

24 份忍冬属材料叶绿体基因组密码子使用偏好性分析

王杰敏^{1,2,3}, 马东来^{1,2,3}, 韩晓伟^{1,2,3}, 严玉平^{1,2,3}, 郑玉光^{1,2,3,①}, 孙会改^{1,2,3,①}

(1. 河北中医学院, 河北 石家庄 050200; 2. 河北省中药炮制技术创新中心, 河北 石家庄 050200;
3. 河北省中药资源利用与质量评价国际联合研究中心, 河北 石家庄 050200)

摘要: 使用 CodonW1.4.2 和 SPSS 22.0 等软件对 24 份忍冬属 (*Lonicera* Linn.) 材料叶绿体基因组进行密码子偏好性分析, 并基于叶绿体基因组序列构建了系统发育树。结果显示: 24 份忍冬属材料叶绿体基因组中 1 314 个蛋白质编码基因的 GC 含量为 38.7%, 密码子第 3 碱基位置的 GC 含量 (GC_3) 最低, 仅为 31.4%, 表明其密码子偏好以 A 或 U 结尾; 根据基因功能将 1 314 个基因分为 5 类, 每类显示出特定的密码子特征; 24 份忍冬属材料叶绿体基因组的有效密码子数 (ENC) 为 50.2, 表明密码子使用偏好性较低。密码子适应指数 (CAI)、密码子偏好指数 (CBI) 和最优密码子使用频率 (FOP) 与 GC_3 值、总 GC 含量 (GC_{all}) 和同义密码子第 3 碱基位置的 GC 含量 (GC_{3s}) 呈显著或极显著相关, 说明密码子第 3 碱基位置的碱基组成对基因表达有很大影响。中性绘图、PR2-plot 和 ENC-plot 以及对应性分析结果表明影响密码子使用模式的因子并不单一, 自然选择起主要作用, 突变压力起次要作用。基于相对同义密码子使用度 (RSCU) 和高、低表达组的 RSCU 差值 ($\Delta RSCU$), 共鉴定出 29 个高频密码子和 18 个最优密码子, 且大部分以 A 或 U 结尾。系统发育分析结果显示: 24 份忍冬属材料聚为一支, 各物种间的进化相似且亲缘关系较近。综合分析结果表明: 忍冬属植物叶绿体基因组密码子使用偏好性低且富含 A 或 U, 受到自然选择和突变压力的双重影响, 其中自然选择占主导地位。

关键词: 忍冬属; 叶绿体基因组; 密码子使用偏好性; 自然选择; 突变压力

中图分类号: Q943.2; Q949.781.1 文献标志码: A 文章编号: 1674-7895(2023)03-0012-12
DOI: 10.3969/j.issn.1674-7895.2023.03.02

Analysis on codon usage bias of chloroplast genomes of 24 *Lonicera* materials WANG Jiemin^{1,2,3}, MA Donglai^{1,2,3}, HAN Xiaowei^{1,2,3}, YAN Yuping^{1,2,3}, ZHENG Yuguang^{1,2,3,①}, SUN Huigai^{1,2,3,①} (1. Hebei University of Chinese Medicine, Shijiazhuang 050200, China; 2. Traditional Chinese Medicine Processing Technology Innovation Center of Hebei Province, Shijiazhuang 050200, China; 3. International Joint Research Center on Resource Utilization and Quality Evaluation of Traditional Chinese Medicine of Hebei Province, Shijiazhuang 050200, China), *J. Plant Resour. & Environ.*, 2023, 32(3): 12-23

Abstract: The codon usage bias of chloroplast genomes of 24 *Lonicera* Linn. materials were analyzed by using softwares including CodonW1.4.2 and SPSS 22.0, and a phylogenetic tree was constructed based on the chloroplast genome sequences. The results show that the GC content of 1 314 protein-coding genes in chloroplast genomes of 24 *Lonicera* materials is 38.7%, and the GC content at the third base position of codon (GC_3) is the lowest, which is only 31.4%, indicating that their codons prefer to end with A or U; 1 314 genes are divided into 5 categories based on gene function, and each category shows specific codon characteristics; the effective number of codon (ENC) of chloroplast genomes of 24 *Lonicera* materials is 50.2, indicating that the codon usage bias is relatively low. The codon adaptation index (CAI), codon

收稿日期: 2022-10-03

基金项目: 河北省中医药管理局科研计划项目 (2023355); 河北省二期现代农业产业技术体系创新团队项目 (HBCT2018060205); 河北中医学院科研能力提升项目 (KTY2019077; KTZ2019006)

作者简介: 王杰敏 (2000—), 男, 河北张家口人, 硕士研究生, 主要从事中药资源学研究。

①通信作者 E-mail: zyg314@163.com; sunhuigai66@163.com

引用格式: 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.

bias index (CBI), and frequency of optimal codon (FOP) show significant or extremely significant correlations with GC_3 value, total GC content (GC_{all}), and GC content at the third base position of synonymous codon (GC_{3s}), indicating that the base composition of the third base position of codon has a great influence on gene expression. The neutral mapping, PR2-plot, ENC-plot, and correspondence analysis results show that the factors affecting codon usage pattern are not single, and natural selection plays a dominant role, while mutation pressure plays a secondary role. Based on relative synonymous codon usage (RSCU) and RSCU difference between high and low expression groups ($\Delta RSCU$), 29 high frequency codons and 18 optimal codons are identified in total, and most of them end with A or U. The phylogenetic analysis result shows that 24 *Lonicera* materials can be clustered into one branch, and the evolution among each species is similar and their genetic relationships are relatively close. The comprehensive analysis result shows that the codon usage bias of chloroplast genomes of 24 *Lonicera* materials is low and they are rich in A or U, and they are under dual influences of natural selection and mutation pressure, in which natural selection plays a major role.

Key words: *Lonicera* Linn.; chloroplast genome; codon usage bias; natural selection; mutation pressure

叶绿体是植物重要的细胞器,在光合作用、生物合成和碳封存中有十分重要的作用^[1]。自烟草(*Nicotiana tabacum* Linn.)叶绿体基因组被公布后,其结构和功能受到广泛关注^[2]。与线粒体和核基因组相比,叶绿体基因组序列因其较低的核苷酸取代率、适中的序列长度和无旁系同源干扰的优势被广泛用于植物系统发育研究^[3]。随着新一代测序技术的进步,越来越多的植物叶绿体基因组被测序和组装。目前,叶绿体基因组已广泛应用于阐明系统发育关系^[4]、澄清种间和种内变异^[5]以及物种鉴定^[6]等方面。因此,叶绿体基因组成为不同层次系统发育研究的理想资源^[7]。

遗传密码是核酸与蛋白质间的联系纽带,对生物体遗传信息传递有重要影响。密码子使用偏好性(CUB)是指蛋白质编码过程中同义密码子使用频率的差异,这种现象在生物体中相当普遍^[8-9]。然而,不同的基因组具有其特有的同义密码子使用模式。分析叶绿体基因组中的密码子使用偏好性,有助于了解密码子使用偏好性的潜在分子机制以及相关物种的进化和环境适应,为分子育种提供依据^[10]。目前已有多个物种叶绿体基因组密码子偏好性的研究报告,如玉米(*Zea mays* Linn.)^[11]和菠萝[*Ananas comosus* (Linn.) Merr.]^[12]等单子叶植物中密码子常以G或C结尾,而木兰科(Magnoliaceae)^[13]和芍药科(Paeoniaceae)^[14]等双子叶植物中密码子常以A或U结尾。

忍冬属(*Lonicera* Linn.)是忍冬科(Caprifoliaceae)中最大的属,在北美洲、欧洲、亚洲和非洲北部地区均有分布,忍冬属植物具有重要的药用和观赏价值^[15]。

忍冬(*L. japonica* Thunb.)汤剂中的miRNA可有效抑制新冠病毒复制,加速新冠患者转阴^[16];被称作“第三代果品”的蓝靛果(*L. caerulea* Linn.)具有较高的营养和食用价值^[17];忍冬属植物还在生态系统物质循环、植被更新和生物多样性维护等方面发挥着重要作用^[18]。加强忍冬属植物叶绿体基因组的基础研究对其高效利用具有重要意义。虽然忍冬属植物种类较多,但目前NCBI(<https://www.ncbi.nlm.nih.gov/>)数据库中仅有24份忍冬属材料的叶绿体基因组数据,且这些材料在分类学水平上的进化特征和遗传多样性尚未研究清楚。鉴于此,本研究基于已测序的24份忍冬属材料的叶绿体基因组序列信息,利用生物信息学方法分析其密码子使用偏好性,并评估影响密码子使用偏好性的关键因子,为进一步的遗传进化研究和转基因研究提供参考。

1 材料和方法

1.1 材料

24份忍冬属材料的叶绿体基因组原始编码序列(CDS)来自NCBI数据库。为避免取样偏差,提高分析准确性^[19],利用Galaxy在线程序(<https://usegalaxy.org/>)的Filter sequences by length工具将长度小于300 bp的序列以及重复序列剔除。最终共筛选到1 314条符合条件的编码序列用于后续分析,24份忍冬属材料叶绿体基因组的基本信息见表1。

1.2 方法

1.2.1 密码子相关参数的计算 使用CodonW1.4.2软件对筛选出的1 314条编码序列进行密码子使用

表 1 24 份忍冬属材料叶绿体基因组的基本信息¹⁾
Table 1 Basic information of chloroplast genomes of 24 *Lonicera* Linn. materials¹⁾

材料 Material	l/bp	登录号 Accession number	n_1	n_2
忍冬 <i>L. japonica</i>	155 078	KJ170923.1	130	55
忍冬 <i>L. japonica</i>	155 060	MH028738.1	130	57
金银忍冬 <i>L. maackii</i>	155 318	NC_039636.1	130	57
金银忍冬 <i>L. maackii</i>	155 318	MH028741.1	130	57
大花忍冬 <i>L. macrantha</i>	154 897	MH579750.1	130	55
大花忍冬 <i>L. macrantha</i>	155 515	MW493344.1	129	56
大花忍冬 <i>L. macrantha</i>	155 126	MW795592.1	123	51
红脉忍冬 <i>L. nervosa</i>	154 862	MK176510.1	128	55
红脉忍冬 <i>L. nervosa</i>	154 862	NC_040961.1	128	55
细毡毛忍冬 <i>L. similis</i>	155 207	MZ241297.1	121	51
细毡毛忍冬 <i>L. similis</i>	155 463	MW970104.1	129	56
盘叶忍冬 <i>L. tragophylla</i>	155 545	MG738667.1	128	54
盘叶忍冬 <i>L. tragophylla</i>	155 545	NC_037953.1	128	54
葱皮忍冬 <i>L. ferdinandi</i>	154 513	MK176512.1	128	55
葱皮忍冬 <i>L. ferdinandi</i>	155 182	MH028743.1	128	55
紫花忍冬 <i>L. maximowiczii</i>	155 123	MH028742.1	130	57
紫花忍冬 <i>L. maximowiczii</i>	155 584	MN986996.1	129	55
刚毛忍冬 <i>L. hispida</i>	154 553	MK176511.1	128	54
长白忍冬 <i>L. ruprechtiana</i>	154 611	MW296954.1	131	57
冠果忍冬 <i>L. stephanocarpa</i>	155 056	MG738668.1	128	55
新疆忍冬 <i>L. tatarica</i>	154 677	MW340876.1	131	55
华南忍冬 <i>L. confusa</i>	155 157	MW795591.1	123	51
早花忍冬 <i>L. praeflorens</i>	154 892	MH028740.1	130	56
淡红忍冬 <i>L. acuminata</i>	155 249	MZ241298.1	119	51

¹⁾ l: 基因组长度 Length of genome; n_1 : 原始编码序列数 Number of original coding sequence; n_2 : 选取的编码序列数 Number of selected coding sequence.

情况分析,包括相对同义密码子使用度(RSCU)、有效密码子数(ENC)、密码子适应指数(CAI)、密码子偏好指数(CBI)、最优密码子使用频率(FOP)和 GC_{3s} (同义密码子第3碱基位置的GC含量,不包括甲硫氨酸、色氨酸和3个终止密码子)。RSCU、ENC、CAI、CBI和FOP值可以评估基因组密码子使用偏好性。RSCU值大于1表示对应密码子的使用频率高于预期频率,其中RSCU值大于2表示对应密码子的使用频率极高,RSCU值小于1则表示对应密码子的使用频率低于预期频率;ENC值小于30和大于55的序列分别对应高表达和低表达基因,ENC值越低表示密码子使用偏好性越高;CAI值是指编码区同义密码子与最优密码子使用频率的相符程度,取值在0~1之间,该值越大表示适应性越强,基因表达水平也越高;CBI值为1表示仅有最优密码子被使用,CBI值为负数则表示没有最优密码子被使用;FOP值是最优密

码子数与其同义密码子数的比值,取值在0~1之间,FOP值为1表示仅有最优密码子被使用,FOP值为0则表示没有最优密码子被使用^[20]。使用EMBOSS在线程序的CUSP工具(<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>)计算各基因GC含量和各类基因总GC含量(GC_{all}),密码子第1、第2和第3碱基位置的GC含量(GC_1 、 GC_2 和 GC_3),以及密码子第3碱基位置的T(U,下同)、C、A和G含量(T_3 、 C_3 、 A_3 和 G_3)。使用SPSS 22.0软件和EXCEL 2019软件对叶绿体基因组密码子使用偏好性相关指标间的相关性进行分析。

1.2.2 中性绘图分析 中性绘图分析是定量分析突变压力和自然选择对密码子使用偏好性影响的一种方法。同义密码子的突变通常发生在第3碱基位置,而第1和第2碱基位置的突变是非同义密码子的突变,非同义密码子的突变率较低^[21]。在中性绘图中,以各基因的 GC_1 和 GC_2 的平均值(GC_{12})为纵坐标,以 GC_{3s} 值为横坐标,每个点代表1个独立的基因。若所有点均是对角分布,表示密码子3个位置的碱基没有显著差异,密码子使用偏好性只受突变压力的影响;若所有点不是对角分布,且 GC_{12} 值和 GC_{3s} 值间的相关性很低,即回归曲线的斜率(回归系数)越小,甚至平行于水平轴,则表示自然选择是密码子使用偏好性的主要影响因素^[22]。

1.2.3 PR2-plot分析 PR2-plot分析被广泛用于解释DNA双链中是否存在突变压力和自然选择对核苷酸组成的影响。以 $G_3/(G_3+C_3)$ 比为横坐标,以 $A_3/(A_3+T_3)$ 比为纵坐标, A_3 值等于 T_3 值且 G_3 值等于 C_3 值的位置为中心,图的中心位置表示2条互补DNA链间没有突变压力和自然选择的偏差。若 G_3 和 C_3 值或 A_3 和 T_3 值接近,表示叶绿体基因组密码子使用偏好性只受突变压力的影响;若 G_3 和 C_3 值或 A_3 和 T_3 值差异较大,则表示密码子使用偏好性主要是由自然选择引起的^[23]。

1.2.4 ENC-plot分析 ENC值可在基因组水平上评价密码子使用偏好性。参考Wright^[19]发表的算法计算ENC值,取值范围从20(极端偏好性,每个氨基酸仅使用1个密码子)至61(完全无偏好性,密码子随机使用)。ENC-plot图由散点图和标准曲线构成,散点图以 GC_{3s} 值为横坐标,以ENC值为纵坐标。标准曲线代表密码子使用偏好性的决定因子是突变压力,标准曲线公式为 $ENC = 2 + GC_{3s} + 29 / [GC_{3s}^2 + (1 - GC_{3s})^2]$ 。

基因位于标准曲线上或附近,表示基因组的密码子使用偏好性受突变压力的影响较大,而远离标准曲线,在保证密码子最佳使用的情况下,密码子使用偏好性受自然选择的影响较大。同时为了准确评估观察值(ENC_{obs})与期待值(ENC_{exp})间的差异,采用公式“ENC 比值=($ENC_{exp}-ENC_{obs}$)/ ENC_{exp} ”计算 ENC 比值,根据 ENC 比值分布对观察值和期待值间的差异进行量化分析。

1.2.5 对应性分析 对应性分析通常被用作密码子使用模式的多变量统计分析。基于 RSCU 值在 59 维向量空间中绘制所有基因。第 1 轴代表密码子使用的大部分变化,之后的每个轴解释的变化量逐渐减少^[24]。基因之间的位置,特别是不同功能基因之间的位置可被预期和比较。

1.2.6 最优密码子分析 使用 EXCEL 2019 软件对 ENC 值排序,选择 ENC 值小于 30 和大于 55 的序列分别建立高、低表达组,并从各组取前 10% 的基因,

参考 Sharp 等^[25]发表的算法计算 2 个数据集的 RSCU 值,并用 $\Delta RSCU$ 值(高、低表达组的 RSCU 差值)进行比较。RSCU 值大于 1 的密码子被定义为高频密码子; $\Delta RSCU$ 值大于 0.08 且 RSCU 值大于 1 的密码子被定义为最优密码子^[26]。

1.2.7 系统发育分析 从 NCBI 公共数据库中以 FASTA 格式检索 24 份忍冬属材料叶绿体基因组序列,以同科不同属植物六道木 [*Zabelia tyaihyoni* (Nakai) Hisauti et H. Hara] 为外类群,使用 MEGA 6 软件,基于最大似然法构建系统发育树^[27]。

2 结果和分析

2.1 叶绿体基因组的碱基组成

结果(表 2 和表 3)显示:筛选到的 1 314 个蛋白质编码基因中包含 614 个光合作用相关基因、392 个核糖体蛋白相关基因、92 个自我复制相关基因、112

表 2 24 份忍冬属材料叶绿体基因组蛋白质编码基因的 GC 含量($\bar{X}\pm SD$)¹⁾

Table 2 GC content in protein-coding genes of chloroplast genomes of 24 *Lonicera* Linn. materials ($\bar{X}\pm SD$)¹⁾

基因 Gene	GC 含量/% GC content	基因 Gene	GC 含量/% GC content	基因 Gene	GC 含量/% GC content	基因 Gene	GC 含量/% GC content	基因 Gene	GC 含量/% GC content	基因 Gene	GC 含量/% GC content
PRG		<i>ndhE</i>	33.3±0.3	<i>psaA</i>	43.1±0.1	<i>rpl2</i>	41.8±0.2	<i>rps4</i>	39.4±0.1	<i>ycf2</i>	38.7±0.1
<i>atpA</i>	41.0±0.1	<i>ndhF</i>	32.5±0.1	<i>psaB</i>	41.2±0.2	<i>rpl20</i>	38.1±0.2	<i>rps7</i>	38.5±0.2	<i>ycf3</i>	37.5±0.1
<i>atpB</i>	42.9±0.1	<i>ndhG</i>	35.7±0.3	<i>psbA</i>	42.6±0.1	<i>rpl22</i>	35.8±0.2	<i>rps8</i>	36.2±0.3	<i>ycf4</i>	39.1±0.2
<i>atpE</i>	40.6±0.1	<i>ndhH</i>	38.7±0.1	<i>psbB</i>	44.6±0.2	<i>rps11</i>	46.0±0.3	SRRG		其他基因 Other genes	
<i>atpF</i>	35.9±0.2	<i>ndhI</i>	35.5±0.1	<i>psbC</i>	43.9±0.1	<i>rps12</i>	40.4±0.0	<i>rpoC1</i>	38.5±0.1	<i>accD</i>	41.8±1.2
<i>atpI</i>	38.1±0.1	<i>ndhJ</i>	40.0±0.3	<i>psbD</i>	43.2±0.1	<i>rps14</i>	42.7±0.3	<i>rpoC2</i>	38.2±0.1	<i>ccsA</i>	32.9±0.2
<i>ndhA</i>	34.1±0.1	<i>ndhK</i>	39.1±0.1	<i>rbcL</i>	45.0±0.2	<i>rps16</i>	35.6±0.3	<i>rpoA</i>	36.2±0.8	<i>cemA</i>	34.5±0.2
<i>ndhB</i>	37.7±0.1	<i>petA</i>	40.6±0.2	RPRG		<i>rps18</i>	35.8±0.4	<i>rpoB</i>	39.7±0.1	<i>clpP</i>	42.9±0.2
<i>ndhC</i>	36.5±0.2	<i>petB</i>	38.0±1.3	<i>rpl14</i>	41.4±0.2	<i>rps2</i>	38.7±0.2	<i>ycf</i> 基因 <i>ycf</i> genes		<i>matK</i>	35.1±0.2
<i>ndhD</i>	36.2±0.0	<i>petD</i>	38.1±0.3	<i>rpl16</i>	36.0±2.5	<i>rps3</i>	33.4±0.2	<i>ycf1</i>	32.2±0.3		

¹⁾ PRG: 光合作用相关基因 Photosynthesis related genes; RPRG: 核糖体蛋白相关基因 Ribosomal protein related genes; SRRG: 自我复制相关基因 Self-replicating related genes.

表 3 24 份忍冬属材料叶绿体基因组不同功能基因的碱基组成和表达特征¹⁾

Table 3 Base composition and expression characteristics of different functional genes of chloroplast genomes of 24 *Lonicera* Linn. materials¹⁾

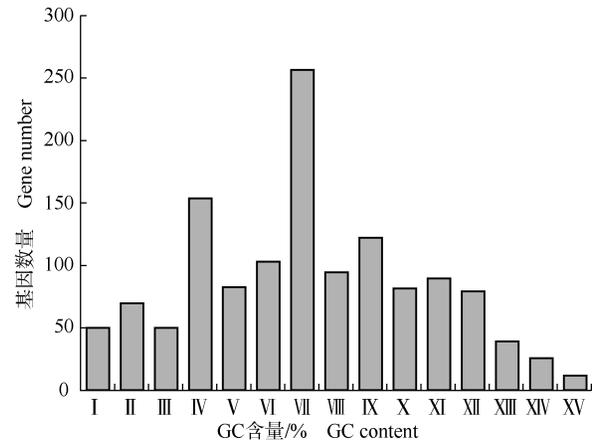
功能基因 Functional gene	GC ₁ /%	GC ₂ /%	GC ₃ /%	GC _{all} /%	GC _{3s} /%	ENC	CAI	CBI	FOP
光合作用相关基因 Photosynthesis related genes	47.4	39.1	31.2	39.2	28.1	49.8	0.18	-0.08	0.36
核糖体蛋白相关基因 Ribosomal protein related genes	44.3	41.3	30.8	38.8	28.4	49.4	0.16	-0.07	0.37
自我复制相关基因 Self-replicating related genes	46.5	38.0	30.2	38.2	28.2	50.8	0.15	-0.15	0.33
<i>ycf</i> 基因 <i>ycf</i> genes	41.5	36.6	34.4	37.5	31.9	52.2	0.16	-0.12	0.35
其他基因 Other genes	43.0	34.2	33.0	36.7	29.8	52.6	0.16	-0.15	0.32
总体 Overall	45.6	39.1	31.4	38.7	31.4	50.2	0.17	-0.09	0.36

¹⁾ GC₁: 密码子第 1 碱基位置的 GC 含量 GC content at the first base position of codon; GC₂: 密码子第 2 碱基位置的 GC 含量 GC content at the second base position of codon; GC₃: 密码子第 3 碱基位置的 GC 含量 GC content at the third base position of codon; GC_{all}: 总 GC 含量 Total GC content; GC_{3s}: 同义密码子第 3 碱基位置的 GC 含量 GC content at the third base position of synonymous codon; ENC: 有效密码子数 Effective number of codon; CAI: 密码子适应指数 Codon adaptation index; CBI: 密码子偏好指数 Codon bias index; FOP: 最优密码子使用频率 Frequency of optimal codon.

个 *ycf* 基因和 104 个其他基因。1 314 个基因的 GC 含量为 38.7%，其中，*ycf1* 基因的 GC 含量最低 (32.2%)，*rps11* 基因的 GC 含量最高 (46.0%)。进一步统计 (表 3 和图 1) 发现，GC 含量 (C_{GC}) 在 $38\% < C_{GC} \leq 39\%$ 范围内的基因数量最多 (257)，GC 含量在 $35\% < C_{GC} \leq 36\%$ 和 $40\% < C_{GC} \leq 41\%$ 范围内的基因数量也较多，分别有 154 和 122 个。按功能分类，每类基因的总 GC 含量 (GC_{all}) 不同，其中光合作用相关基因的 GC_{all} 值最高 (39.2%)。

结果 (表 3) 显示：24 份忍冬属材料叶绿体基因组密码子 3 个位置的碱基组成特征不同。总体上看，密码子第 1 碱基位置的 GC 含量 (GC_1) 最高 (45.6%)，密码子第 2 碱基位置的 GC 含量 (GC_2) 和密码子第 3 碱基位置的 GC 含量 (GC_3) 分别为 39.1% 和 31.4%。表明 24 份忍冬属材料的叶绿体基因组在密码子末端优先使用 A 或 U。各类功能基因的碱基组成也不同，光合作用相关基因的 GC_1 值最高 (47.4%)，核糖体蛋白相关基因的 GC_2 值最高 (41.3%)，*ycf* 基因的 GC_3 值最高 (34.4%)。24 份忍冬属材料叶绿体基因组的有效密码子数 (ENC) 为 50.2，其中，其他基因的 ENC 值最高 (52.6)，*ycf* 基因的 ENC 值次之 (52.2)，核糖体蛋白相关基因的 ENC 值最低 (49.4)。密码子适应指数 (CAI)、密码子偏好指数 (CBI) 和最优密码子使用频率 (FOP) 均较小，说明基因的表达水平较低。总体上看，与其他基因相比，光合作用相关基因的 FOP 值 (0.36) 较高、CAI 值 (0.18) 最高，表明这类基因的表达水平更高。

相关性分析结果 (表 4) 显示： GC_1 、 GC_2 和 GC_3 值间呈极显著或显著相关，表明密码子第 1、第 2 和第 3 碱基位置的碱基组成相似。ENC 值与 GC_1 、 GC_2 和 GC_3 值呈极显著相关，表明密码子不同位置的碱基组成对密码子使用偏好性有显著影响。CAI、CBI 和 FOP 值与 GC_3 值、 GC_{all} 值和同义密码子第 3 碱基位置的 GC 含量 (GC_{3s}) 呈显著或极显著相关，说明叶绿体基因组的 GC 含量，尤其是密码子第 3 碱基位置的碱基组成对基因表达有很大影响。



I: $C_{GC} \leq 33\%$; II: $33\% < C_{GC} \leq 34\%$; III: $34\% < C_{GC} \leq 35\%$; IV: $35\% < C_{GC} \leq 36\%$; V: $36\% < C_{GC} \leq 37\%$; VI: $37\% < C_{GC} \leq 38\%$; VII: $38\% < C_{GC} \leq 39\%$; VIII: $39\% < C_{GC} \leq 40\%$; IX: $40\% < C_{GC} \leq 41\%$; X: $41\% < C_{GC} \leq 42\%$; XI: $42\% < C_{GC} \leq 43\%$; XII: $43\% < C_{GC} \leq 44\%$; XIII: $44\% < C_{GC} \leq 45\%$; XIV: $45\% < C_{GC} \leq 46\%$; XV: $46\% < C_{GC} \leq 47\%$. C_{GC} : GC 含量 GC content.

图 1 24 份忍冬属材料叶绿体基因组不同 GC 含量水平的基因分布
Fig. 1 Gene distribution of chloroplast genomes of 24 *Lonicera* Linn. materials at different GC content levels

表 4 24 份忍冬属材料叶绿体基因组密码子使用偏好性相关指标的相关性分析¹⁾

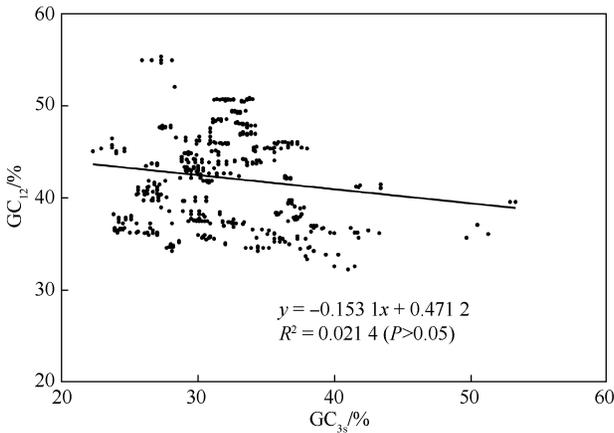
Table 4 Correlation analysis of related indexes of codon usage bias of chloroplast genomes of 24 *Lonicera* Linn. materials¹⁾

指标 Index	相关系数 Correlation coefficient								
	GC_1	GC_2	GC_3	GC_{all}	GC_{3s}	ENC	CAI	CBI	FOP
GC_1	1.000								
GC_2	0.316 **	1.000							
GC_3	-0.266 **	-0.063 *	1.000						
GC_{all}	0.702 **	0.724 **	0.266 **	1.000					
GC_{3s}	-0.321 **	-0.076 **	0.976 **	0.209 **	1.000				
ENC	-0.152 **	-0.304 **	0.410 **	-0.034	0.394 **	1.000			
CAI	0.499 **	0.135 **	-0.072 **	0.364 **	-0.157 **	-0.018	1.000		
CBI	-0.044	0.014	-0.174 **	-0.100 **	-0.158 **	-0.205 **	-0.027	1.000	
FOP	0.331 **	0.346 **	0.070 *	0.425 **	0.058 *	-0.097 **	0.614 **	-0.374 **	1.000

¹⁾ GC_1 : 密码子第 1 碱基位置的 GC 含量 GC content at the first base position of codon; GC_2 : 密码子第 2 碱基位置的 GC 含量 GC content at the second base position of codon; GC_3 : 密码子第 3 碱基位置的 GC 含量 GC content at the third base position of codon; GC_{all} : 总 GC 含量 Total GC content; GC_{3s} : 同义密码子第 3 碱基位置的 GC 含量 GC content at the third base position of synonymous codon; ENC: 有效密码子数 Effective number of codon; CAI: 密码子适应指数 Codon adaptation index; CBI: 密码子偏好指数 Codon bias index; FOP: 最优密码子使用频率 Frequency of optimal codon. *: $P < 0.05$; **: $P < 0.01$.

2.2 中性绘图分析

结果(图 2)显示:24 份忍冬属材料的大部分叶绿体基因分布相对集中但并不密集,小部分点沿对角线分布。密码子第 1 和第 2 碱基位置 GC 含量的平均值(GC_{12})与同义密码子第 3 碱基位置的 GC 含量(GC_{3s})的平均值分别为 42.3% 和 31.4%。 GC_{12} 值与 GC_{3s} 值的回归系数较小(-0.153 1),且二者间未达到显著水平,表明密码子第 1 和第 2 碱基位置的 GC 含量与同义密码子第 3 碱基位置的 GC 含量差异较大。说明 24 份忍冬属材料叶绿体基因组的密码子使用偏好性受自然选择和突变压力的双重影响,其中自然选择的影响更大。



GC_{3s} : 同义密码子第 3 碱基位置的 GC 含量 GC content at the third base position of synonymous codon; GC_{12} : 密码子第 1 和第 2 碱基位置 GC 含量的平均值 Average of GC content at the first and second base positions of codon.

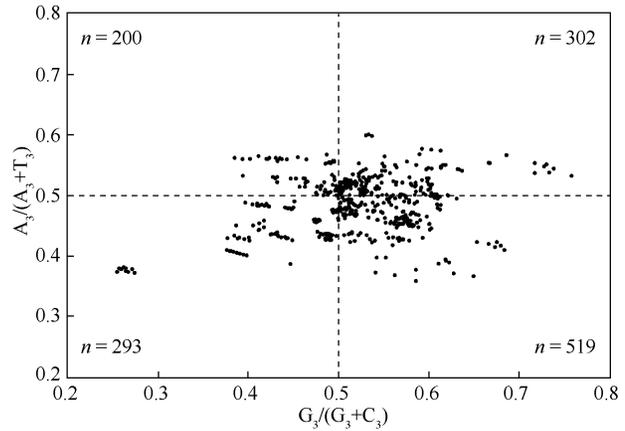
图 2 24 份忍冬属材料叶绿体基因组中性绘图分析
Fig. 2 Neutral mapping analysis on chloroplast genomes of 24 *Lonicera* Linn. materials

2.3 PR2-plot 分析

结果(图 3)显示:24 份忍冬属材料叶绿体基因不均等的分布在 4 个区域内,左上角区域有 200 个基因,左下角区域有 293 个基因,右上角区域有 302 个基因,右下角区域有 519 个基因。右下角区域基因分布较多,即密码子第 3 碱基位置的碱基使用频率为 G 大于 C, T 大于 A,说明 24 份忍冬属材料叶绿体基因组的密码子使用偏好性受自然选择和突变压力的双重影响,其中自然选择的影响更大。

2.4 ENC-plot 分析

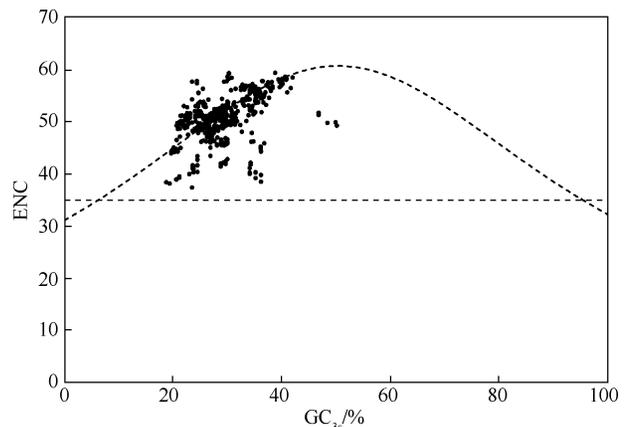
结果(图 4 和表 5)显示:24 份忍冬属材料的大部分叶绿体基因与标准曲线之间均具有一定的距离



T_3, C_3, A_3, G_3 : 分别为密码子第 3 碱基位置的 T、C、A 和 G 含量 The content of T, C, A, and G at the third base position of codon, respectively. n : 基因数量 Gene number.

图 3 24 份忍冬属材料叶绿体基因组 PR2-plot 分析
Fig. 3 PR2-plot analysis on chloroplast genomes of 24 *Lonicera* Linn. materials

偏差,且多数分布于标准曲线下方,说明基因的密码子使用偏好性主要受自然选择的影响。分析有效密码子数(ENC)比值的频数分布发现,621 个基因(占比 47.3%)分布在 $[-0.05, 0.05]$, 剩余 693 个基因(占比 52.7%)均分布在这个范围外侧,即观察值(ENC_{obs})与期待值(ENC_{exp})差异较小,也说明自然选择对 24 份忍冬属材料叶绿体基因组的密码子使用偏好性的贡献更大,突变压力的贡献较小。



GC_{3s} : 同义密码子第 3 碱基位置的 GC 含量 GC content at the third base position of synonymous codon; ENC: 有效密码子数 Effective number of codon.

图 4 24 份忍冬属材料叶绿体基因组 ENC-plot 分析
Fig. 4 ENC-plot analysis on chloroplast genomes of 24 *Lonicera* Linn. materials

表 5 24 份忍冬属材料叶绿体基因组有效密码子数 (ENC) 比值分布
Table 5 Distribution of ratios of effective number of codon (ENC) in chloroplast genomes of 24 *Lonicera* Linn. materials

ENC 比值 ENC ratio		基因数量 Gene number	占比/% Proportion
范围 Range	中值 Mid-value		
[-0.15, -0.05)	-0.08	54	4.1
[-0.05, 0.05)	0.00	621	47.3
[0.05, 0.15)	0.10	494	37.6
[0.15, 0.25)	0.20	119	9.1
[0.25, 0.35)	0.29	26	2.0
合计 Total		1 314	100.0

2.5 对应性分析

基于相对同义密码子使用度 (RSCU) 进行对应性分析。结果(图 5 和图 6)显示:第 1、第 2、第 3 和第 4 轴分别占 24 份忍冬属材料叶绿体基因总变异的 10.6%、8.6%、7.4% 和 7.0%,之后各轴的贡献率逐渐减小,说明更多的因子参与了叶绿体基因组密码子使用偏好性的调控。各功能基因在前 2 轴上的分布没有明显的偏好性,说明叶绿体基因间密码子使用模式较为接近。

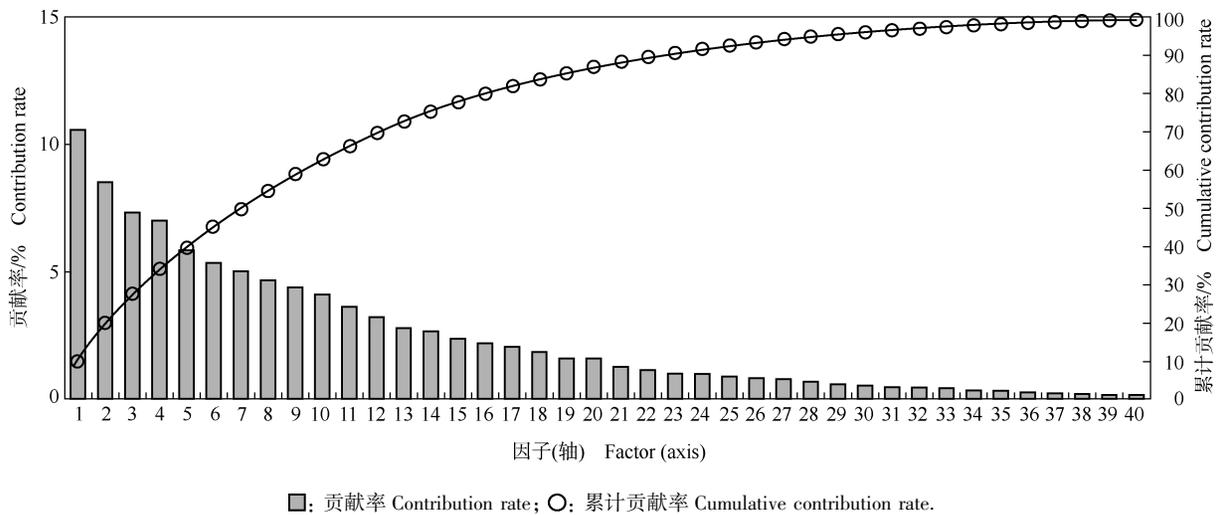
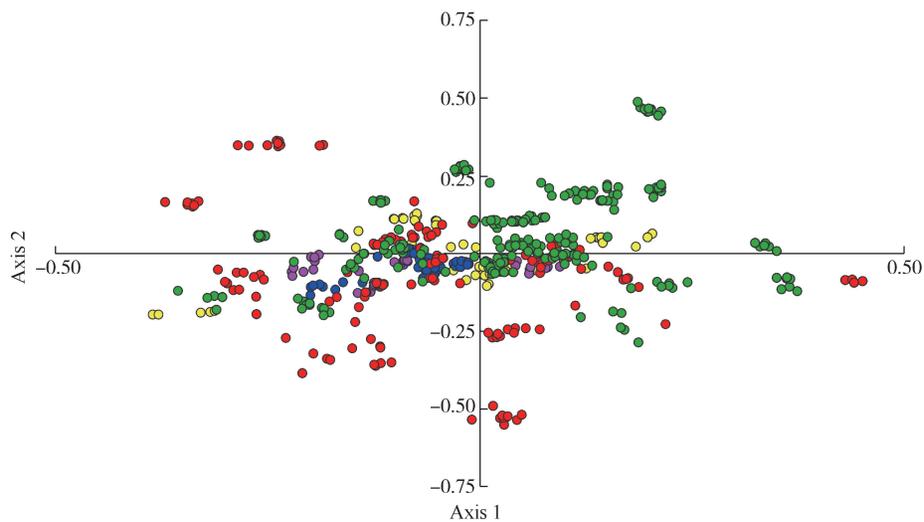


图 5 24 份忍冬属材料叶绿体基因组对应性分析的轴贡献率
Fig. 5 Axis contribution rate of correspondence analysis on chloroplast genomes of 24 *Lonicera* Linn. materials



●: 光合作用相关基因 Photosynthesis related genes; ●: 核糖体蛋白相关基因 Ribosomal protein related genes; ●: 自我复制相关基因 Self-replicating related genes; ●: *ycf* 基因 *ycf* genes; ●: 其他基因 Other genes.

图 6 24 份忍冬属材料叶绿体基因组的对应性分析
Fig. 6 Correspondence analysis on chloroplast genomes of 24 *Lonicera* Linn. materials

2.6 最优密码子分析

24 份忍冬属材料叶绿体基因组相对同义密码子使用度 (RSCU) 见表 6; 通过比较高、低表达组的 RSCU 差值 (Δ RSCU) 对叶绿体基因组最优密码子进

行分析, 结果见表 7。

结果(表 6) 显示: 24 份忍冬属材料叶绿体基因组密码子的 RSCU 值是高度可变的。例如: 编码 Leu 的 UUA 的 RSCU 值(1.73) 是其同义密码子 CUG

表 6 24 份忍冬属材料叶绿体基因组相对同义密码子使用度 (RSCU)
Table 6 Relative synonymous codon usage (RSCU) in chloroplast genomes of 24 *Lonicera* Linn. materials

氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU	氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU	氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU
Ala	GCU	12 927	1.63	Arg	CGU	7 379	1.48	Val	AUA	15 664	0.95
	GCC	5 162	0.73		CGC	2 319	0.43		GUU	11 931	1.47
	GCA	8 516	1.12		CGA	8 538	1.27		GUC	4 336	0.51
	GCG	3 697	0.52		CGG	3 269	0.43		GUA	11 499	1.43
Tyr	UAU	17 724	1.51	AGA	12 597	1.60	GUG	5 203	0.60		
	UAC	5 386	0.42		AGG	5 905		0.78	Ser	UCU	14 086
His	CAU	11 716	1.50	Gly	GGU	12 113	1.25	UCC		8 171	1.01
	CAC	3 530	0.60		GGC	4 551	0.45	UCA		9 109	1.14
Gln	CAA	15 997	1.55	GGA	14 928	1.50	UCG	4 698		0.51	
	CAG	5 830	0.45		GGG	8 482	0.81	AGU	9 512	1.31	
Asn	AAU	22 070	1.45	Phe	UUU	23 168	1.35	AGC	3 586	0.42	
	AAC	7 833	0.51		UUC	13 724	0.66		Pro	CCU	9 487
Lys	AAA	24 300	1.41	Leu	UUA	17 781	1.73	CCC		4 883	0.73
	AAG	10 620	0.55		UUG	14 080	1.34	CCA		7 926	1.20
Asp	GAU	21 214	1.55	CUU	13 822	1.22	CCG	3 788		0.59	
	GAC	5 363	0.45	CUC	5 612	0.50	Thr	ACU	5 905	0.78	
Glu	GAA	23 126	1.44	CUA	9 470	0.84		ACC	9 512	1.31	
	GAG	10 356	0.53	CUG	4 255	0.38		ACA	15 664	0.95	
Cys	UGU	5 312	1.29	Ile	AUU	23 277		1.44	ACG	10 952	0.61
	UGC	2 911	0.55		AUC	10 952	0.61				

表 7 24 份忍冬属材料叶绿体基因组最优密码子分析¹⁾
Table 7 Optimal codon analysis on chloroplast genomes of 24 *Lonicera* Linn. materials¹⁾

氨基酸 Amino acid	密码子 Codon	高表达组 High expression group		低表达组 Low expression group		Δ RSCU	氨基酸 Amino acid	密码子 Codon	高表达组 High expression group		低表达组 Low expression group		Δ RSCU
		数量 Number	RSCU	数量 Number	RSCU				数量 Number	RSCU	数量 Number	RSCU	
Ala	GCU	233	0.90	334	1.71	-0.81	Glu	GAC	0	0.00	210	0.54	-0.54
	GCC	197	0.76	154	0.79	-0.03		GAA	294	1.44	675	1.43	0.01
	GCA	523	2.02	155	0.79	1.23		GAG	113	0.56	267	0.57	-0.01
	GCG	82	0.32	137	0.70	-0.38		Cys	UGU	34	2.00	320	1.31
Tyr	UAU	212	1.25	561	1.32	-0.07	UGC		0	0.00	170	0.69	-0.69
	UAC	127	0.75	289	0.68	0.07	Arg	CGU	497	3.03	133	0.48	2.55
His	CAU	99	1.19	409	1.55	-0.36		CGC	48	0.29	38	0.14	0.15
	CAC	68	0.81	119	0.45	0.36	CGA	192	1.17	294	1.05	0.12	
Gln	CAA	285	1.89	364	1.11	0.78	CGG	0	0.00	299	1.07	-1.07	
	CAG	17	0.11	292	0.89	-0.78	AGA	240	1.46	543	1.95	-0.49	
Asn	AAU	212	1.38	698	1.27	0.11	AGG	6	0.04	366	1.31	-1.27	
	AAC	96	0.62	400	0.73	-0.11	Gly	GGU	164	1.05	265	0.94	0.11
Lys	AAA	881	1.66	1 007	1.43	0.23		GGC	48	0.31	116	0.41	-0.10
	AAG	178	0.34	398	0.57	-0.23		GGA	289	1.86	371	1.32	0.54
Asp	GAU	305	2.00	562	1.46	0.54		GGG	121	0.78	371	1.32	-0.54

续表7 Table 7 (Continued)

氨基酸 Amino acid	密码子 Codon	高表达组 High expression group		低表达组 Low expression group		Δ RSCU	氨基酸 Amino acid	密码子 Codon	高表达组 High expression group		低表达组 Low expression group		Δ RSCU
		数量 Number	RSCU	数量 Number	RSCU				数量 Number	RSCU	数量 Number	RSCU	
Phe	UUU	286	1.38	605	1.10	0.28	Ser	UCU	65	0.54	397	1.51	-0.97
	UUC	127	0.62	494	0.90	-0.28		UCC	198	1.65	209	0.79	0.86
Leu	UUA	393	2.27	468	1.28	0.99	UCA	131	1.09	320	1.21	-0.12	
	UUG	176	1.02	440	1.21	-0.19	UCG	116	0.96	128	0.49	0.47	
	CUU	216	1.25	324	0.89	0.36	AGU	195	1.62	231	0.88	0.74	
	CUC	82	0.47	326	0.89	-0.42	AGC	17	0.14	297	1.13	-0.99	
Ile	CUA	99	0.57	452	1.24	-0.67	Pro	CCU	17	0.20	230	1.03	-0.83
	CUG	72	0.42	179	0.49	-0.07		CCC	113	1.32	188	0.84	0.48
	AUU	509	1.75	785	1.14	0.61		CCA	156	1.82	292	1.31	0.51
Val	AUC	17	0.06	587	0.85	-0.79	CCG	56	0.65	185	0.83	-0.18	
	AUA	345	1.19	697	1.01	0.18	Thr	ACU	356	3.02	360	1.44	1.58
	GUU	311	2.07	293	0.91	1.16		ACC	0	0.00	249	1.00	-1.00
	GUC	34	0.23	213	0.67	-0.44		ACA	99	0.84	230	0.92	-0.08
	GUA	205	1.36	432	1.35	0.01		ACG	17	0.14	162	0.65	-0.51
	GUG	51	0.34	343	1.07	-0.73							

¹⁾ RSCU: 相对同义密码子使用度 Relative synonymous codon usage; Δ RSCU: 高、低表达组的 RSCU 差值 RSCU difference between high and low expression groups.

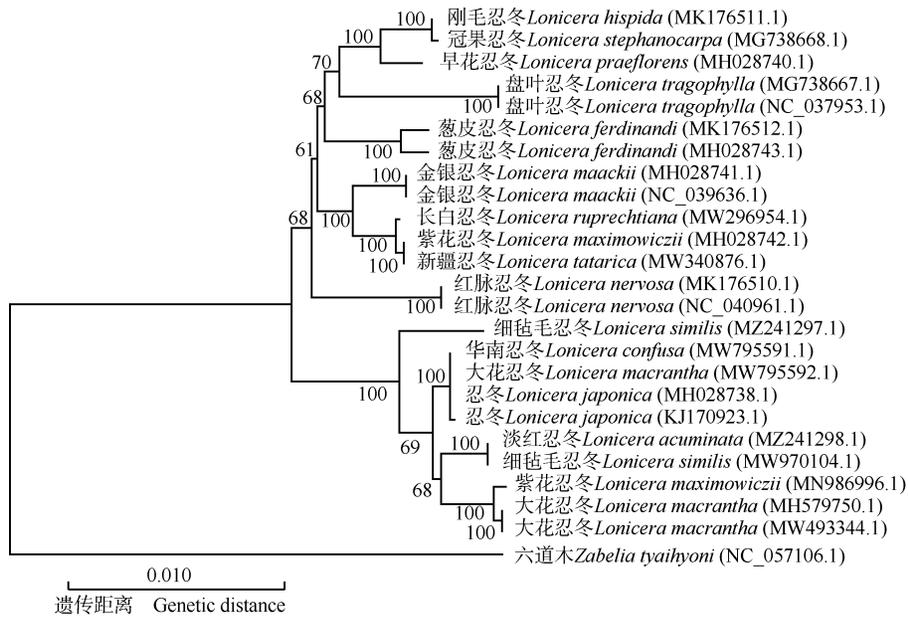
(0.38) 的 4.6 倍。RSCU 值大于 1 的 29 个密码子被定义为高频密码子, 分别为 UUA、GCU、UCU、AGA、CAA、GAU、UAU、CAU、GGA、CGU、GUU、AAU、GAA、AUU、GUA、AAA、UUU、UUG、CCU、AGU、ACC、UGU、CGA、GGU、CUU、CCA、UCA、GCA 和 UCC, 其中, 26 个密码子以 A 或 U 结尾, 2 个密码子以 C 结尾, 仅 1 个密码子以 G 结尾。

结果(表 7)还显示: 高、低表达组的 Δ RSCU 值大于 0.08 的密码子有 24 个, 分别为 CGU、ACU、GCA、GUU、UUA、UCC、CAA、AGU、UGU、AUU、GAU、GGA、CCA、CCC、UCG、CAC、CUU、UUU、AAA、AUA、CGC、CGA、AAU 和 GGU, 其中, 11 个密码子以 U 结尾, 8 个密码子以 A 结尾, 4 个密码子以 C 结尾, 仅 1 个密码子以 G 结尾。最终筛选出 18 个最优密码子, 即 GCA、CAA、AAU、AAA、GAU、UGU、CGU、CGA、GGU、GGA、UUU、UUA、CUU、AUU、GUU、UCC、AGU 和 CCA, 其中, 10 个密码子以 U 结尾, 7 个密码子以 A 结尾, 仅 1 个密码子以 C 结尾, 没有以 G 结尾的密码子。证明了 24 份忍冬属材料叶绿体基因组密码子偏好以 A 或 U 结尾。

2.7 系统发育分析

以六道木(NC_057106.1)为外类群, 基于 24 份忍冬属材料叶绿体基因组序列构建系统发育树。结

果(图 7)显示: 24 份忍冬属材料聚为一支, 进一步可分为 2 个分支, 第 1 个分支包括刚毛忍冬(*Lonicera hispida* Pall. ex Roem. et Schult.) (MK176511.1)、冠果忍冬(*L. stephanocarpa* Franch.) (MG738668.1)、早花忍冬(*L. praeflorens* Batal.) (MH028740.1)、盘叶忍冬(*L. tragophylla* Hemsl.) (MG738667.1 和 NC_037953.1)、葱皮忍冬(*L. ferdinandi* Franch.) (MK176512.1 和 MH028743.1)、金银忍冬 [*L. maackii* (Rupr.) Maxim.] (MH028741.1 和 NC_039636.1)、长白忍冬(*L. ruprechtiana* Regel) (MW296954.1)、紫花忍冬 [*L. maximowiczii* (Rupr.) Regel] (MH028742.1)、新疆忍冬(*L. tatarica* Linn.) (MW340876.1) 和 红脉忍冬(*L. nervosa* Maxim.) (MK176510.1 和 NC_040961.1), 第 2 个分支包括细毡毛忍冬(*L. similis* Hemsl.) (MZ241297.1 和 MW970104.1)、华南忍冬 [*L. confusa* (Sweet) DC.] (MW795591.1)、大花忍冬 [*L. macrantha* (D. Don) Spreng.] (MW795592.1、MH579750.1 和 MW493344.1)、忍冬 (MH028738.1 和 KJ170923.1)、淡红忍冬(*L. acuminata* Wall.) (MZ241298.1) 和 紫花忍冬 [*L. maximowiczii* (Rupr.) Regel] (MN986996.1), 聚为一支的物种间进化相似且亲缘关系较近。此外, 除细毡毛忍冬、紫花忍冬和大花忍冬外, 同一物种的不同材料均聚在一起。



系统发育树分支上的数据为自展支持率,括号内编号为 GenBank 登录号 The data on the branches of phylogenetic tree are the bootstrap values, and the Nos. in brackets are GenBank accession numbers.

图7 24份忍冬属材料叶绿体全基因组序列的系统发育树

Fig. 7 Phylogenetic tree of chloroplast genome sequences of 24 *Lonicera* Linn. materials

3 讨论和结论

3.1 忍冬属植物叶绿体基因组密码子特征

叶绿体是植物细胞中与光合作用相关的特殊细胞器,具有独立于核基因组的完整基因表达体系。因此,叶绿体基因组对于系统发育和分类研究具有十分重要的意义^[3,7]。基因组中蛋白质编码基因的GC含量在物种内的个体间大多是保守的。核基因组中蛋白质编码基因的GC含量较高,一般在40%以上,如拟南芥 [*Arabidopsis thaliana* (Linn.) Heynh.] (44.44%) 和烟草(43.01%)^[28],而叶绿体基因组中蛋白质编码基因的GC含量一般小于40%,如马尾松 (*Pinus massoniana* Lamb.) (39.81%)^[29]、陆地棉 (*Gossypium hirsutum* Linn.) (37.89%)^[26] 和糜子 (*Panicum miliaceum* Linn.) (39.36%)^[30]。密码子第1、第2和第3碱基位置的GC含量(GC₁、GC₂和GC₃)也不同。在叶绿体基因组中,一般的规律是GC₁值最高、GC₂值次之、GC₃值最低,如蒺藜苜蓿 (*Medicago truncatula* Gaertn.) 中GC₁、GC₂和GC₃值分别为45.50%、36.80%和26.90%^[31]。本研究通过对24份忍冬属材料叶绿体基因组的碱基组成分析发现,

1314个蛋白质编码基因的GC含量为38.7%,密码子3个位置的碱基组成显示出不同的特征,GC₁值最高(45.6%),其次是GC₂值(39.1%),GC₃值最低(31.4%),由此可见24份忍冬属材料的叶绿体基因组蛋白质编码区序列富含A或U,密码子第3位碱基以A或U为主。相反,在核基因组中GC₃值与GC₂值相似或高于GC₂值,如双子叶植物拟南芥(GC₃值为43.08%,GC₂值为35.53%)和烟草(GC₃值为38.23%,GC₂值为36.72%)^[28];单子叶植物核基因组密码子第3碱基位置对G或C有极高的偏好性,如水稻(*Oryza sativa* Linn.) (GC₃值为70.14%)和玉米(GC₃值为69.77%)^[28]。因此,密码子在叶绿体基因组中比在核基因组中保守性更强,表明叶绿体基因组可能是独立于核基因组起源和进化的。

3.2 忍冬属植物叶绿体基因组密码子使用偏好性

密码子的有效密码子数(ENC)一般在20~61之间,ENC值较低代表密码子使用偏好性较高。ENC值为35通常被认为是密码子使用偏好性的阈值^[19]。在24份忍冬属材料叶绿体基因组中,ENC值为50.2,表明密码子使用偏好性较低。筛选出的最优密码子GCA、CAA、AAU、AAA、GAU、UGU、CGU、CGA、GGU、GGA、UUU、UUA、CUU、AUU、GUU、UCC、AGU和CCA

的表现形式为 NNA 或 NNU 的模式。这与陆地棉^[26]、蒙古韭(*Allium mongolicum* Regel)^[32]和油茶(*Camellia oleifera* Abel.)^[33]等物种的研究结果一致,不同物种间密码子使用偏好性相同,原因可能与叶绿体基因组的相对保守性有关。

Liu 等^[34]发现物种内遗传距离与密码子使用偏好性特征之间可能存在一定的关系,遗传关系越紧密,密码子使用偏好性越相似。本研究中,除细毡毛忍冬、紫花忍冬和大花忍冬外,同一物种不同材料均聚在一起,推测可能是上述 3 个物种不同材料的基因组大小相差较大,且基因组中含有较多的非编码序列,影响了基因组间遗传距离计算结果的准确度。但整体上 24 份忍冬属材料聚为一支,各物种间的进化相似且亲缘关系较近,说明基于叶绿体基因组序列构建的系统发育树在物种鉴定和系统发育方面具有一定的代表性。

3.3 密码子使用偏好性与基因表达的关系

GC 含量高的基因比 GC 含量低的基因有更高的基因表达水平^[35]。密码子中含有 3 个氢键对的 G 和 C 末端可以有效地增加蛋白质的结合强度,促进蛋白质翻译的准确性和效率。本研究中,光合作用相关基因的总 GC 含量最高,为 39.2%。说明其在植物细胞中扮演着不可或缺的角色。

本研究考虑了最优密码子的数量和类型与基因表达的关系。24 份忍冬属材料叶绿体基因组中的 18 个最优密码子多以 A 或 U 结尾,但核基因中的最优密码子表现不同。例如:单子叶植物水稻和玉米的最优密码子多以 G 或 C 结尾,双子叶植物拟南芥和烟草的最优密码子多以 A 或 T 结尾^[28]。相关分析结果显示:最优密码子使用频率(FOP)与密码子适应指数(CAI)呈极显著正相关。光合作用相关基因具有较高的 FOP 值(0.36)和最高的 CAI 值(0.18)。在 54 种蛋白质编码基因中,*rbcL* 基因的 FOP 值最高(0.48),表达水平(CAI 值为 0.27)也最高。密码子使用偏好性表现出基因和物种的特异性,这种特异性可以影响基因的表达水平,通常情况下密码子使用偏好性越强,基因表达水平越高^[36]。Gustafsson 等^[37]发现密码子的碱基组成对基因在细胞中的表达水平有很大的影响,通过对特定密码子的修饰,可以使基因表达水平提高 1 000 倍以上。本研究 24 份忍冬属材料叶绿体基因组中包含最优密码子的基因可能表现出较高的表达水平。

3.4 密码子使用偏好性的影响因子分析

对应性分析结果表明多种因子参与了 24 份忍冬属材料叶绿体基因组密码子使用偏好性的调控。中性绘图、PR2-plot 和 ENC-plot 分析结果表明 24 份忍冬属材料叶绿体基因组的密码子使用偏好性受突变压力和自然选择的影响,其中自然选择的影响较大。因此,在决定 24 份忍冬属材料叶绿体基因组密码子使用偏好性时,自然选择起主要作用。这与王占军等^[9]的研究结果相似,但在乳油木(*Vitellaria paradoxa* C. F. Gaertn.)中密码子使用偏好性主要受突变压力的影响^[38],在芒果(*Mangifera indica* Linn.)中突变压力与自然选择的影响相对均衡^[39]。故推测不同物种间密码子使用偏好性的影响因子有差异,但自然选择和突变压力均发挥了一定作用。

3.5 结论

24 份忍冬属材料叶绿体基因组共有 1 314 个蛋白质编码基因,其中光合作用相关基因的总 GC 含量最高;密码子 3 个位置的碱基组成显示出不同的特征,第 1 碱基位置的 GC 含量最高、第 2 碱基位置的 GC 含量次之、第 3 碱基位置的 GC 含量最低;不同忍冬属材料叶绿体基因组密码子使用模式较为接近,密码子使用偏好性较低,密码子末端优先使用 A 或 U;密码子使用偏好性受突变压力和自然选择的双重影响,其中自然选择发挥了主要作用。忍冬属植物叶绿体基因组的上述特征为探究其进化规律奠定了基础。本研究仅选取了 NCBI 数据库中已有的 24 份忍冬属材料,研究结果存在一定的局限性,后期本课题组计划进行忍冬属 100 余种植物的收集和测序工作,为忍冬属植物叶绿体的进一步研究和应用奠定基础。

参考文献:

- [1] 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.
- [2] SUGIURA M, SHINOZAKI K, ZAITA N, et al. Clone bank of the tobacco (*Nicotiana tabacum*) chloroplast genome as a set of overlapping restriction endonuclease fragments; mapping of eleven ribosomal protein genes[J]. Plant Science, 1986, 44: 211-217.
- [3] 舒军霞, 杨林, 周涛, 等. 刺柏属 4 种药用植物叶绿体基因组密码子偏好性分析[J]. 中草药, 2022, 53(23): 7507-7515.
- [4] 赵月梅, 杨振艳, 赵永平, 等. 木犀科植物叶绿体基因组结构特征和系统发育关系[J]. 植物学报, 2019, 54(4): 441-454.
- [5] LI S F, SU T, CHENG G Q, et al. Chromosome evolution in

- connection with repetitive sequences and epigenetics in plants[J]. *Genes*, 2017, 8: 290.
- [6] NOCK C J, WATERS D L E, EDWARDS M A, et al. Chloroplast genome sequences from total DNA for plant identification[J]. *Plant Biotechnology Journal*, 2011, 9: 328–333.
- [7] CHEN X L, ZHOU J G, CUI Y X, et al. Identification of *Ligularia* herbs using the complete chloroplast genome as a super-barcode[J]. *Frontiers in Pharmacology*, 2018, 9: 695.
- [8] 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 Research*, 1987, 15 (3): 1281–1295.
- [9] 王占军, 吴子琦, 王朝霞, 等. 3个茶树品种 *WOX* 基因家族的进化及密码子偏好性比较[J]. *南京林业大学学报(自然科学版)*, 2022, 46(2): 71–80.
- [10] 严其伟, 梁湘兰, 陶秋云, 等. 老班瑶药牛耳风叶绿体基因组特征及其密码子使用偏好性分析[J]. *中国农业科技导报*, 2022, 24(8): 74–86.
- [11] 刘汉梅, 何 瑞, 张怀渝, 等. 玉米叶绿体基因密码子使用频率分析[J]. *四川农业大学学报*, 2010, 28(1): 10–14.
- [12] 杨祥燕, 蔡元保, 谭秦亮, 等. 菠萝叶绿体基因组密码子偏好性分析[J]. *热带作物学报*, 2022, 43(3): 439–446.
- [13] 季凯凯, 宋希强, 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.
- [14] 穆赢通, 樊东昌, 吕丽娟, 等. 毛茛科和芍药科叶绿体基因组密码子特征和系统发育比较[J]. *植物研究*, 2022, 42(6): 964–975.
- [15] 中国科学院中国植物志编辑委员会. *中国植物志: 第七十二卷* [M]. 北京: 科学出版社, 1988: 143.
- [16] ZHOU L K, ZHOU Z, JIANG X M, et al. Absorbed plant MIR2911 in honeysuckle decoction inhibits SARS-CoV-2 replication and accelerates the negative conversion of infected patients[J]. *Cell Discovery*, 2020, 6: 54.
- [17] 刘安成, 毕 毅, 尉 倩, 等. 中国忍冬属植物育种研究进展[J]. *北方园艺*, 2022(5): 133–139.
- [18] HAN A X, QIU J, CAO R M, et al. Shrubs should be valued: the functional traits of *Lonicera fragrantissima* var. *lancifolia* in a Qinling Huangguan Forest dynamics plot, China [J]. *Forests*, 2022, 13(7): 1147.
- [19] WRIGHT F. The ‘effective number of codons’ used in a gene[J]. *Gene*, 1990, 87: 23–29.
- [20] OUYANG T L, ZHONG J C, CHAI Z X, et al. Codon usage bias and cluster analysis of the *MMP-2* and *MMP-9* genes in seven mammals[J]. *Genetics Research*, 2022, 2022: 2823356.
- [21] WEI L, HE J, JIA X, et al. Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution [J]. *BMC Evolutionary Biology*, 2014, 14: 262.
- [22] ZHANG W J, ZHOU J, LI Z F, et al. Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L. [J]. *Journal of Integrative Plant Biology*, 2007, 49(2): 246–254.
- [23] 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: 53–58.
- [24] ZHOU M, TONG C F, SHI G S. Analysis of codon usage between different poplar species [J]. *Journal of Genetics and Genomics*, 2007, 34(6): 555–561.
- [25] SHARP P M, TUOHY T M F, MOSURSKI K R. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes [J]. *Nucleic Acids Research*, 1986, 14(13): 5125–5143.
- [26] 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.
- [27] KUMAR S, STECHER G, LI M, et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms [J]. *Molecular Biology and Evolution*, 2018, 35(6): 1547–1549.
- [28] KAWABE A, MIYASHITA N T. Patterns of codon usage bias in three dicot and four monocot plant species[J]. *Genes and Genetic Systems*, 2003, 78: 343–352.
- [29] 叶友菊, 倪州猷, 白天道, 等. 马尾松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*, 2018, 37(10): 4464–4471.
- [30] 刘 慧, 王梦醒, 岳文杰, 等. 糜子叶绿体基因组密码子使用偏性的分析[J]. *植物科学学报*, 2017, 35(3): 362–371.
- [31] 杨国锋, 苏昆龙, 赵怡然, 等. 蒺藜苜蓿叶绿体密码子偏好性分析[J]. *草业学报*, 2015, 24(12): 171–179.
- [32] 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.
- [33] 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.
- [34] LIU S X, XUE D Y, CHENG R, et al. The complete mitogenome of *Apocheima cinerarius* (Lepidoptera: Geometridae: Ennominae) and comparison with that of other lepidopteran insects [J]. *Gene*, 2014, 547: 136–144.
- [35] GUO X Y, BAO J D, FAN L J. Evidence of selectively driven codon usage in rice: implications for GC content evolution of *Gramineae* genes [J]. *FEBS Letters*, 2007, 581: 1015–1021.
- [36] ZHOU Z P, DANG Y K, ZHOU M, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(41): E6117–E6125.
- [37] GUSTAFSSON C, GOVINDARAJAN S, MINSHULL J. Codon bias and heterologous protein expression [J]. *Trends in Biotechnology*, 2004, 22(7): 346–353.
- [38] 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.
- [39] 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.