

## 怀玉山产三叶青叶绿体基因组特征及其系统进化关系

洪森荣<sup>1,2,3,4</sup>, 刘雯莉<sup>1</sup>, 宋冰雁<sup>1</sup>, 颜玉情<sup>1</sup>

1. 上饶师范学院生命科学学院, 江西 上饶 334001

2. 上饶农业技术创新研究院, 江西 上饶 334001

3. 上饶市药食同源植物资源保护与利用重点实验室, 江西 上饶 334001

4. 上饶市三叶青保育与利用技术创新中心, 江西 上饶 334001

**摘要:** 目的 解析怀玉山产三叶青 *Tetrastigma hemsleyanum* 叶绿体基因组信息序列特征和确定其在崖爬藤属的系统位置。方法 用 Illumina 高通量测序平台 NovaSeq6000 进行测序获得怀玉山产三叶青叶绿体基因组序列, 借助 GeSeq、tRNAscan-SE、MISA、VISTA tools、DNADnaSP6.0、JSHYCloud、CodonW1.4.2、Pasteur Galaxy、mafft 7.0、fasttree 2.1.10 等生物信息学工具进行序列分析、密码子偏好分析、崖爬藤属基因组比较分析和系统发育研究。结果 怀玉山产三叶青 *Tetrastigma hemsleyanum* 叶绿体基因组为共价闭合双链环状分子, 长 160 165 bp, 包含 1 个大单拷贝区 (large single copy region, LSC)、1 个反向重复区 a (inverted repeat region a, IRa)、1 个反向重复区 b (inverted repeat region b, IRb) 和 1 个小单拷贝区 (small single copy region, SSC); 怀玉山产三叶青叶绿体基因组注释到光合作用基因、自我复制基因、其他基因和未知功能基因 4 类共 133 个基因, 包括 88 个 CDS 基因、37 个 tRNA 基因和 8 个 rRNA 基因; 怀玉山产三叶青叶绿体基因组检测到 63 个 SSR 位点, 其中单碱基重复 56 个, 双碱基重复 7 个; 怀玉山产三叶青叶绿体基因组核苷酸多样性的变化范围为 0~0.234 18, 基因间隔区 *trnS-GCU-trnG-UCC*、*rps16-trnG-UCC* 以及基因 *ycf1* 变异率最高。怀玉山产三叶青叶绿体基因组密码子偏好以 A 或 U 结尾, 具有较强的偏好性。突变对密码子偏好性的影响较强, 而其他因素如自然选择对密码子的偏好性影响较小。怀玉山产三叶青叶绿体基因组有 31 个高频密码子, 16 个以 U 结尾, 13 个以 A 结尾, 2 个以 G 结尾, 其中 13 个为最优密码子。怀玉山产三叶青与福建产三叶青 (MW375708) 和广西产三叶青 (MW375709) 亲缘关系较近, 与浙江产三叶青 (MT827073)、浙江产三叶青 (MW375707)、浙江产三叶青 KT033563 (NC\_029339)、四川产三叶青 (MW375710) 和浙江产三叶青 (MW375711) 亲缘关系较远。结论 三叶青之间的亲缘关系与地理位置相关性不大。首次采用生物信息学分析方法对怀玉山产三叶青叶绿体基因组进行了全面、深度解析, 结果将为崖爬藤属药用植物的遗传多样性分析以及新品系的遗传育种研究等提供理论依据。

**关键词:** 三叶青; 叶绿体基因组特征; 系统进化关系; 大单拷贝区; 小单拷贝区

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2023)16-5358-14

DOI: 10.7501/j.issn.0253-2670.2023.16.025

## Phylogenetic relationship and characterization of chloroplast genome of *Tetrastigma hemsleyanum* from Huaiyu Mountain

HONG Sen-rong<sup>1,2,3,4</sup>, LIU Wen-li<sup>1</sup>, SONG Bing-yan<sup>1</sup>, YAN Yu-qing<sup>1</sup>

1. College of Life Sciences, Shangrao Normal University, Shangrao 334001, China

2. Shangrao Agricultural Technology Innovation Research Institute, Shangrao 334001, China

3. Key Laboratory of protection and utilization of medicinal and edible plant resources in Shangrao City, Shangrao 334001, China

4. Shangrao *Tetrastigma hemsleyanum* Diels et Gilg Conservation and Utilization Technology Innovation Center, Shangrao 334001, China

**Abstract: Objective** To analyze the genomic information sequence characteristics of the chloroplast of *Tetrastigma hemsleyanum* from Huaiyu Mountain and determine its systematic position in the genus of *Tetrastigma*. **Methods** The chloroplast genome sequence of *T. hemsleyanum* from Huaiyu Mountain was obtained by sequencing with Illumina high throughput sequencing platform NovaSeq6000, and

收稿日期: 2023-02-03

基金项目: 国家自然科学基金资助项目 (31960079); 2022 年上饶市科技专项项目 (饶科发[2023]5 号社发类) (2022A008); 江西省科技厅重点研发计划一般项目 (20202BBG73010); 江西省教育厅科学技术研究项目 (GJJ201704); 上饶市科技局平台载体建设项目 (2020J001)

作者简介: 洪森荣 (1974—), 教授, 主要从事植物生物技术研究工作。Tel: (0793)8153721 E-mail: hongsenrong@163.com

bioinformatics tools such as GeSeq, tRNAscan SE, MISA, VISTA tools, DNADnaSP6.0, JSHYCloud, CodonW1.4.2, Pasteur Galaxy, Mafft 7.0, fasttree 2.1.10 were used for sequence analysis, codon preference analysis, genome comparative analysis and phylogenetic research of *Tetrastigma*. **Results** The chloroplast genome of *T. hemsleyanum* from Huaiyu Mountain is a covalently closed double stranded ring molecule with a length of 160 165 bp, including 1 LSC, 1 IRa, 1 IRb and 1 SSC; the chloroplast genome of *T. hemsleyanum* from Huaiyu Mountain has been annotated with 133 genes in four categories: photosynthesis genes, self-replication genes, other genes and unknown functional genes, including 88 CDS genes, 37 tRNA genes and eight rRNA genes; Sixty three SSR loci were detected in the chloroplast genome of *T. hemsleyanum* from Huaiyu Mountain, including 56 single base repeats and seven double base repeats; The variation range of nucleotide diversity of the chloroplast genome of *T. hemsleyanum* from Huaiyu Mountain was 0—0.234 18, and the variation rates of *trnS-GCU-trnG UCC*, *rps16-trnG UCC* and gene *ycf1* were the highest. The codon of chloroplast genome of *T. hemsleyanum* from Huaiyu Mountain had a strong preference for ending in A or U. Mutation had a strong impact on codon preference, while other factors such as natural selection had a small impact on codon preference. The chloroplast genome of *T. hemsleyanum* from Huaiyu Mountain had 31 high-frequency codons, 16 ending in U, 13 ending in A, and two ending in G, of which 13 were the optimal ones. *T. hemsleyanum* from Huaiyu Mountain was close to *T. hemsleyanum* from Fujian MW375708 and *T. hemsleyanum* from Guangxi MW375709, and far from *T. hemsleyanum* from Zhejiang MT827073, *T. hemsleyanum* from Zhejiang MW375707, *T. hemsleyanum* from Zhejiang KT033563 (NC\_029339), *T. hemsleyanum* from Sichuan MW375710 and *T. hemsleyanum* from Zhejiang MW375711, which showed that there was little correlation between the genetic relationship and the geographical location. **Conclusion** The genetic relationship between *T. hemsleyanum* is not closely related to its geographical location. In this study, the chloroplast genome of *T. hemsleyanum* from Huaiyu Mountain is analyzed comprehensively and deeply for the first time by using bioinformatics analysis method, which will provide theoretical basis for genetic diversity analysis of medicinal plants and genetic breeding research of new strains of *Tetrastigma*.

**Key words:** *Tetrastigma hemsleyanum* Diels et Gilg; characterization of chloroplast genome; phylogenetic relationship; large single copy region; small single copy region

三叶青 *Tetrastigma hemsleyanum* Diels et Gilg 为我国特有珍稀濒危葡萄科崖爬藤属药用植物<sup>[1]</sup>, 其新鲜或干燥块根可清热解毒、消肿止痛、化痰散结, 常用于治疗小儿高热惊风、百日咳、毒蛇咬伤等<sup>[2]</sup>; 其干燥全草也可清热解毒、消肿止痛、活血祛风, 常用于治疗流行性感、肝炎、泌尿系统结石、跌打损伤等<sup>[3]</sup>。

叶绿体是为绿色植物提供能量的光合细胞器, 在光合作用及相关代谢活动中起重要作用<sup>[4]</sup>。随着生物技术的开发和应用, 叶绿体基因组大小已经确定, 被子植物中典型的叶绿体基因组为双链共价闭合环状分子, 具有典型的四分体结构, 包括一个大单拷贝区 (large single copy region, LSC), 小单拷贝区 (small single copy region, SSC), 反向重复区 a (inverted repeat region a, IRa), 反向重复区 b (inverted repeat region b, IRb), 植物叶绿体基因组的大小差异由 IRa 和 IRb 决定<sup>[5]</sup>。被子植物的叶绿体基因组的长度一般在 120~170 kb, 通常编码 110~130 个基因, 约 40 个基因是参与光合作用、转录和翻译<sup>[6]</sup>。叶绿体基因结构高度保守, 确保在进化群体之间基因的直接同源性, 能自我复制, 多为母系遗传, 进化不重组, 直接反映植物进化积累的遗传变异, 可为高等植物群体遗传学、系统发育

和基因组进化提供依据<sup>[7]</sup>。

三叶崖爬藤植物在形态上很难区分, 被认为是崖爬藤属中最难分类的一个属, 许多专家学者从 DNA 分子标记和形态学角度对三叶崖爬藤植物进行了相关的研究, 但这些分类系统还是存在较大差异, 究其原因, 一是形态学的传统物种分类方法易受环境因子影响, 二是 DNA 分子标记关注的侧重点存在差异<sup>[8]</sup>。植物叶绿体基因组远小于核基因组, 编码区和非编码区的分子进化速度差异显著, 各类群叶绿体基因组之间共线性良好, 随着系统发育学和基因组学的交融, 在植物系统发育研究中, 基于叶绿体基因组的系统发育基因组学研究优势渐渐凸显, 为一些分类困难类群的系统学问题提供了新的解决方案<sup>[9-10]</sup>。

国内已经有浙江三叶崖爬藤<sup>[9-10]</sup>叶绿体基因组的相关报道, 但未涉及到崖爬藤属植物叶绿体基因组密码子偏好性分析。三叶青种质遗传多样性十分丰富, 浙江产三叶青种质之间的遗传信息有一定差异<sup>[11]</sup>。怀玉山产三叶青为江西省上饶市玉山全县特产, 2016年3月31日原中华人民共和国农业部正式批准对“怀玉山产三叶青”实施农产品地理标志登记保护<sup>[12]</sup>。本研究对怀玉山产三叶青叶绿体基因组进行测序, 对其 GC 含量、SSR 位点等进

行分析,并与不同产地三叶崖爬藤以及崖爬藤属其他物种的基因组进行多重比较,以期为怀玉山产三叶青的系统发育和不同产地三叶青的亲缘关系提供数据支撑,同时丰富三叶崖爬藤的叶绿体基因组数据库,为怀玉山产三叶青的遗传多样性保护、生产性状等相关研究提供理论基础,也为怀玉山产三叶青种质鉴定、DNA条形码开发提供新思路。

## 1 材料与仪器

### 1.1 材料

怀玉山产三叶青栽培种“怀玉2号”试管苗(由上饶市三叶青保育与利用技术创新中心提供)。由上饶市红日农业开发有限公司陈荣华高级农艺师鉴定为三叶青 *T. hemsleyanum* Diels et Gilg 栽培种“怀玉2号”试管苗。

### 1.2 仪器

高通量二代测序 NextSeq 550AR 基因测序仪(Illumina 公司,美国)。

## 2 方法

### 2.1 DNA提取和测序

用 CTAB 法提取怀玉山产三叶青叶片 DNA, DNA 浓度利用 Qubit2.0 进行精确定量检测合格后,将 DNA 片段化(Covaris 超声波破碎仪随机打断),经过片段纯化、末端修复等途径获得测序文库。文库构建完成后,先用 Qubit2.0 进行初步定量,稀释文库,随后使用 Agilent 2100 对文库的插入片段进行检测,插入片段大小符合预期后,使用 Q-PCR 方法对文库的有效浓度进行准确定量,以保证文库质量。文库检测合格后,在天津诺禾致源科技有限公司用 Illumina 高通量测序平台 NovaSeq6000 进行测序。

### 2.2 基因的组装和注释

首先利用 SOAPdenovo v2.04 (<http://soap.genomics.org.cn/>) 拼接软件对质控后的全部 clean data 进行 novo 初步组装,将组装得到的全部 contig 与叶绿体数据库进行比对,筛选得到来自叶绿体的 contig; 然后利用 MITObim v1.6 通过迭代比对,将测序的所有 clean reads mapping 到这些 contig 上进行延伸,通过 gap close 获得叶绿体的全基因组序列。将怀玉山产三叶青叶绿体基因组序列上传至 NCBI (GenBank 登录号 OP589402)。使用 GeSeq、tRNAscan-SE 对叶绿体基因组进行注释,再经过手工校正后得到最终的叶绿体基因组基因注释结果。使用 OGDRAW 制作叶绿体基因组图谱。

### 2.3 叶绿体基因组特征分析

使用 MISA 在线工具 (<https://webblast.ipk-gatersleben.de/misa/index.php>) 的微卫星定位检测技术对怀玉山产三叶青叶绿体基因组序列中的简单重复序列(Simple sequence repeats, SSR) 进行搜索。用 VISTA tools 绘制结构 mVISTA 结构变异图。利用 DNADnaSP6.0 软件计算近缘物种的基因组 Pi。通过 JSHYCloud 在线工具集分析怀玉山产三叶青叶绿体基因组 IR 区边界结构差异,从 NCBI 中选崖爬藤属浙江产三叶青 *T. hemsleyanum* Diels et Gilg、扁担藤 *T. planicaule* (Hook.) Gagnep.、毛五叶崖爬藤 *T. voinierianum* Pierre ex Gagnep.、*T. lawsonii* Linnaeus、*Tetrastigma rafflesiae* (Miq.) Planch.; 葡萄属祖提葡萄 *Cyphostemma juttiae* (Dinter & Gilg) Desc.、矮葡萄瓮 *C. humile* (N. E. Br.) Desc. ex Wild & R. B. Drumm.、*C. adenopodum* (Sprague) Desc.、桑德森葡萄 *C. sandersonii* Linnaeus 和乌菟莓 *Causonis japonica* (Thunb.) Raf. 的 10 种植物叶绿体基因组来比对不同物种间叶绿体基因组异同。运用软件 CodonW1.4.2 分析怀玉山产三叶青叶绿体基因组的相对同义密码子使用情况(relative synonymous condon usage, RSC)。通过 Pasteur Galaxy 在线工具集 (<https://galaxy.pasteur.fr/CodonW>) 中的 CodonW 模块分析密码子使用情况,设置输出结果为有效密码子数(ENC)和相对同义密码子使用度(RSCU),其他参数设为默认值。ENC>35 表示含有较多种类的稀有密码子,且基因表达量偏低;某一密码子的 RSCU>1.00 表示编码对应的蛋白质时偏好使用该密码子,RSCU<1.00 表示不偏好使用该密码子,RSCU=1.00 表示该密码子没有偏好性。对近缘物种进行 GC<sub>3</sub>~GC<sub>12</sub> 分析(Neutrality-plot 分析)、ENC-plot 分析和 PR2-bias-plot 分析和最优密码子分析。将怀玉山产三叶青叶绿体基因组序列上传至 NCBI 进行 BLASTn 比对,选择 highly similar sequence (megablast) 来比较相似性在 95% 以上的序列,检索获得怀玉山产三叶青的近缘种,以明确怀玉山产三叶青叶绿体基因组序列在葡萄科中的系统关系。序列比对利用 mafft 7.0 软件,建树利用 fasttree 2.1.10 软件。

## 3 结果与分析

### 3.1 怀玉山产三叶青叶绿体基因组的结构特征

测序 raw reads 为 17 808 860 条。测序 clean reads 为 17 695 827 条。经组装和注释后怀玉山产三叶青叶

绿体基因组为共价闭合的双链环状分子(图1),长160 165 bp,具有典型的四分体结构,由1个LSC、2个IR(1个IRa和1个IRb)和1个SSC共4部分组成,其中,LSC、SSC、IRa和IRb区的长度分别为88 172、18 963、26 515和26 515 bp。基因组的总GC含量为37.50%,其中,LSC、SSC、IRa和IRb区的GC含量分别为35.50%、31.76%、42.89%和42.89%。

### 3.2 怀玉山产三叶青叶绿体基因类型分析

对怀玉山产三叶青叶绿体基因组进行在线注释,结果见表1。由表1可知,共注释到光合作用基因、自我复制基因、其他基因和未知功能基因4类,包括88个CDS基因、37个tRNA基因和8个rRNA基因,共133个基因。对有多个外显子的叶绿体基因进行结构分析,由2个外显子构成的基因有19个,包括8个CDS基因和13个tRNA基因,其中有2个基因在

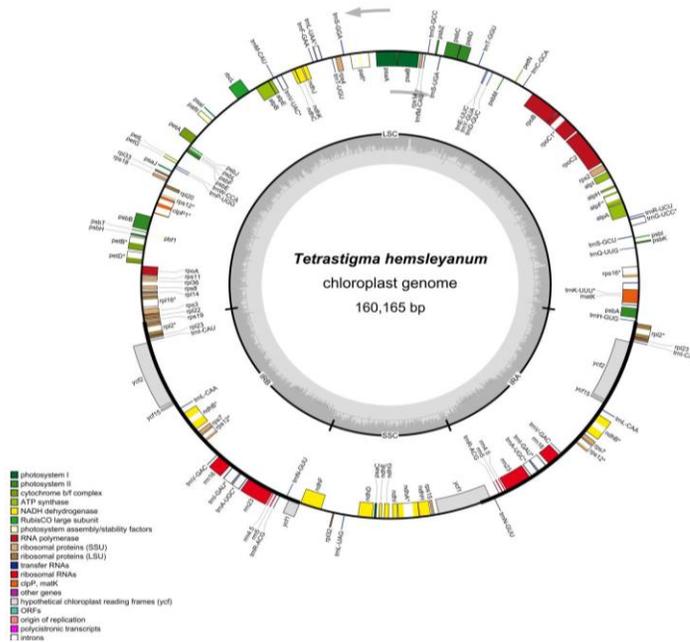


图1 怀玉山产三叶青叶绿体基因组图谱

Fig. 1 Chloroplast genome map of *T. hemsleyanum* from Huaiyu Mountain

表1 怀玉山产三叶青叶绿体基因功能分类

Fig.1 Chloroplast gene functional classification of *T. hemsleyanum* from Huaiyu Mountain

基因功能	基因类型	基因名	基因数量
光合作用	光系统	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	5
	光系统II	<i>psaJ</i> , <i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbT</i> , <i>psbZ</i>	14
自我复制	NADH 脱氢酶	<i>ndhA</i> , <i>ndhB</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>	12
	细胞色素 b/f 复合体	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>	6
	ATP 合成酶	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>	6
	核糖体大亚基蛋白质	<i>rpl14</i> , <i>rpl16</i> , <i>rpl2</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	11
	核糖体小亚基蛋白质	<i>rps11</i> , <i>rps12</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> , <i>rps18</i> , <i>rps19</i> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps8</i>	14
	核糖体大亚基	<i>rbcL</i>	1
	RNA 聚合酶	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>	4
	核糖体 RNA	<i>rrn16</i> , <i>rrn16</i> , <i>rrn23</i> , <i>rrn23</i> , <i>rrn4.5</i> , <i>rrn4.5</i> , <i>rrn5</i> , <i>rrn5</i>	8
	转运 RNA	<i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnF-GAA</i> , <i>trnG-GCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU</i> , <i>trnI-CAU</i> , <i>trnK-UUU</i> , <i>trnTERM-UUA</i> , <i>trnL-CAA</i> , <i>trnL-CAA</i> , <i>trnL-UAA</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> , <i>trnN-GUU</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> , <i>trnR-ACG</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> , <i>trnV-GAC</i> , <i>trnV-UAC</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i> , <i>trnY-CAU</i>	30
	其他基因	成熟酶	<i>matK</i>
蛋白酶		<i>clpP1</i>	1
囊膜蛋白		<i>cemA</i>	1
乙酰辅酶 A 羧化酶		<i>accD</i>	1
C-型细胞色素合成基因		<i>ccsA</i>	1
翻译起始因子		<i>infA</i>	1
未知功能基因	保守假设叶绿体阅读框架	<i>yef1</i> , <i>yef1</i> , <i>yef15</i> , <i>yef15</i> , <i>yef2</i> , <i>yef2</i>	6

IR 区重复；由 4 个外显子构成的基因有 4 个，均为 4 个 tRNA 基因；完全在 LSC 区的基因有 82 个；完全在 SSC 区的基因有 12 个；完全在 IRb 和 IRa 区的基因有 17 个；在 SSC-IRb junction 的基因有 1 个；在 LSC-IRb junction 的基因有 1 个；在 SSC-IRa junction 的基因有 1 个；在 LSC-IRa junction 无基因；rps12 有 2 个拷贝，每个拷贝具有 3 个外显子，且 2 个拷贝共享第一个外显子，第一个外显子位于 LSC 区域，另外 2 个外显子位于 IR 区域。

### 3.3 怀玉山产三叶青叶绿体基因重复序列分析

怀玉山产三叶青叶绿体基因组中 SSR 的类型

及分布见表 2。由表 2 可知，在怀玉山产三叶青叶绿体基因组中共检测到 63 个 SSR 位点，其中，单碱基重复有 56 个，双碱基重复有 7 个，其中，重复单元为 A、重复频率为 10 的 SSR 位点数量最多（7），重复单元为 T、重复频率为 10 的 SSR 位点数量最多（11），重复单元为 A/T 的最高重复频率为 18。

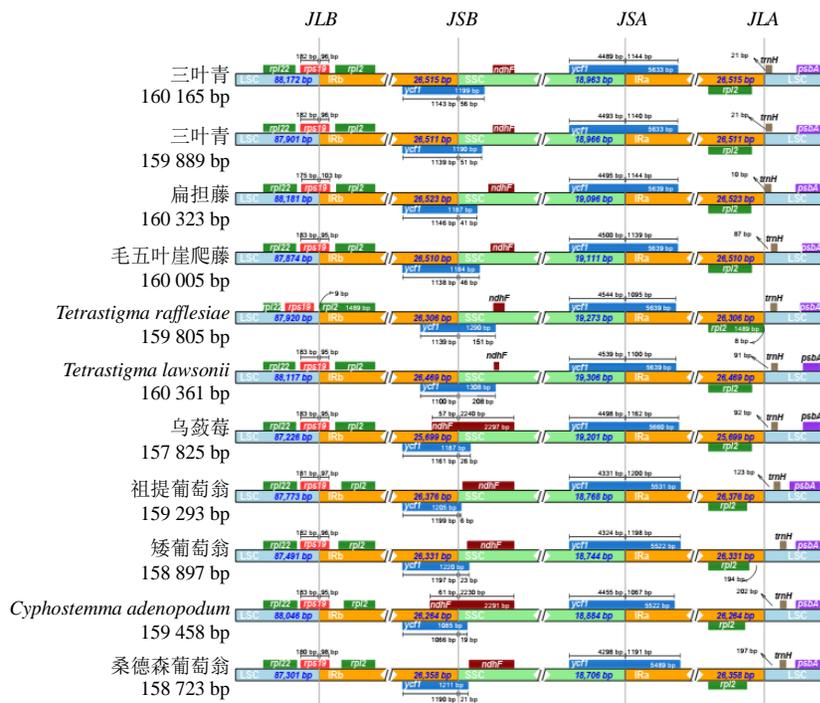
### 3.4 怀玉山产三叶青 IR 区边界结构差异分析

怀玉山产三叶青 IR 区边界结构差异分析见图 2。由图 2 可知，11 种葡萄科植物叶绿体基因组结构从 LSC 中间呈线性展开，均由 1 个 LSC 区、1 个

表 2 怀玉山产三叶青叶绿体基因组中 SSR 的类型及分布

Table 2 Type and distribution of SSR in chloroplast genome of *T. hemsleyanum* from Huaiyu Mountain

重复单元碱基类型	重复单元重复次数														总数	
	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19
A	-	-	-	-	-	7	5	5	2	2	-	2	-	-	-	23
C	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1
T	-	-	-	-	-	11	8	5	4	3	-	-	-	-	-	32
AT	-	-	1	1	-	2	-	-	-	-	-	-	-	-	-	4
CT	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
TA	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	2
A/T	-	-	-	-	-	18	13	10	6	5	-	2	-	-	1	55
C/G	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1
AG/CT	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AT/AT	-	1	2	1	-	2	-	-	-	-	-	-	-	-	-	6



JLB-LSC 和 IRb 区的边界 JSB-IRb 和 SSC 区的边界 JSA-SSC 和 IRa 区的边界 JLA-IRa 和 LSC 区的边界

JLB-boundary of LSC and IRb regions JSB-boundary of IRb and SSC regions JSA-boundary of SSC and IRa regions JLA-boundary of IRa and LSC regions

图 2 11 种葡萄科植物叶绿体基因组 LSC、SSC 和 IR 边界位置的比较

Fig. 2 Comparison on boundary locations of LSC, SSC and IR in chloroplast genome of 11 species of Vitaceae

SSC 和 2 个反向重复区 (IRa 和 IRb) 4 部分组成, 各个区域间对应的连接基因相同。除了 *T. rafflesiae* 的 *rps19* 基因位于接近 JLB 边界的 LSC 区外, 其他 10 种植物浙江产三叶青、扁担藤、毛五叶崖爬藤、*T. lawsonii*、乌苡莓、祖提葡萄翁、矮葡萄瓮、*C. adenopodum*、桑德森葡萄翁的 *rps19* 基因横跨 JLB 边界; *rpl22* 基因位于接近 JLB 边界的 LSC 区; *rpl2* 基因位于接近 JLB 边界的 IRb 区和接近 JLA 边界的 IRa 区; *ycf1* 基因横跨 JSB 边界和 JSA 边界; 除了乌苡莓和 *C. adenopodum* 的 *ndhF* 基因横跨 JSB 边界外, 其他 9 种植物浙江产三叶青、扁担藤、毛五叶崖爬藤、*T. lawsonii*、*T. rafflesiae*、祖提葡萄翁、矮葡萄瓮、桑德森葡萄翁的 *ndhF* 基因位于接近 JSB 边界的 SSC 区; *trnH* 基因和 *psbA* 基因位于接近 JLA 边界的 LSC 区。由 IR 区边界扩张和收缩情况看, 葡萄科植物 *rps19*、*rpl2*、*ycf1* 和 *ndhF* 4 个基因位置和长度在属间差异明显, 基因位置和长度不太一致。怀玉山产三叶青和浙江产三叶青 *rps19* 基因和 *trnH* 基因一致, 但具有位置和长度的特异性, 与乌苡莓属和葡萄翁属组间差异明显, 说明 *rps19* 基因和 *ndhF* 基因不同程度的扩张和伸缩导致了不同物种间的 IR 和 SSC 区长度差异。

### 3.5 叶绿体基因组比对分析

以怀玉山产三叶青为参考, 使用 mVISTA 在线软件对 11 种葡萄科植物叶绿体基因组进行全序列比对, 以分析 11 种葡萄科植物叶绿体基因组序列的差异。结果表明, 崖爬藤属、乌苡莓属、葡萄翁属植物的基因组变化较大, 但崖爬藤属植物的基因组较为保守, 特别是怀玉山产三叶青和浙江产三叶青之间的基因组较为稳定 (图 3)。使用滑动窗口计算高度可变区域的核苷酸多样性, 以估计所分析 11 种葡萄科植物叶绿体基因组中不同区域的差异水平。结果表明, 核苷酸多样性的变化范围为 0~0.234 18, 基因间隔区 *trnS-GCU-trnG-UCC*、*rps16-trnG-UCC* 以及基因 *ycf1* 变异率最高 (图 4), 以上结果与 mVISTA 软件分析结果一致。

### 3.6 怀玉山产三叶青 RSCU 分析

怀玉山产三叶青氨基酸 RSCU 见表 3 和图 5。由表 3 和图 5 可知 RSCU 值大于 1.00 的密码子共有 30 个 (终止密码子除外), 其中, 有 29 个以 A 或 U 结尾, 有 1 个以 G 结尾, 说明怀玉山产三叶青叶绿体基因组的密码子偏好以 A 或 U 结尾。RSCU 值等于 1 的密码子有 1 个, 表明该密码子使用无偏好性。RSCU 值大于 1.60 的密码子为编码 Ala 的 GCU、编码 Arg 的 AGA、编码 Gly 的 GGA、编码 Leu 的 UUA、

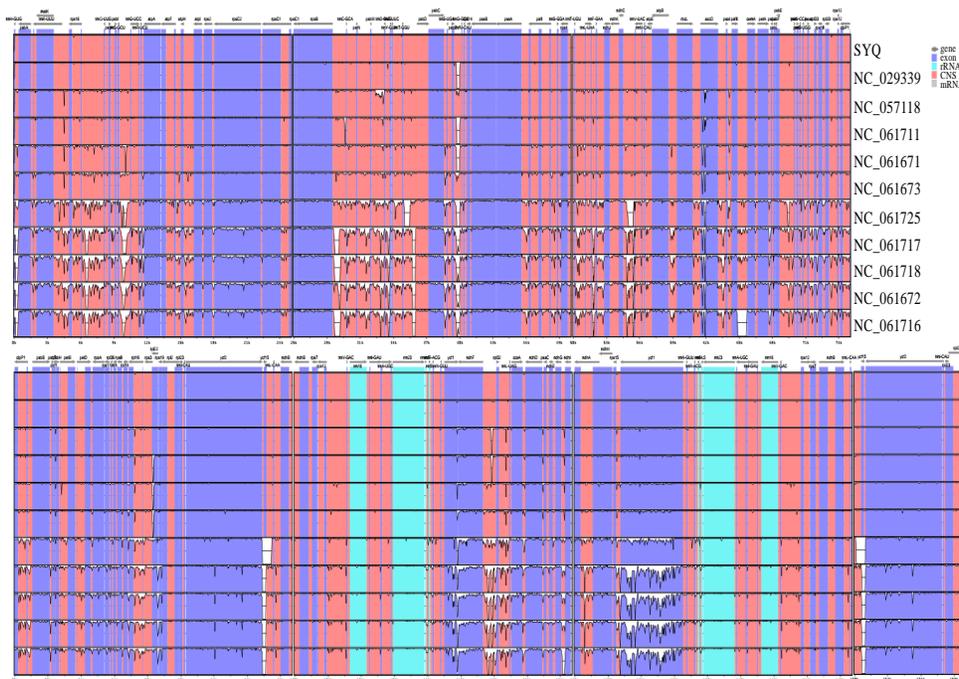


图 3 11 种葡萄科植物叶绿体基因组的可视化对比图

Fig. 3 Comparison of chloroplast genomes of 11 species of Vitaceae

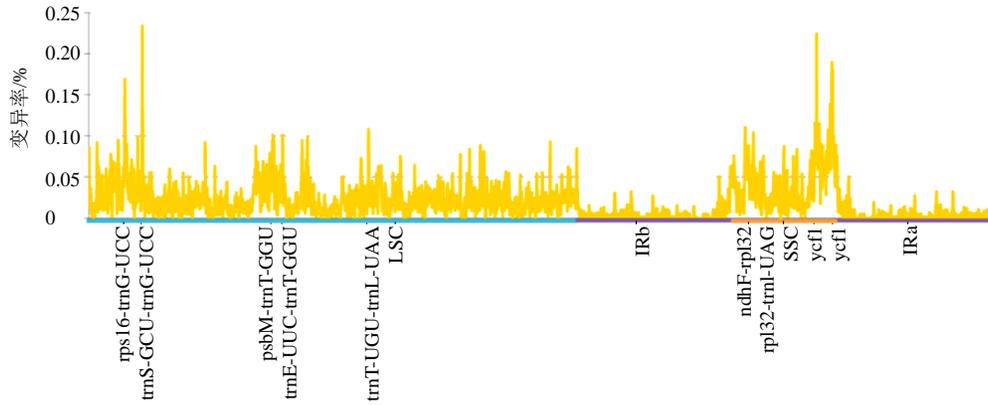


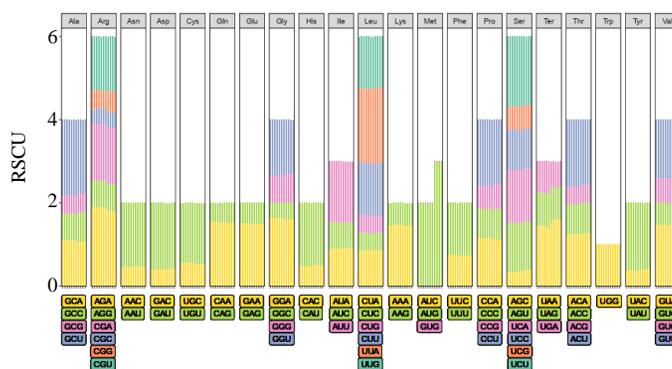
图4 11种葡萄科植物叶绿体基因组核苷酸多样性

Fig. 4 Nucleotide diversity of chloroplast genomes of 11 species of Vitaceae

表3 怀玉山产三叶青氨基酸的RSCU

Table 3 RSCU of amino acids of *T. hemsleyanum* from Huaiyu Mountain

密码子	氨基酸	RSCU	数量	密码子	氨基酸	RSCU	数量
GCA	Ala	1.101 780 0	387	UUA	Leu	1.791 490 0	842
GCC	Ala	0.643 416 0	226	UUG	Leu	1.238 300 0	582
GCG	Ala	0.438 434 0	154	AAA	Lys	1.467 320 0	1044
GCU	Ala	1.816 370 0	638	AAG	Lys	0.532 677 0	379
AGA	Arg	1.911 730 0	527	AUG	Met	1.993 800 0	643
AGG	Arg	0.634 825 0	175	GUG	Met	0.006 201 5	2
CGA	Arg	1.356 710 0	374	UUC	Phe	0.748 021 0	567
CGC	Arg	0.362 757 0	100	UUU	Phe	1.251 980 0	949
CGG	Arg	0.449 819 0	124	CCA	Pro	1.153 500 0	325
CGU	Arg	1.284 160 0	354	CCC	Pro	0.716 948 0	202
AAC	Asn	0.457 358 0	303	CCG	Pro	0.546 584 0	154
AAU	Asn	1.542 640 0	1022	CCU	Pro	1.582 960 0	446
GAC	Asp	0.403 950 0	225	AGC	Ser	0.353 444 0	124
GAU	Asp	1.596 050 0	889	AGU	Ser	1.162 950 0	408
UGC	Cys	0.577 640 0	93	UCA	Ser	1.282 660 0	450
UGU	Cys	1.422 360 0	229	UCC	Ser	0.963 420 0	338
CAA	Gln	1.543 270 0	740	UCG	Ser	0.544 418 0	191
CAG	Gln	0.456 726 0	219	UCU	Ser	1.693 110 0	594
GAA	Glu	1.509 540 0	1068	UAA	Ter	1.465 910 0	43
GAG	Glu	0.490 459 0	347	UAG	Ter	0.784 091 0	23
GGA	Gly	1.640 860 0	747	UGA	Ter	0.750 000 0	22
GGC	Gly	0.366 831 0	167	ACA	Thr	1.255 300 0	429
GGG	Gly	0.645 799 0	294	ACC	Thr	0.711 046 0	243
GGU	Gly	1.346 510 0	613	ACG	Thr	0.418 435 0	143
CAC	His	0.486 983 0	159	ACU	Thr	1.615 220 0	552
CAU	His	1.513 020 0	494	UGG	Trp	1.000 000 0	471
AUA	Ile	0.898 172 0	688	UAC	Tyr	0.378 270 0	188
AUC	Ile	0.629 243 0	482	UAU	Tyr	1.621 730 0	806
AUU	Ile	1.472 580 0	1128	GUA	Val	1.488 950 0	539
CUA	Leu	0.859 574 0	404	GUC	Val	0.527 624 0	191
CUC	Leu	0.425 532 0	200	GUG	Val	0.585 635 0	212
CUG	Leu	0.434 043 0	204	GUU	Val	1.397 790 0	506
CUU	Leu	1.251 060 0	588				



下面方块代表编码每种氨基酸的所有密码子，上方柱子的高度代表所有密码子 RSCU 值的总和

Box below represented all codons encoding each amino acid, and the height of the column above represented the sum of the RSCU values of all codons

图 5 RSCU 柱状图

Fig. 5 Relative synonymous codon usage (RSCU)

编码 Tyr 的 UAU、编码 Thr 的 ACU、编码 Ser 的 UCU、编码 Met 的 AUG。RSCU 值小于 0.60 的密码子包括编码 Ala 的 GCG、编码 Arg 的 CGC 和 CGG、编码 Asn 的 AAC、编码 Asp 的 GAC、编码 Cys 的 UGC、编码 Gln 的 CAG、编码 Glu 的 GAG、编码 Gly 的 GGC、编码 His 的 CAC、编码 Leu 的 CUC 和 CUG、编码 Lys 的 AAG、编码 Met 的 GUG、编码 Pro 的 CCG、编码 Ser 的 AGC 和 UCG、编码 Thr 的 ACG、编码 Tyr 的 UAC、编码 Val 的 GUC 和 GUG。综上所述，怀玉山产三叶青叶绿体基因组高频率使用 GCU、AGA、GGA、UUA、UAU、ACU、UCU、AUG 分别编码 Ala、Arg、Gly、Leu、Tyr、Thr、Ser 和 Met，低频率使用 GCG、CGC 和 CGG、AAC、GAC、UGC、CAG、GAG、GGC、CAC、CUC 和 CUG、AAG、GUG、CCG、AGC 和 UCG、ACG、UAC、GUC 和 GUG 分别编码 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Leu、Lys、Met、Pro、Ser、Thr、Tyr 和 Val。

### 3.7 密码子组成成分分析

利用 Codon W 和 CUSP 在线软件对 11 种葡萄科植物叶绿体基因组的 88 条 CDS 进行密码子组成成分分析 (图 6)。由图 6 可知，11 种葡萄科植物的 T<sub>3</sub> 和 A<sub>3</sub> 含量高于 G<sub>3</sub> 含量，表明 11 种葡萄科植物的第 3 位碱基以 A/T 为主。对 11 种葡萄科植物各基因密码子的 GC 含量进行统计分析，发现 GC<sub>3</sub> 含量 (29.18%) 均低于 GC<sub>1</sub> (45.90%) 和 GC<sub>2</sub> (39.87%)，表明 GC 在密码子不同位置上的分布并不均匀，GC<sub>1</sub>、GC<sub>2</sub>、GC<sub>3</sub> 呈现不断降低的趋势。大部分基因的 ENC 值都在 40 以上。ENC 的取值大小侧面反映了密码子偏性的强弱，值越高偏性越弱，怀玉山产

三叶青各基因的 ENC 值介于 26.167~61.000，平均值为 46.88，其中 ENC 值大于 45 的有 59 个，小于 45 的仅有 29 个，表明怀玉山产三叶青叶绿体基因组密码子具有较强的偏好性。

### 3.8 GC<sub>3</sub>-GC<sub>12</sub> 分析

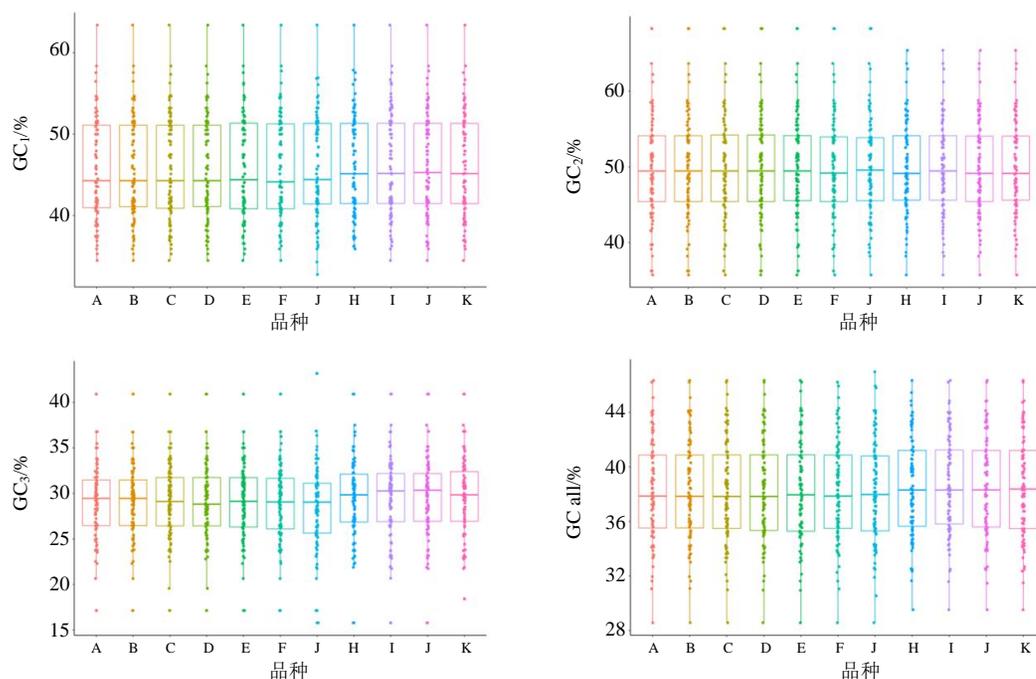
11 种葡萄科植物叶绿体基因组密码子 GC<sub>3</sub>-GC<sub>12</sub> 分析分析结果见图 7。由图 7 可知，GC<sub>3</sub>-GC<sub>12</sub> 分析发现 11 种葡萄科植物叶绿体基因的 GC<sub>3</sub> 含量分布在 0.15~0.50，GC<sub>12</sub> 含量分布在 0.30~0.60，二者的分布范围都比较集中，且沿着对角线上方分布。两者的相关系数  $r=0.01$  ( $R^2=0.0001$ )，相关显著，表明密码子偏好性主要受突变影响，受选择因素影响较小。回归曲线斜率为 0.0152，说明密码子第 1、2 位与第 3 位碱基组成偏好性差异较小，突变对密码子偏好性的影响较强，而其他因素，如自然选择可能对密码子的偏好性影响较小。

### 3.9 ENC-plot 分析

ENC-plot 分析主要用于判断密码子偏好性是否被突变或选择等因素影响，11 种葡萄科植物的 ENC-plot 分析见图 8。由图 8 可知，部分基因位点沿标准曲线分布或落在标准曲线附近，表明这些基因实际 ENC 值与预期值相近；另外一部分基因位点落在期望值的下方，说明实际 ENC 值与预期 ENC 值之间存在差异，说明 11 种葡萄科植物叶绿体基因组有一部分基因的密码子偏好性受到突变的影响较大，而受到选择效应的影响相对较小。

### 3.10 PR2-plot 分析

通过 PR2-plot 分析，11 种葡萄科植物叶绿体基因组中由 4 个密码子编码的氨基酸家族中第 3 位上嘌呤 (A、G) 和嘧啶 (T、C) 的使用情况见图 9。



GC<sub>12</sub>表示密码子第1、2位GC含量的平均值; GC<sub>3</sub>表示密码子第3位的GC含量 A~K 分别代表怀玉山产三叶青、浙江产三叶青、扁担藤、毛五叶崖爬藤、*Tetrastigma rafflesiae*、*Tetrastigma lawsonii*、乌菝莓、祖提葡萄翁、矮葡萄翁、*Cyphostemma adenopodum* 和桑德森葡萄翁, 下同

GC<sub>12</sub> means average GC value of the first and second position of these codons; GC<sub>3</sub> means the GC value of the third position of these codons; Letters A~K represent SYQ、*Tetrastigma hemsleyanum*, *Tetrastigma planicaule*, *Tetrastigma voinierianum*, *Tetrastigma rafflesiae*, *Tetrastigma lawsonii*, *Causonis japonica*, *Cyphostemma juttae*, *Cyphostemma humile*, *Cyphostemma adenopodum* and *Cyphostemma sandersonii* respectively, same as below

图6 11种葡萄科植物叶绿体基因组密码子组成成分分析

Fig. 6 Composition analysis of chloroplast genome codons of 11 species of Vitaceae

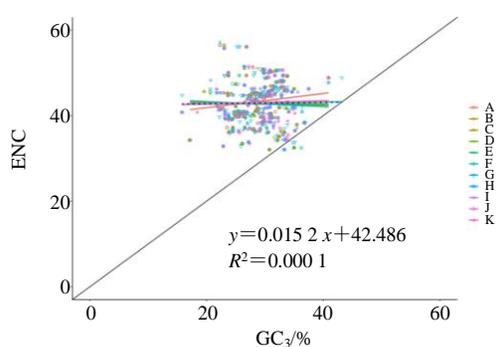


图7 11种葡萄科植物叶绿体基因组密码子 GC<sub>3</sub>-GC<sub>12</sub>分析  
Fig. 7 GC<sub>3</sub>-GC<sub>12</sub> analysis of chloroplast genome codons of 11 species of Vitaceae

由图9可知, 基因不均匀地分布于平面图的4个区域内, 平面图左上方基因分布较多, 说明密码子第3位碱基的使用存在偏性, 而且T的使用频率低于A, C的使用频率高于G。当密码子使用偏好只受突变影响时, 4种碱基的使用频率应相等。因此, 11种葡萄科植物叶绿体基因组密码子使用模式

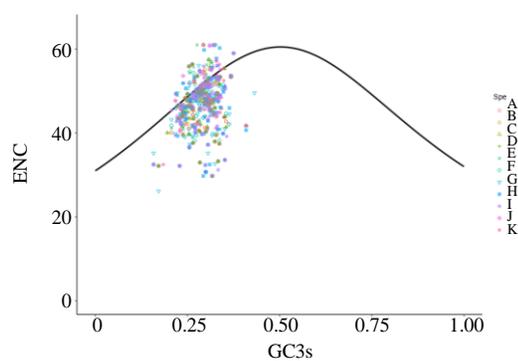


图8 11种葡萄科植物叶绿体基因组密码子 ENC-plot分析  
Fig. 8 ENC-plot analysis of chloroplast genome codons of 11 species of Vitaceae

不仅受到突变影响, 同时还受到如选择压力等其他因素的影响。

### 3.11 最优密码子确定

对怀玉山产三叶青叶绿体基因的 ENC 值进行排序, 分别选取两端基因各 5 个, 构建高表达基因库 (ENC 值小) 和低表达基因库 (ENC 值大), 并计算

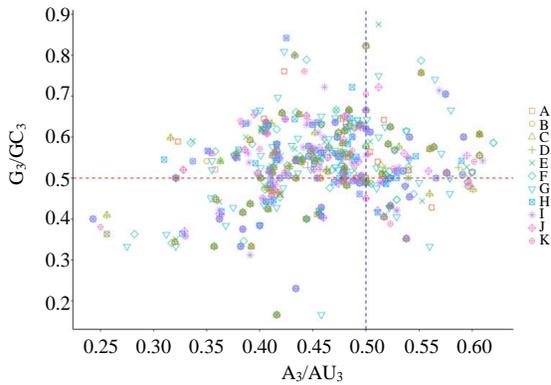


图9 11种葡萄科植物叶绿体基因组密码子PR2-plot分析  
Fig. 9 PR2-plot analysis of chloroplast genome codons of 11 species of Vitaceae

二者的RSCU差值( $\Delta$ RSCU)。筛选 $\Delta$ RSCU $\geq$ 0.08的高表达密码子,且RSCU>1的高频率密码子定义为怀玉山产三叶青叶绿体基因组的最优密码子(表4)。由表4可知,怀玉山产三叶青有31个高频密码子,16个以U结尾,13个以A结尾,2个以G结尾,其中有13个为最优密码子(6个以U结尾,7个以A结尾)。

### 3.12 葡萄科植物系统发育分析

基于6个崖爬藤属种类、1个乌葭莓属种类、4个葡萄翁属种类、1个牛果藤属种类和1个菱叶藤属种类的叶绿体基因组构建的系统发育树见图10。由图10可知,在崖爬藤属植物中,怀玉山产三叶青(SYQ)、浙江产三叶青(NC\_029339)、扁担藤、毛

表4 怀玉山产三叶青叶绿体基因最优密码子筛选

Table 4 Screening of optimal codons for chloroplast genes of *T. hemsleyanum* from Huaiyu Mountain

密码子	氨基酸	RSCU	数量	RSCU 低表达	数量	RSCU 高表达	数量	RSCU 差值
GCU	Ala	1.816 37	638			1.655 170	12	1.655 170
UUA	Leu	1.791 49	842	1.034 480	5	1.866 670	14	0.832 190
UCU	Ser	1.693 11	594	0.272 727	1	1.785 710	10	1.512 983
GGA	Gly	1.640 86	747	1.333 330	6	2.074 070	14	0.740 740
GAU	Asp	1.596 05	889	0.800 000	2	1.428 570	10	0.628 570
CCU	Pro	1.582 96	446	1.538 460	5	1.882 350	8	0.343 890
CAA	Gln	1.543 27	740	1.200 000	3	1.733 330	13	0.533 330
GUA	Val	1.488 95	539	1.142 860	4	1.454 550	12	0.311 690
UAA	Ter	1.465 91	43	0.600 000	1	1.200 000	3	0.600 000
CGU	Arg	1.284 16	354	0.833 333	5	1.315 790	5	0.482 457
UCA	Ser	1.282 66	450	1.636 360	6	1.964 290	11	0.327 930
UUU	Phe	1.251 98	949	1.230 770	8	1.680 000	21	0.449 230
CCA	Pro	1.153 50	325	0.615 385	2	0.705 882	3	0.090 497

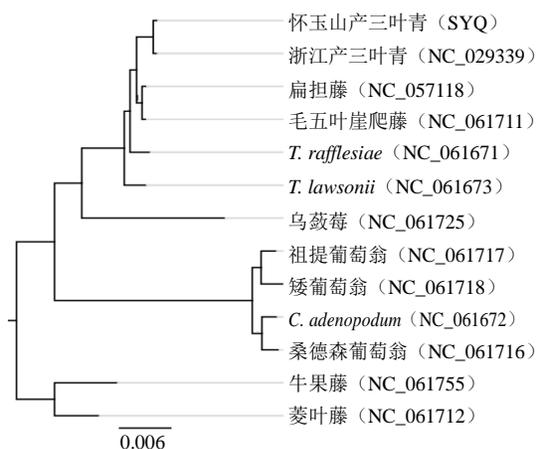


图10 基于叶绿体基因组的葡萄科13个种类的系统发育树  
Fig. 10 Phylogenetic tree of 13 species in Vitaceae based on chloroplast genome

五叶崖爬藤、*T. rafflesiae*、*T. lawsonii*和乌葭莓聚为一大支;祖提葡萄翁、矮葡萄翁、*Cyphostemma adenopodum*、桑德森葡萄翁聚为一支;牛果藤和菱叶藤聚为一大支,但怀玉山产三叶青与浙江产三叶青聚为一小支。说明聚在一支的崖爬藤属植物种类亲缘关系较近,崖爬藤属植物与乌葭莓属、葡萄翁属、牛果藤属和菱叶藤属植物亲缘关系较远。

基于8个三叶崖爬藤[怀玉山产三叶青(SYQ)、福建产三叶青(MW375708)、广西产三叶青(MW375709)、浙江产三叶青(MT827073)、浙江产三叶青(MW375707)、浙江产三叶青(KT033563)(NC\_029339)、四川三叶青(MW375710)、浙江产三叶青(MW375711)]、扁担藤、毛五叶崖爬藤、*T. rafflesiae*、*T. lawsonii*、*Tetrastigma sp.* Wen12461、牛果藤和菱叶藤的叶绿体基因组构建的系统发育树见图11。由图11可知,三叶青聚为2大类,其中一类包括

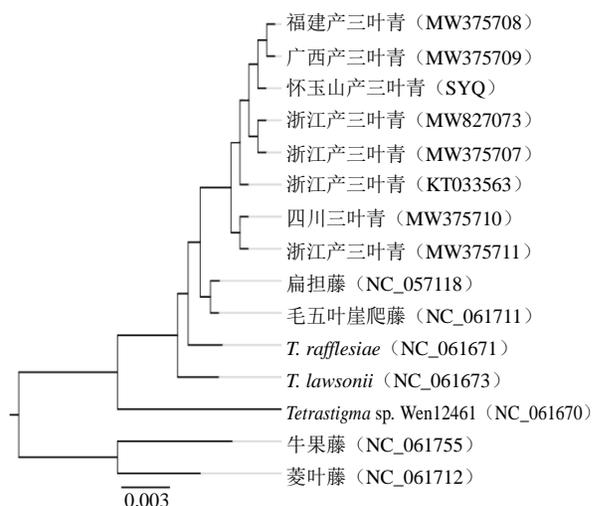


图 11 基于叶绿体基因组的葡萄科 15 个种类的系统发育树  
Fig. 11 Phylogenetic tree of 15 species in Vitaceae based on chloroplast genome

福建产三叶青 (MW375708)、广西产三叶青 (MW375709)、浙江产三叶青 (MW827073)、浙江产三叶青 (MW375707) 和浙江产三叶青 (KT033563) (NC\_029339), 另一类包括四川三叶青 (MW375710) 和浙江产三叶青 (MW375711)。而在第一类中, 浙江产三叶青 KT033563 (NC\_029339) 单独成一小支, 浙江产三叶青 (MW827073) 和浙江产三叶青 (MW375707) 单独成一小支, 怀玉山产三叶青 (SYQ)、福建产三叶青 MW375708 和广西产三叶青 (MW375709) 单独成一小支。说明怀玉山产三叶青 (SYQ)、福建产三叶青 (MW375708) 和广西产三叶青 (MW375709) 亲缘关系较近。

#### 4 讨论

叶绿体是高等植物光合作用和碳固定的重要细胞器, 起到能量转换的作用<sup>[13]</sup>。在叶绿体基因组中以若干小簇的形式排列的光合作用基因如 *psaA-psaB*、*psbD-psbC*、*psbE-psbF*、*atpB-atpE*、*atpI-atpH-atpF-atpA* 和 *psbB-psbH-petB-petD* 等属于转录单位和调控单位, 如叶绿体中的 *ndh* 基因都具转录活性, 可编码 NADH-质体醌 (PQ) 氧化还原酶, 当用 PQ 氧化抑制剂阻断光合系统 I 和 II 之间的电子传递通道时, 叶绿体呼吸的电子传递则不受影响; 叶绿体的 *mbpX* 和 *mbpY* 基因可编码膜上输送各种代谢物质的蛋白质; 约有一半的叶绿体基因包括 rRNA、tRNA、RNA 多聚酶和核糖体蛋白质等基因协同核基因参与到叶绿体蛋白质的合成; 叶绿体基因如 tRNA 和蛋白质基因

均含有内含子, 在表达过程中需要对转录后的 RNA 分子进行拼接加工<sup>[14]</sup>。

植物叶绿体基因组长约 107~218 kb, 其长短取决于 IR 区的收缩和扩张<sup>[15]</sup>。本研究所获得的怀玉山产三叶青叶绿体基因组大小与结构与已报道的被子植物研究结果相符。怀玉山产三叶青叶绿体基因组长 160 165 bp, 是一个典型的四分体结构, 与浙江产三叶青 (159 914 bp)<sup>[9-10]</sup>、浙江舟山三叶青 (152 281 bp) 差异小。怀玉山产三叶青叶绿体基因组 GC 含量为 37.50%, 与浙江舟山三叶青叶绿体基因组中 GC 含量 (37.5%) 一致, 与浙江产三叶青叶绿体基因组中 GC 含量 (37.55%) 有些差异, 怀玉山产三叶青叶绿体基因组 IRs 区中 GC 含量达到最高 (42.98%), 与浙江舟山三叶青 IRs 中 GC 含量 42.9% 有差异<sup>[9-10]</sup>。怀玉山产三叶青叶绿体基因组 LSC 区 88 172 bp, SSC 区 18 963 bp, IRs 区 26 515 bp, 与浙江产三叶青 (LSC: 87 927 bp、SSC: 18 967 bp、IRs: 26 510 bp)、浙江舟山三叶青 (LSC: 88 185 bp、SSC: 18 966 bp、IRs: 26 519 bp) 遗传差异很小, LSC 区域 10~250 bp, SSC 和 IRs 区域相差 1~10 bp<sup>[9-10]</sup>。

通过 IRscope 比较分析, 发现葡萄科的 11 个物种叶绿体基因组结构和大小差异较大, SC 和 IR 边界有明显区别, *rps19* 基因和 *ndhF* 基因不同程度的扩张和伸缩导致了 11 个物种间的 IR 和 SSC 区长度差异。反向重复区域经常出现假基因化, 在怀玉山产三叶青叶绿体基因组中, *ycf1* 基因有 2 个, 分别横跨 JSB 边界和 JSA 边界, 位于 JSA 边界的 *ycf1* 基因长度为 5633 bp (浙江舟山三叶青为 5634 bp), 为正常基因; 位于 JSB 边界的 *ycf1* 基因长度为 1199 bp (浙江舟山三叶青为 1200 bp), 位于 JS 边界的 *ycf1* 基因的 3' 端 SS 区发生缺失, 仅有 56 bp, 位于 JSB 边界的 *ycf1* 基因作为一个假基因<sup>[9]</sup>。有研究表明, 阴暗潮湿等环境因子的自然选择, *ycf1* 等基因选择压力存在差异, 导致 *ycf1* 等基因的进化速率也存在差异<sup>[16-17]</sup>, 这类进化速率存在差异的 *ycf1* 等基因, 可以组合成为 DNA 条形码用来分析崖爬藤属植物的系统发育<sup>[18]</sup>。

怀玉山产三叶青叶绿体基因组共注释出 133 个基因, 包括 88 个 CDS 基因、37 个 tRNA 基因和 8 个 rRNA 基因, 这与浙江舟山三叶青叶绿体基因组的基因注释结果<sup>[9]</sup>一致。怀玉山产三叶青

叶绿体基因组没有检测到内含子,仅个别基因具1~2个含子。其中 *clpP1* 基因具2个内含子,而浙江舟山三叶青叶绿体基因组 *ycf3* 和 *clpP1* 基因具2个内含子,且怀玉山产三叶青叶绿体基因组没有注释到 *ycf3* 基因。值得一提的是怀玉山产三叶青叶绿体基因组注释到6个未知功能基因 [*ycf1* (2)、*ycf15* (2)、*ycf2* (2)], 而浙江舟山三叶青叶绿体基因组未知功能基因有7个,注释到了1个 *ycf4* 基因<sup>[9]</sup>。*ycf3* 基因和 *ycf4* 基因可能是造成怀玉山产三叶青与浙江舟山三叶青品质差异的2个候选基因,其功能需进一步利用先进的分子技术手段对其功能进行鉴定。在本研究中,基于 VISTA tools 和 DNADnaSP6.0 软件分析结果表明,在11种葡萄科植物存在一些变异率较高的区段,如 *trnS-GCU-trnG-UCC*、*rps16-trnG-UCC* 以及基因 *ycf1* 等,今后这些变异热点区段,或许可以也作为新的 DNA 条形码,用于崖爬藤属植物的物种鉴定和系统发育关系等方面的研究<sup>[9]</sup>。

叶绿体 SSR 为短的单核苷酸串联重复序列,位于叶绿体基因组非编码区域时,通常显示种内变异重复次数<sup>[19]</sup>,在同一物种表现高度变异,广泛用于群体遗传学和系统发育分析等领域<sup>[20-21]</sup>。怀玉山产三叶青叶绿体基因组共检测到63个 SSR 位点,单碱基重复有56个,双碱基重复有7个,其中,单重复单元为 A 或 T 重复频率为10的 SSR 位点数量最多,重复单元为 A/T 的最高重复频率最高,这与其他植物科<sup>[22]</sup>中的报道一致,怀玉山产三叶青叶绿体基因组 SSR 位点可为崖爬藤属物种分子标记开发及物种鉴定提供理论依据。

核酸组成成分显著影响核基因密码子使用偏好性,自然选择和变异方向在叶绿体基因组密码子使用偏好性中起到重要作用<sup>[23]</sup>。在本研究中,怀玉山产三叶青叶绿体基因组 RSCU 值大于1.00的密码子共有30个(终止密码子除外),其中有29个以 A 或 U 结尾,有1个以 G 结尾,这一结果与大多数被子植物叶绿体基因组的研究结果类似,都偏好使用 A 或 U 结尾的密码子<sup>[24-25]</sup>。以叶绿体基因组为对象,RSCU 值 >1.60 时,为高使用频率密码子,RSCU 值 <0.60 时为低使用频率密码子<sup>[26]</sup>,在本研究中,怀玉山产三叶青叶绿体基因组高频率使用 GCU、AGA、GGA、UUA、UAU、ACU、UCU、AUG 分别编码 Ala、Arg、Gly、Leu、Tyr、Thr、Ser 和 Met,低频率使用

GCG、CGC 和 CGG、AAC、GAC、UGC、CAG、GAG、GGC、CAC、CUC 和 CUG、AAG、GUG、CCG、AGC 和 UCG、ACG、UAC、GUC 和 GUG 分别编码 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Leu、Lys、Met、Pro、Ser、Thr、Tyr 和 Val。假设核苷酸碱基变异发生在密码子第3位,那么在基因或基因组中简并密码子 A、T 和 C、G 所占比例相当<sup>[27]</sup>。本研究通过 PR2-plot 分析发现,基因不均匀地分布于平面图的4个区域内,平面图左上方基因分布较多,密码子第3位碱基的使用存在偏性,T 使用频率低于 A,C 使用频率高于 G,说明怀玉山产三叶青叶绿体基因组受碱基突变和自然选择的双重影响。本研究 GC3-GC12 分析表明,密码子第1、2位与第3位碱基组成偏好性差异较小,突变对密码子偏好性的影响较强,而其他因素如自然选择可能对密码子的偏好性影响较小。有研究表明,低等生物体中如细菌和古生菌,富含 GC 的物种通常拥有 GC 含量较高的最优密码子<sup>[28]</sup>;反之,富含 AU 的物种,最优密码子 AU 含量较高<sup>[29]</sup>。本研究结果与此一致,怀玉山产三叶青有31个高频密码子,16个以 U 结尾,13个以 A 结尾,2个以 G 结尾,其中有13个为最优密码子(6个以 U 结尾,7个以 A 结尾),最优密码子多以 A 或 U 碱基结尾,可为叶绿体是通过内共生方式从蓝细菌进化而来的观点提供依据<sup>[30]</sup>。

叶绿体基因组信息量较大,且其结构、大小和基因组成相对恒定,已广泛应用于系统进化分析。蒋明等<sup>[9]</sup>和 Li 等<sup>[10]</sup>分别构建了三叶崖爬藤、赤苍藤、葡萄、夏葡萄、东北蛇葡萄、槭叶葡萄等12种植物和三叶崖爬藤、葡萄等40种植物叶绿体基因组的系统发生树,均发现崖爬藤属的三叶崖爬藤单独聚为一组,与葡萄属较为接近。本研究结果与此研究结果一致,本研究构建了6个崖爬藤属种类、1个乌莓属种类、4个葡萄翁属种类、1个牛果藤属种类和1个菱叶藤属种类叶绿体基因组的系统发育树,发现怀玉山产三叶青与浙江产三叶青聚为一小支,说明聚在一支的崖爬藤属植物种类亲缘关系较近,崖爬藤属植物与乌莓属、葡萄翁属、牛果藤属和菱叶藤属植物亲缘关系较远。本研究还构建了怀玉山产三叶青、福建产三叶青、广西产三叶青、浙江产三叶青、四川三叶青叶绿体基因组构建的系统发育树,发

现怀玉山产三叶青、福建产三叶青和广西产三叶青聚为一小支,说明怀玉山产三叶青与福建产三叶青和广西产三叶青亲缘关系较近,与浙江产三叶青、四川三叶青和浙江产三叶青亲缘关系较远,说明三叶青之间的亲缘关系与地理位置相关性不大,这也证实了尹明华等<sup>[31]</sup>利用 SSR 荧光标记分析三叶青 64 个种质遗传多样性和亲缘关系的结论。本研究获得了怀玉山产三叶青叶绿体基因组的基本信息及其系统发育关系,为怀玉山产三叶青的遗传多样性保护、生产性状、系统演化等相关研究提供数据基础,也为怀玉山产三叶青种质鉴定、DNA 条形码开发提供了新思路。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] Zhan L H, Pu J B, Zheng J R, *et al.* *Tetrastigma hemsleyanum* Diels et Gilg ameliorates lipopolysaccharide induced sepsis via repairing the intestinal mucosal barrier [J]. *Biomed Pharmacother*, 2022, 148: 112741.
- [2] Yang L, Shi Y S, Ruan X, *et al.* Salt interferences to metabolite accumulation, flavonoid biosynthesis and photosynthetic activity in *Tetrastigma hemsleyanum* [J]. *Environ Exp Bot*, 2022, 194: 104765.
- [3] Liu Y Y, Li Y L, Chen W, *et al.* *Tetrastigma hemsleyanum* flavones exert antihepatic carcinoma property both in vitro and in vivo [J]. *Food Qual Saf*, 2021, 5: 9-16.
- [4] Bobik K, Burch-Smith T M. Chloroplast signaling within, between and beyond cells [J]. *Front Plant Sci*, 2015, 6: 781.
- [5] Shahzadi I, Abdullah, Mehmood F, *et al.* Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: Comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae [J]. *Genomics*, 2020, 112(2): 1454-1463.
- [6] Liu Y C, Li Y T, Feng S X, *et al.* Complete chloroplast genome structure of four *Ulmus* species and *Hemiptelea davidii* and comparative analysis within Ulmaceae species [J]. *Sci Rep*, 2022, 12(1): 15953.
- [7] Meng J, Li X P, Li H T, *et al.* Comparative analysis of the complete chloroplast genomes of four *Aconitum* medicinal species [J]. *Molecules*, 2018, 23(5): 1015.
- [8] 孙健, 沈晓霞, 陈加红, 等. 药用植物三叶青种质多样性与栽培管理的研究进展 [J]. *科技通报*, 2018, 34(1): 13-17.
- [9] 蒋明, 王军峰, 应梦豪, 等. 三叶崖爬藤叶绿体基因组的组装与序列分析 [J]. *中草药*, 2020, 51(2): 461-468.
- [10] Li M Z, Chen Q Y, Yang B X, *et al.* The complete chloroplast genome sequence of *Tetrastigma hemsleyanum* Diels at Gilg [J]. *Mitochondrial DNA A DNA Mapp Seq Anal*, 2016, 27(5): 3729-3730.
- [11] 尹明华, 谢妮妮, 徐文慧, 等. 三叶青种质资源遗传多样性的 ISSR 分析 [J]. *中草药*, 2018, 49(20): 4884-4891.
- [12] 林国卫, 闻静, 石光禹, 等. 侵染怀玉山产三叶青的病毒 RT-PCR 鉴定 [J]. *分子植物育种*, 2020, 18(3): 968-975.
- [13] Neuhaus H E, Emes M J. Nonphotosynthetic metabolism in plastids [J]. *Annu Rev Plant Physiol Plant Mol Biol*, 2000, 51: 111-140.
- [14] Grevich J J, Daniell H. Chloroplast genetic engineering: Recent advances and future perspectives [J]. *Crit Rev Plant Sci*, 2005, 24(2): 83-107.
- [15] Yu J Y, Xia M Z, Xu H, *et al.* The complete chloroplast genome sequence of *Neopallasia pectinata* (Asteraceae) [J]. *Mitochondrial DNA B Resour*, 2021, 6(2): 430-431.
- [16] Yang X F, Wang Y T, Chen S T, *et al.* PBR1 selectively controls biogenesis of photosynthetic complexes by modulating translation of the large chloroplast gene *Ycf1* in *Arabidopsis* [J]. *Cell Discov*, 2016, 2: 16003.
- [17] Vitti J J, Grossman S R, Sabeti P C. Detecting natural selection in genomic data [J]. *Annu Rev Genet*, 2013, 47: 97-120.
- [18] 裴男才. 利用植物 DNA 条形码构建亚热带森林群落系统发育关系: 以鼎湖山样地为例 [J]. *植物分类与资源学报*, 2012, 34(3): 263-270.
- [19] Provan J, Powell W, Hollingsworth P M. Chloroplast microsatellites: New tools for studies in plant ecology and evolution [J]. *Trends Ecol Evol*, 2001, 16(3): 142-147.
- [20] Wills D M, Hester M L, Liu A Z, *et al.* Chloroplast SSR polymorphisms in the Compositae and the mode of organellar inheritance in *Helianthus annuus* [J]. *Theor Appl Genet*, 2005, 110(5): 941-947.
- [21] Ebrahimi A, Zarei A, Zamani F M, *et al.* Evaluation of genetic variability among early mature *Juglans regia* using microsatellite markers and morphological traits [J]. *Peer J*, 2017, 5: e3834.
- [22] 孙哲, 李澳旋, 杜晓蓉, 等. 房山紫堇叶绿体基因组特征及其系统进化关系 [J]. *草地学报*, 2022, 30(8): 1982-1989.
- [23] Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding [J]. *Nat Struct Mol Biol*, 2013,

- 20(2): 237-243.
- [24] 杨亚蒙, 焦健, 樊秀彩, 等. 桑叶葡萄叶绿体基因组及其特征分析 [J]. 园艺学报, 2019, 46(4): 635-648.
- [25] Zhou M, Long W, Li X. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants [J]. *For Stud China*, 2008, 10(4): 235-242.
- [26] Mukhopadhyay P, Basak S, Ghosh T C. Nature of selective constraints on synonymous codon usage of rice differs in GC-poor and GC-rich genes [J]. *Gene*, 2007, 400(1/2): 71-81.
- [27] Liu Q P, Xue Q Z. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species [J]. *J Genet*, 2005, 84(1): 55-62.
- [28] Nekrutenko A, Li W H. Assessment of compositional heterogeneity within and between eukaryotic genomes [J]. *Genome Res*, 2000, 10(12): 1986-1995.
- [29] Hershberg R, Petrov D A. General rules for optimal codon choice [J]. *PLoS Genet*, 2009, 5(7): e1000556.
- [30] Raven J A, Allen J F. Genomics and chloroplast evolution: What did cyanobacteria do for plants? [J]. *Genome Biol*, 2003, 4(3): 209.
- [31] 尹明华, 徐文慧, 谢妮妮, 等. 三叶青种质资源遗传多样性的 SSR 荧光标记分析 [J]. 中草药, 2018, 49(23): 5649-5656.

[责任编辑 时圣明]