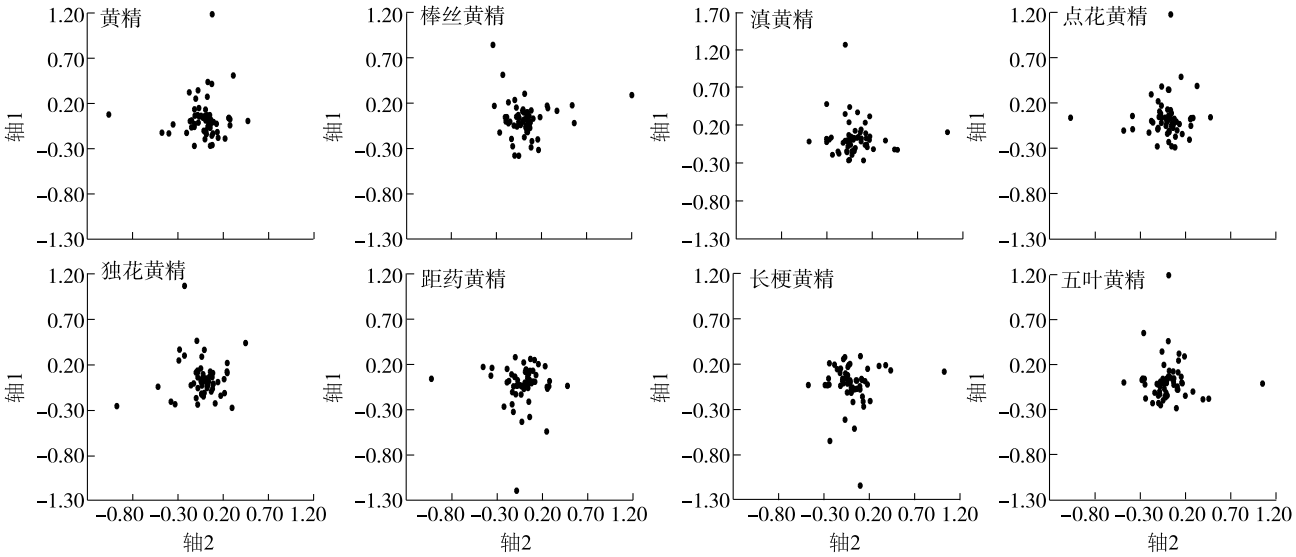
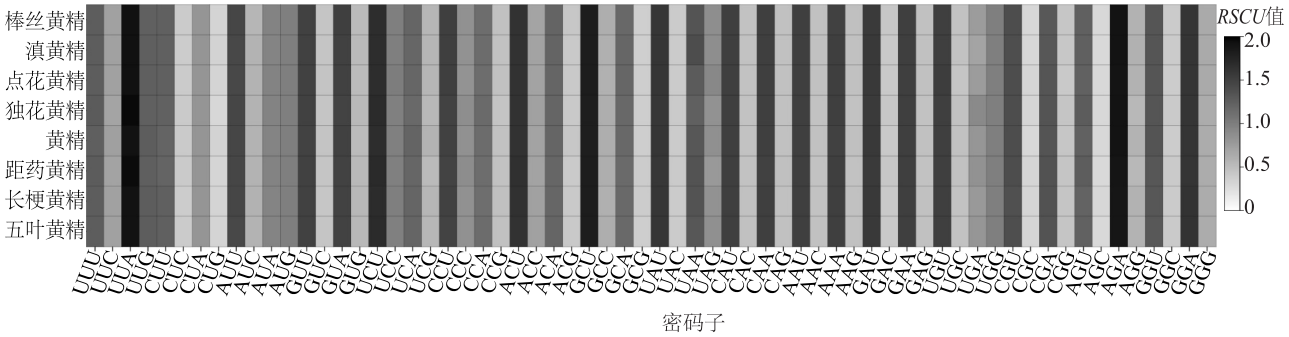


图 5 黄精属 8 种植物叶绿体基因组的对应分析结果

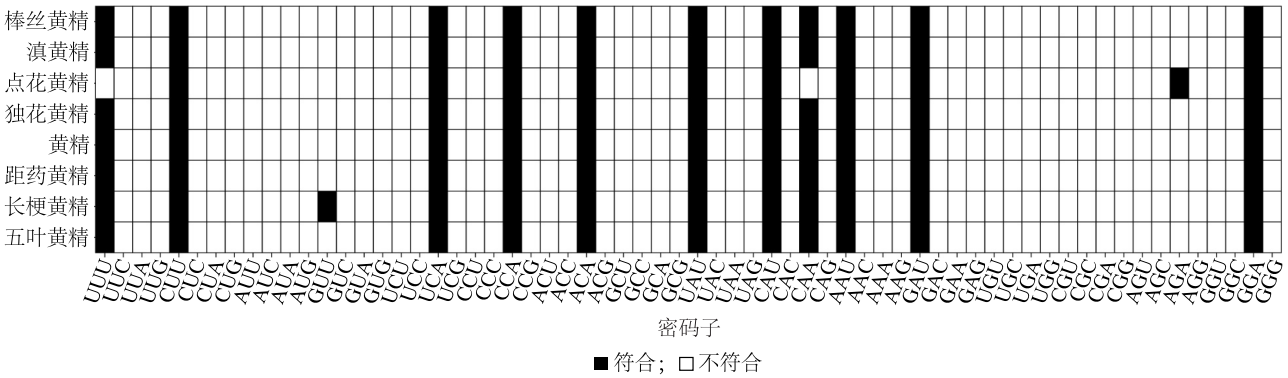
Fig.5 Correspondence analysis (COA) of chloroplast genomes from eight *Polygonatum* species



RSCU:相对同义密码子使用数。

图 6 黄精属植物叶绿体基因组同义密码子分析结果

Fig.6 Synonymous codon usage analysis in chloroplast genomes of eight *Polygonatum* species



黑色背景:相对同义密码子使用数(*RSCU*)>1 和相对同义密码子使用数差异(△*RSCU*)≥0.08;白色背景:*RSCU*>1。

图 7 黄精属植物叶绿体基因组最优密码子分析结果

Fig.7 Optimal codon analysis in chloroplast genomes of eight *Polygonatum* species

2.10 系统发育分析

从图 8 可知,基于邻接(NJ)法构建的系统发育树将 8 种黄精属植物分为 2 大分支,独花黄精和距药黄精以具有 100%的支持率独立为 1 个分支,表明它们亲缘关系密切。其他 6 种独立为另一分支,其中长梗黄精单独为 1 个亚分支得到支持,其与滇黄精、棒丝黄精、点花黄精、五叶黄精和黄精的亲缘关系较远。

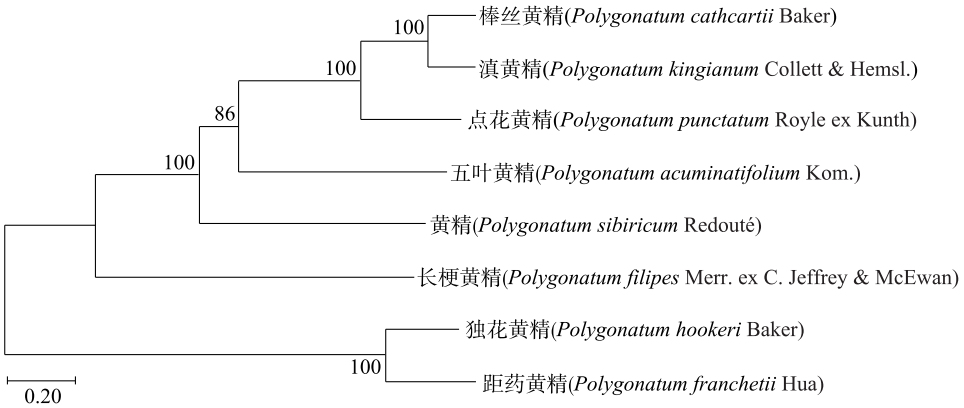


图 8 基于叶绿体基因组编码区序列的黄精属植物的系统发育分析结果

Fig.8 Phylogenetic analysis of eight *Polygonatum* species based on chloroplast genome coding region sequences

3 讨 论

数千年来,黄精的药用和食用价值得到中国人民广泛认可[22-26]。黄精作为一种集药用、食用、观赏和保健于一体的植物,具有显著的经济价值和社会价值。本研究对黄精属植物的叶绿体基因组进行了详细注释,发现 8 种黄精属植物叶绿体基因组均呈现典型的四分体环状结构,4 个区域(SSC 区、LSC 区、IRa 区、IRb 区)虽存在部分扩张与收缩情况,但整体仍然较为保守。基于黄精叶绿体基因组对其他 7 种黄精属植物叶绿体基因组进行变异分析,结果显示 LSC 区具有丰富的序列多样性,且 *ycf1*、*psbN*、*rpl16*、*ycf2* 等基因编码区有明显变异。SSR 是植物叶绿体基因组不可或缺的一部分,其在基因表达、转录调控、染色体结构和生理代谢的研究中有着重要作用[27];8 种黄精属植物叶绿体基因组的单核苷酸重复次数最多,其中以 A 和 T 组成的单核苷酸重复序列占比最多,即黄精属植物叶绿体基因组的碱基组成中 A 和 T 被频繁使用。

生物进化过程中,密码子的使用偏好是基因突变、自然选择和遗传漂移等多种因素综合作用的产物,它在研究物种的进化历程以及药用植物的分子育种中占据举足轻重的地位[5]。密码子使用偏好性显著受到其 GC 含量的影响,特别是由于密码子第 3 碱基位的 GC_3 在选择过程中受到的压力较小,GC_3 常被作为衡量密码子偏好性的关键指标。当自然选择成为密码子偏好性形成的关键因素时,GC_3 与密码子前两个碱基位 GC_{12} 之间通常不会展现出明显的相关性[27]。研究发现,8 种黄精属植物叶绿体基因的密码子 GC 含量在不同碱基位置呈 $GC_1>GC_2>GC_3$ 的梯度分布,且密码子末位多为碱基 A 或碱基 T。此外,*ENC* 比值均大于 35,*CAI* 均远小于 1,*CBI* 均小于 0,*Fop* 均远小于 1。总体来看,这 8 种黄精属植物叶绿体基因组显示出较低的密码子使用偏好性。相关性分析结果显示,GC_3、GC_{3s} 与 *ENC* 比值的相关性为极显著($P<0.01$),说明黄精属植物叶绿体基因的有效密码子使用及密码子偏好性主要受 GC_3 变化的影响。

综合中性绘图分析、RP2-plot 分析、*ENC*-plot 分析和对应分析的数据可以得出:8 种黄精属植物叶绿体基因的密码子使用偏好性是由碱基突变和自然选择等多种因素共同塑造的,其中,中性绘图分析结果说明自然选择压力是主要因素,而 *ENC*-plot 分析结果表明突变压力影响较大,RP2-plot 分析结果显示长梗黄精叶绿体基因更多受到突变压力的影响。

基于叶绿体基因组的 CDS 序列构建的系统发育树显示,独花黄精和距药黄精聚为一分支,独立于其他 6 种黄精属植物,而在另一分支内,长梗黄精又以 100%的支持率区别于其他 5 种黄精属植物,其中棒丝黄精与滇黄精的亲缘关系最近。系统发育树分析

结果可为今后扩大黄精药源的探索提供理论依据。

4 结 论

本研究对8种黄精属植物叶绿体基因组特征、密码子偏好性和系统发育进行分析,结果显示基因突变和自然选择共同影响8种黄精属植物叶绿体基因密码子的使用偏好;8种黄精属植物叶绿体基因共有9个相同的最优密码子,最优密码子倾向于碱基A或碱基U结尾,具体为CUU、UAU、CAU、AAU、GAU、UCA、CCA、ACA和GGA。因此,筛选出的最优密码子将有助于后续黄精属植物分子育种研究,选取最优密码子用于基因表达载体的构建,提高基因的表达效率。本研究结果将为物种进化研究和药材种质鉴定提供重要的理论基础,并通过科学的方法推动黄精种植向规模化、产业化方向迈进,为乡村经济发展贡献力量。

参考文献:

[1] 陶弘景. 名医别录[M]. 尚志钧,辑校. 北京:中国中医药出版社,2013.

[2] 葛 洪. 抱朴子内篇[M]. 梅全喜,郝近大,冉懋雄,等,编译. 北京:中国中医药出版社,1997.

[3] 中国药典委员会. 中华人民共和国药典[S]. 北京:中国医学出版社,2015.

[4] MA C D, CHANG H, YANG Y C, et al. Herbal textual research on polygonati rhizoma in famous classical formulas[J]. Chinese Journal of Experimental Traditional Medical Formulae, 2022, 28(10):193-206.

[5] GRANTHAM R, GAUTIER C, GOUY M, et al. Codon catalog usage is a genome strategy modulated for gene expressivity[J]. Nucleic Acids Research,1981,9(1):43-74.

[6] 刘蒙蒙,邢咏梅,郭顺星. 基于转录组数据分析药用真菌猪苓密码子使用偏好性[J]. 药学学报,2020,55(5):1050-1055.

[7] 王杰敏,马东来,韩晓伟,等. 24份忍冬属材料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,2023,32(3):12-23.

[8] GUSTAFSSON C, GOVINDARAJAN S, MINSHULL J. Codon bias and heterologous protein expression[J]. Trends in Biotechnology,2004,22(7):346-353.

[9] THODE V A, LOHMANN L G, SANMARTÍN I. Evaluating character partitioning and molecular models in plastid phylogenomics at low taxonomic levels:a case study using *Amphilophium* (Bignonieae, Bignoniaceae)[J]. Journal of Systematics and Evolution, 2020,58(6):1071-1089.

[10] 覃雨倩. 天门冬科黄精属系统发育基因组学研究[D]. 吉首:吉首大学,2023.

[11] 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.

[12] BEIER S, THIEL T, MÜNCH T, et al. MISA-web:a web server for microsatellite prediction[J]. Bioinformatics, 2017, 33(16):2583-2585.

[13] AMIRYOUSEFI A, HYVÖNEN J, POCZAI P. IRscope:an online program to visualize the junction sites of chloroplast genomes[J]. Bioinformatics,2018,34(17):3030-3031.

[14] FRAZER K A, PACHTER L, POLIAKOV A, et al. VISTA:computational tools for comparative genomics[J]. Nucleic Acids Research,2004,32:273-279.

[15] SUEOKA N. Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G+C content of third codon position[J]. Gene,1999,238(1):53-58.

[16] SUEOKA N. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses[J]. Journal of Molecular Evolution,2001,53(4/5):469-476.

[17] NOVEMBRE J A. Accounting for background nucleotide composition when measuring codon usage bias[J]. Molecular Biology and Evolution,2002,19(8):1390-1394.

[18] ROMERO H, ZAVALA A, MUSTO H, et al. The influence of translational selection on codon usage in fishes from the family Cyprinidae[J]. Gene,2003,317:141-147.

[19] XIANG H, ZHANG R Z, BUTLER R B, et al. Comparative analysis of codon usage bias patterns in microsporidian genomes[J]. PLoS One,2015,10(6):0129223.

[20] KUMAR S, STECHER G, TAMURA K. MEGA7:molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. Molecular Biology and Evolution,2016,33(7):1870-1874.

[21] WANG W W, LANFEAR R. Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants[J]. Genome Biology and Evolution,2019,11(12):3372-3381.

[22] 韩 凤,黄 靖,章文伟,等. 多花黄精白绢病病原菌鉴定及室内药剂筛选[J]. 江苏农业学报,2024,40(12):2244-2253.

[23] 何艳艳,龙晓明,杨虎权,等. 思茅松林下黄精田节肢动物群落组成结构及多样性分析[J]. 南方农业学报,2024,55(1):117-127.

[24] 马菁华,刘 芳,任启飞,等. 多花黄精致病真菌筛选及致病力鉴定[J]. 江苏农业学报,2023,39(7):1472-1482.

[25] 钱丽华,严建立,吴晓疆,等. 基于滇黄精转录组序列的SSR标记开发及其在黄精属资源分析中的应用[J]. 江苏农业学报,2023,39(5):1120-1131.

[26] 黄 博,邓利娟,杨双琳,等. 云南3种黄精根系丛枝菌根真菌群落结构特征差异研究[J]. 南方农业学报,2023,54(6):1656-1666.

[27] SUEOKA N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons[J]. Journal of Molecular Evolution,1996,42:323.

(责任编辑:黄克玲)