

## ‘怀玉山’高山马铃薯叶绿体基因组特征及密码子使用偏好性分析

洪森荣<sup>1,2,3,4</sup>, 张牧彤<sup>1</sup>, 徐子林<sup>1</sup>, 张钦荣<sup>1</sup>, 罗雨欣<sup>1</sup>, 田文慧<sup>1</sup>, 王心雨<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. 上饶师范学院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. 上饶师范学院上饶市薯芋类作物种质保存与利用重点实验室, 江西 上饶 334001)

**摘要:** 【目的】分析‘怀玉山’高山马铃薯 *Solanum tuberosum* var. *cormosus* ‘Huaiyushan’ 叶绿体基因组特征及密码子使用偏好性, 为开展‘怀玉山’高山马铃薯叶绿体基因组密码子优化、叶绿体基因组改造, 探索物种进化和增加外源基因表达等研究提供参考依据和理论基础。【方法】采用高通量测序技术对‘怀玉山’高山马铃薯叶绿体基因组进行测序, 并利用生物信息学分析软件对组装和注释后的叶绿体基因组进行结构、基因组成及密码子偏好性分析。【结果】‘怀玉山’高山马铃薯叶绿体基因组大小为 155 296 bp, 为经典的 4 段式结构。大单拷贝区 (LSC)、小单拷贝区 (SSC) 和反向重复区 (IR) 长度分别为 85 737、18 373、25 593 bp, 总鸟嘌呤和胞嘧啶所占的比例 (GC 比例) 为 37.88%, 共注释出 133 个基因, 包含 87 个编码区 (CDS) 基因、37 个 tRNA 基因、8 个 rRNA 基因和 1 个假基因。‘怀玉山’高山马铃薯叶绿体基因组中共检测到 38 个简单重复序列位点 (SSR 位点, 36 个单碱基重复和 2 个双碱基重复) 和 32 个长重复序列 (16 个正向重复和 16 个回文重复)。“怀玉山”高山马铃薯叶绿体基因组核苷酸多样性为 0~0.139 27, 高变区主要分布在大单拷贝区和小单拷贝区, 大单拷贝区 *trnL-UAA-trnF-GAA*、*cemA*、*rps12-exon1-clpP1*、*clpP1* 基因变异率最高, 小单拷贝区 *rpl32-trnL-UAG*、*yef1* 基因变异率最高。‘怀玉山’高山马铃薯叶绿体基因组 87 个 CDS 基因的平均有效密码子数 (ENC) 为 47.29, ENC > 45 的基因有 60 个, 密码子偏性较弱。‘怀玉山’高山马铃薯叶绿体基因组密码子偏好以 A、U 结尾, 使用偏性很大程度上受自然选择的影响, 而受突变压力的影响小。CGU、AAA、CUU、GUU、GGA、GUA、GGU、UCA、GCU、CCU 为‘怀玉山’高山马铃薯叶绿体基因组的 10 个最优密码子。【结论】‘怀玉山’高山马铃薯与马铃薯栽培种 *S. tuberosum* ‘Desiree’ 亲缘关系较近。图 5 表 3 参 41

**关键词:** ‘怀玉山’高山马铃薯; 叶绿体基因组; 序列特征; 密码子偏好性; 最优密码子; 系统发育分析

**中图分类号:** S532 **文献标志码:** A **文章编号:** 2095-0756(2024)01-0092-12

## Chloroplast genome characteristics and codon usage preference of *Solanum tuberosum* var. *cormosus* ‘Huaiyushan’

HONG Senrong<sup>1,2,3,4</sup>, ZHANG Mutong<sup>1</sup>, XU Zilin<sup>1</sup>, ZHANG Qinrong<sup>1</sup>,  
LUO Yuxin<sup>1</sup>, TIAN Wenhui<sup>1</sup>, WANG Xinyu<sup>1</sup>

(1. College of Life Sciences, Shangrao Normal University, Shangrao 334001, Jiangxi, China; 2. Shangrao Agricultural Technology Innovation Research Institute, Shangrao 334001, Jiangxi, China; 3. Key Laboratory of Protection and Utilization of Medicinal and Edible Plant Resources in Shangrao City, Shangrao Normal University, Shangrao 334001, Jiangxi, China; 4. Key Laboratory of Germplasm Conservation and Utilization of Potato and Taro Crops in Shangrao City, Shangrao Normal University, Shangrao 334001, Jiangxi, China)

收稿日期: 2023-02-20; 修回日期: 2023-10-30

基金项目: 国家自然科学基金资助项目 (32060092, 31960079); 2022 年上饶市科技专项项目 (2022A008); 江西省科技厅重点研发计划一般项目 (20192BBF60006, 20202BBF63001); 江西省教育厅科学技术研究项目 (GJJ201704, GJJ211729); 江西省现代农业产业技术体系建设专项 (JXARS-13-赣东站); 上饶市科技局平台载体建设项目 (2020I001, 2020J001)

作者简介: 洪森荣 (ORCID: 0000-0002-9219-8303), 教授, 从事植物生物技术研究。E-mail: hongsenrong@163.com

**Abstract:** [Objective] This study aims to analyze the chloroplast genome characteristics and codon usage preference of *Solanum tuberosum* var. *cormosus* ‘Huaiyushan’, so as to provide a reference and theoretical basis for future research on chloroplast genome codon optimization, chloroplast genome modification, exploration of species evolution, and increased expression of exogenous genes. [Method] The chloroplast genome of *S. tuberosum* var. *cormosus* ‘Huaiyushan’ was sequenced by high-throughput sequencing technology. The structure, gene composition and codon preference of the assembled and annotated chloroplast genome were analyzed using bioinformatics analysis software. [Result] The chloroplast genome size of *S. tuberosum* var. *cormosus* ‘Huaiyushan’ was 155 296 bp, which was a classical 4-segment structure. The length of large single copy region (LSC), small single copy region (SSC) and inverted repeat region (IR) was 85 737, 18 373 and 25 593 bp respectively, and the proportion of total guanine and cytosine (GC ratio) was 37.88%. A total of 133 genes were released, including 87 CDS genes, 37 tRNA genes, 8 rRNA genes and 1 pseudogene. A total of 38 simple sequence repeat (SSR) loci (36 single base repeats and 2 double base repeats) and 32 long repeat sequences (16 forward repeats and 16 palindrome repeats) were detected in the chloroplast genome of *S. tuberosum* var. *cormosus* ‘Huaiyushan’. The variation range of nucleotide diversity of chloroplast genome of *S. tuberosum* var. *cormosus* ‘Huaiyushan’ was 0 – 0.139 27. The hypervariable regions were mainly distributed in LSC and SSC regions. LSC region had the highest mutation rates of *trnL-UAA-trnF-GAA*, *cema*, *rps12-exon1-clpP1* and *clpP1* genes, while SSC region had the highest mutation rates of *rpl32-trnL-UAG* and *ycf1* genes. The average effective number of codon (ENC) value of 87 CDS genes in the chloroplast genome of *S. tuberosum* var. *cormosus* ‘Huaiyushan’ was 47.29, and there were 60 genes with ENC value larger than 45, indicating weak codon preference. The preference for codons ending with A and U in the chloroplast genome of *S. tuberosum* var. *cormosus* ‘Huaiyushan’ was largely affected by natural selection, and less affected by mutation pressure. CGU, AAA, CUU, GUU, GGA, GUA, GGU, UCA, GCU, and CCU were the 10 best codons in the chloroplast genome of *S. tuberosum* var. *cormosus* ‘Huaiyushan’. [Conclusion] *S. tuberosum* var. *cormosus* ‘Huaiyushan’ has a close genetic relationship with *S. tuberosum* ‘Desiree’. [Ch, 5 fig. 3 tab. 41 ref.]

**Key words:** *Solanum tuberosum* var. *cormosus* ‘Huaiyushan’; chloroplast genome; sequence characteristics; codon preference; optimal codons; phylogenetic analysis

‘怀玉山’高山马铃薯 *Solanum tuberosum* var. *cormosus* ‘Huaiyushan’, 又名麻籽洋芋, 茄科 Solanaceae 茄属 *Solanum* 1 年生草本植物, 主要种植区域为江西省玉山县怀玉乡<sup>[1]</sup>。‘怀玉山’高山马铃薯食用、药用皆优, 获批为国家地理标志农产品<sup>[2-3]</sup>。已有研究表明: ‘怀玉山’高山马铃薯与云南德宏和曲靖以及湖北恩施的高山马铃薯种质存在差异<sup>[4]</sup>。但‘怀玉山’高山马铃薯的进化来源尚无相关研究报道。

叶绿体是高等植物细胞内一种重要的与光合作用和物质代谢相关的细胞器, 叶绿体基因组是一套具有母系遗传特征的独立基因组, 是高等植物细胞质基因组的组成成分之一<sup>[5]</sup>。与核基因组相比, 叶绿体基因组全长序列短、易测序获得、基因直系同源、基因结构稳定、保守性较高、进化速率适中, 目前已经广泛应用于植物系统发育分析、物种分类鉴定及分子标记开发等研究中, 在物种起源、进化、演变及比较基因组学等研究领域发挥着越来越大的作用<sup>[6]</sup>。密码子是核酸和蛋白质之间遗传信息传递的桥梁<sup>[7]</sup>, mRNA 上的遗传信息以 tRNA 三重密码子传递。氨基酸一般对应  $\geq 1$  的密码子<sup>[8]</sup>, 这些密码子称为同义密码子<sup>[9]</sup>。在自然选择或突变偏好的情况下, 基因倾向于使用  $\geq 1$  的同义密码子, 即同义密码子使用偏好性<sup>[10-12]</sup>。目前, 关于茄属的叶绿体基因组研究已有报道<sup>[13-19]</sup>, 而针对‘怀玉山’高山马铃薯的研究大多集中在基因克隆<sup>[20]</sup>、转录组分析<sup>[3]</sup>、遗传多样性<sup>[4]</sup>、脱毒快繁<sup>[2]</sup>、DNA 甲基化敏感扩增多态性 (MSAP) 分析<sup>[21]</sup> 等方面, 对‘怀玉山’高山马铃薯叶绿体全基因组及其密码子使用偏好性方面的研究还未见系统报道。本研究通过对‘怀玉山’高山马铃薯叶绿体基因组进行测序和组装, 分析基于叶绿体基

因组的‘怀玉山’高山马铃薯系统进化、结构解析和密码子偏好性等,为‘怀玉山’高山马铃薯叶绿体基因组研究和应用提供科学依据,也为进一步研究‘怀玉山’高山马铃薯遗传背景、种质资源保护与开发利用奠定基础。

## 1 材料与方法

### 1.1 材料

由上饶市薯芋类作物种质保存与利用重点实验室提供的‘怀玉山’高山马铃薯试管苗。

### 1.2 方法

1.2.1 DNA提取和测序 选取‘怀玉山’高山马铃薯(MLS)试管苗叶片组织,利用植物基因组DNA提取试剂盒(北京天根生化科技有限公司)提取‘怀玉山’高山马铃薯试管苗DNA,质量分数为1%琼脂糖凝胶电泳检测DNA的完整性,NanoDrop 2000分光光度计(Thermo Scientific公司)检测DNA浓度和纯度,用超声波将DNA片段化,然后对片段化的DNA进行片段纯化、末端修复、3'端加A、连接测序接头,再用琼脂糖凝胶电泳进行片段大小选择,进行聚合酶链式反应(PCR)扩增形成测序文库。建好的文库先进行文库质检,质检合格的文库用BGISEQ-500平台进行测序。

1.2.2 叶绿体全基因组的组装与注释 通过SOAPnuk 1.3.0对raw data(测序下机的原始数据)进行数据过滤,去除其中的接头序列及低质量reads(高通量测序中一个反应获得的测序序列),获得高质量的clean data(对原始数据进行过滤后并剔除了低质量数据的剩余数据)。采用Noveplastys软件组装叶绿体基因组核心模块,以起始组装序列为起点开始组装叶绿体contigs(很多reads根据序列拼接在一起拼出的片段),如果contigs未环化,则利用CAP3软件连接多个contigs为完整叶绿体基因组,并手动调整环状叶绿体基因组起始位置。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基因注释结果。将注释完成的‘怀玉山’高山马铃薯叶绿体基因组序列提交至美国国家生物信息中心(NCBI),获得登录号:OP589401。使用OGDRAW绘制叶绿体基因组图谱。

1.2.3 叶绿体基因组特征分析 通过JSHYCloud在线工具集分析并统计叶绿体基因组、大单拷贝区(LSC)、小单拷贝区(SSC)和反向重复区(IR)的鸟嘌呤和胞嘧啶所占的比例(GC比例);使用MISA软件进行简单重复序列(SSR)分析,单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸的最小重复值分别设置为10、6、5、5、5、5;利用REPuter软件进行长重复序列(longrepeat)分析,查找正向重复(F)、反向重复(R)、互补重复(C)、回文重复(P)等4种重复类型;通过Pastor Galaxy在线工具集中的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频率(RSCU),其他参数设为默认值。将‘怀玉山’高山马铃薯叶绿体基因组序列上传至美国国家生物技术信息中心(NCBI)进行BLASTn比对,选择highly similar sequence(megablast)比较相似性在95%以上的序列,检索获得‘怀玉山’高山马铃薯的近缘种。利用Gview、VISTA tools、IRscope和DNADnaSP 6.0软件绘制‘怀玉山’高山马铃薯及其10个近缘种(*S. cochoae* NC\_062512、多毛番茄*S. habrochaites* NC\_026879、潘那利番茄*S. pennellii* NC\_035742、*S. bukasovii* MT120867、*S. boliviense* NC\_062870、*S. trisectum* NC\_062469、*S. salamancae* NC\_062480、*S. clivorum* NC\_062513、*S. mortonii* NC\_062426、*S. insanum* MW384851)的变异圈图、mVIST结构变异图、IR结构变异图,计算‘怀玉山’高山马铃薯及其10个近缘种的基因组核酸多样性(Pi),参数设置100 bp滑窗,25 bp的步长,并进行中性绘图分析(GC3-GC12分析)、ENC-plot分析、PR2-bias-plot分析和最优密码子分析;对‘怀玉山’高山马铃薯叶绿体基因的ENC进行排序,分别选取两端基因各5个,构建高表达基因库(ENC小)和低表达基因库(ENC大),并计算两者的RSCU差值( $\Delta$ RSCU)。筛选 $\Delta$ RSCU $\geq$ 0.08的高表达密码子,且将RSCU $>$ 1.00的高频率密码子定义为‘怀玉山’高山马铃薯叶绿体基因组的最优密码子;最后利用mafft 7.0和fasttree 2.1.10软件分别对‘怀玉山’高山马铃薯和18个近缘种以及烟草属*Nicotiana* 2个外类群物种进行序列比对和构建进化树。

## 2 结果与分析

### 2.1 叶绿体基因组序列特征分析

经过测序组装的完整的叶绿体基因组长度为155 296 bp,图1显示:‘怀玉山’高山马铃薯叶绿体

基因组呈典型的四分体结构, 包含 1 个 LSC、1 个 SSC 和 2 个将 LSC 与 SSC 分隔开的 IR (IRa 和 IRb)。基因组的总 GC 比例为 37.88%, A、T、C、G 比例分别为 30.65%、31.47%、19.24%、18.65%。LSC、SSC 和 IR 的长度分别为 85 737、18 373、25 593 bp。LSC 的 GC 比例为 36.01%, A、T、C、G 比例分别为 31.29%、32.70%、18.40%、17.61%; SSC 的 GC 比例为 32.09%, A、T、C、G 比例分别为 33.78%、34.14%、16.69%、15.40%; IRb 的 GC 比例为 43.10%, A、T、C、G 比例分别为 28.57%、28.33%、20.72%、22.39%; IRa 的 GC 比例为 43.10%, A、T、C、G 比例分别为 28.33%、28.57%、22.39%、20.72%。表明 ‘怀玉山’高山马铃薯 IR 的 GC 比例最大, LSC 次之, SSC 最少; 叶绿体基因组总 GC 比例显著低于 AT 比例; 叶绿体基因组各碱基比例从大到小依次为 T、A、C、G。

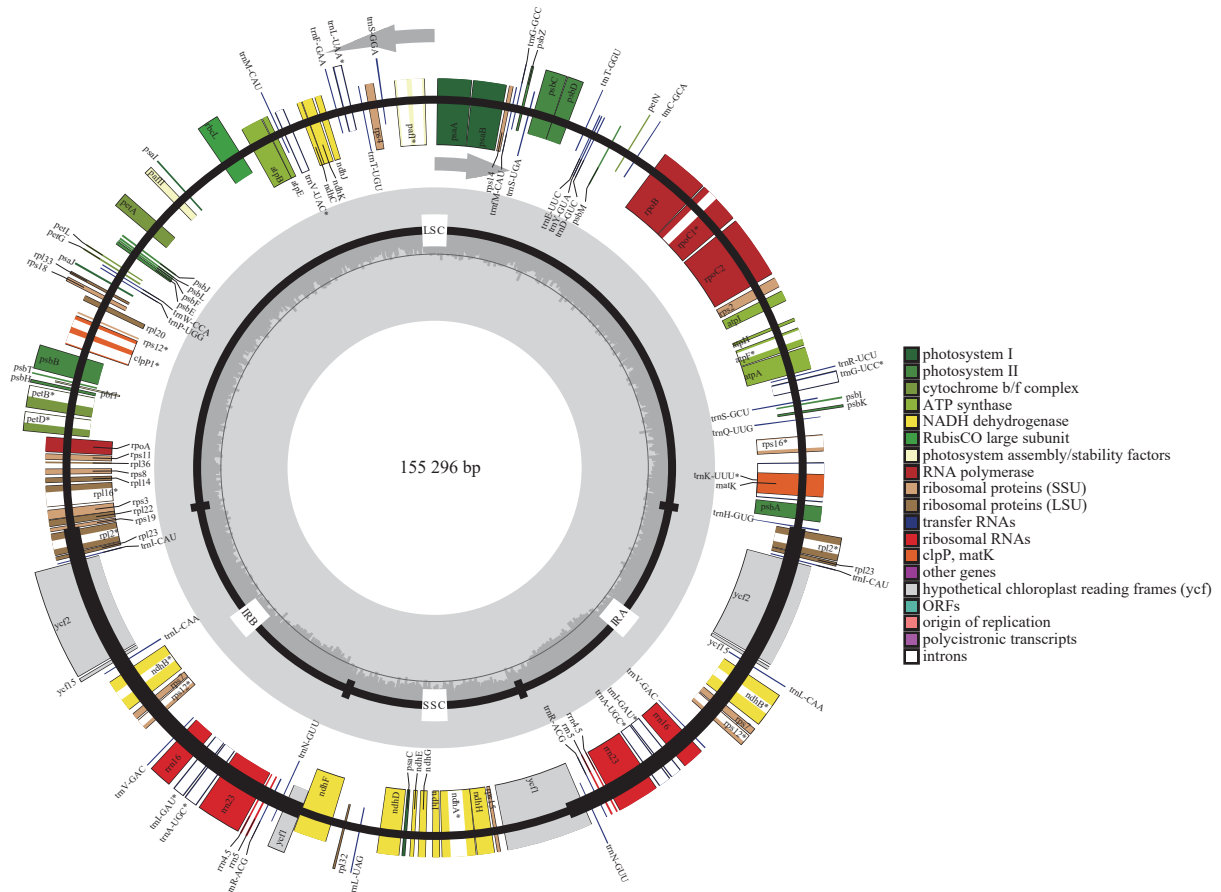


图 1 ‘怀玉山’高山马铃薯叶绿体基因组图谱  
Figure 1 Chloroplast genome map of *S. tuberosum* var. *cormosum* ‘Huaiyushan’

## 2.2 叶绿体基因类型分析

叶绿体基因组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 87 个编码区 (CDS) 基因、37 个 tRNA 基因、8 个 rRNA 基因、1 个假基因, 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 21 个, 包括 13 个 CDS 基因和 8 个 tRNA 基因; 由 3 个外显子构成的基因有 4 个, 为 *clpP1*、*ycf3*、*rps12* (2 个) 基因。LSC 的基因数量最多 (81 个), 其中 CDS 基因 59 个、tRNA 基因 22 个; SSC 的基因数量为 11 个, 其中 CDS 基因 10 个、tRNA 基因 1 个; IR 的基因数量为 17 个, 其中 CDS 基因 6 个、rRNA 基因 4 个、tRNA 基因 7 个; SSC 与 IRb 边界 (JSB) 的基因数量为 2 个 (*ndhF* 和 *ycf1*); LSC 与 IRb 边界 (JLB) 的基因数量为 1 个 (*rps19*); SSC 与 IRa 边界 (JSA) 的基因数量为 2 个 (*ycf1*); LSC 与 IRa 边界 (JLA) 的基因数量为 0。*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 1 个外显子, 第 1 个外显子位于 LSC, 另外 2 个外显子位于 IR (表 1)。

## 2.3 叶绿体基因组重复序列分析

叶绿体基因组中共检测到 38 个 SSR 位点, 其中, 单碱基重复有 36 个, 双碱基重复有 2 个。其中,

表1 ‘怀玉山’高山马铃薯叶绿体基因功能分类

Table 1 Chloroplast gene functional classification of *S. tuberosum* var. *cormosus* ‘Huaiyushan’

基因功能	基因类型	基因名	基因数量/个	
光合作用	光系统 I	<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>	15	
	NADH 脱氢	<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	
自我复制	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>rrn23*</i> , <i>rrn4.5*</i> , <i>rrn5*</i>	8	
	转运RNA	<i>trnA</i> -UGC*, <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnG</i> -GCC, <i>trnG</i> -UCC, <i>trnH</i> -GUG, <i>trnI</i> -CAU*, <i>trnI</i> -GAU*, <i>trnK</i> -UUU, <i>trnL</i> -CAA*, <i>trnL</i> -UAA, <i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnN</i> -GUU*, <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG*, <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC*, <i>trnV</i> -UAC, <i>trnW</i> -CCA, <i>trnY</i> -GUA, <i>trnY</i> -CAU	37	
	成熟酶	<i>matK</i>	1	
	蛋白酶	<i>clpP1</i>	1	
	其他基因	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>ycf1*</i> , <i>ycf15*</i> , <i>ycf2*</i> , <i>ycf3</i> , <i>ycf4</i>	8	

说明: \*表示该基因的数量有2个。

重复单元为 A/T, 重复频率为 10 的 SSR 位点数量最多 (18 个), 重复频率为 11 的 SSR 位点数量次之 (11 个); 重复单元为 AT/AT、重复频率为 6 的 SSR 位点数量为 2 个。

#### 2.4 叶绿体基因组长重复序列分析

叶绿体基因组共鉴定到 32 个长重复序列, 包括 16 个正向重复 (15 个 30~39 bp, 1 个 40~49 bp), 16 个回文重复 (13 个 30~39 bp, 2 个 40~49 bp, 1 个 50~59 bp), 无反向重复和互补重复。

#### 2.5 叶绿体基因组反向重复区边界结构差异分析

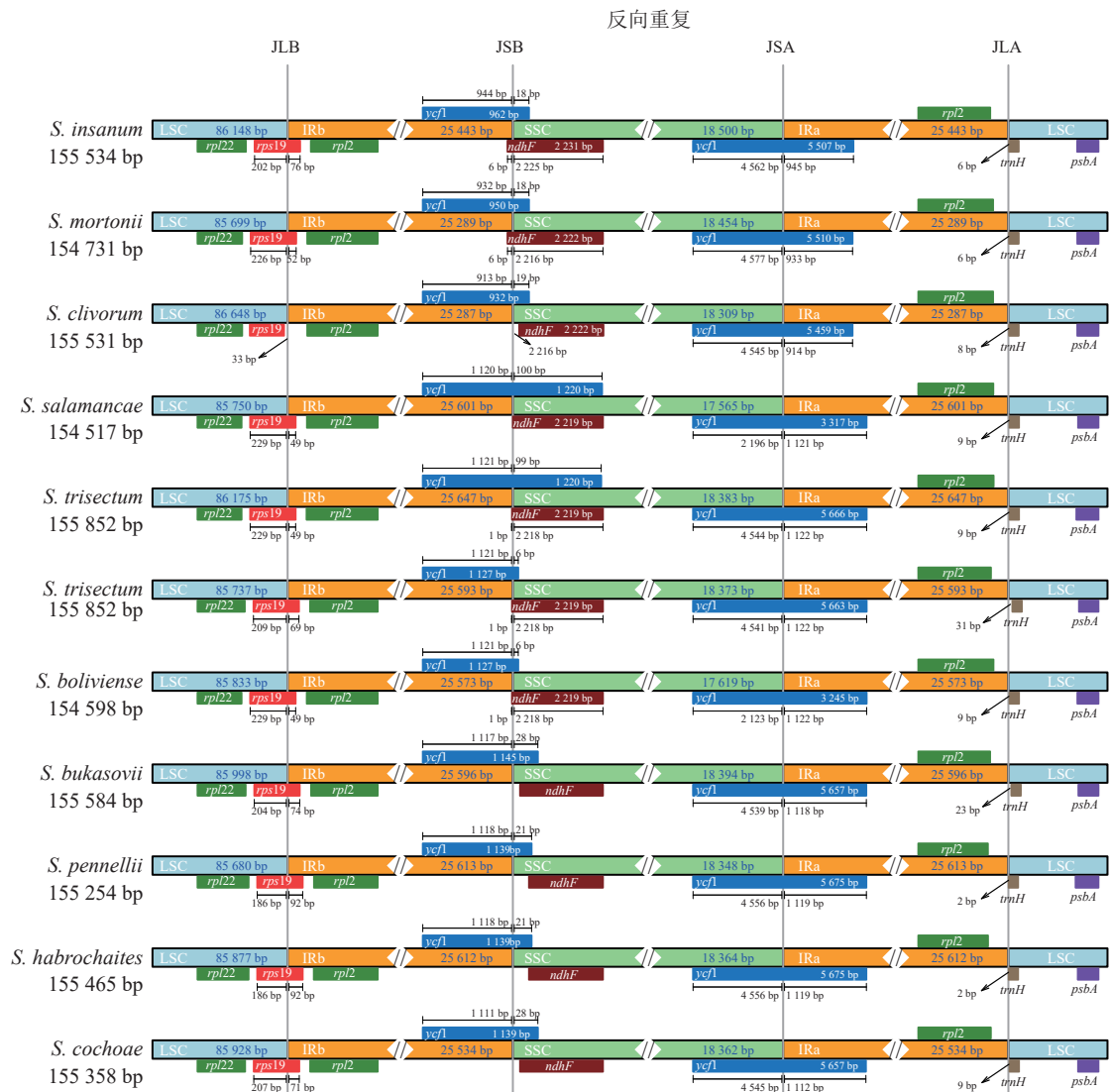
‘怀玉山’高山马铃薯及其 10 个近缘种叶绿体基因组结构从 LSC 中间呈线性展开, 均由 1 个 LSC、1 个 SSC 和 2 个 IR (IRa 和 IRb) 4 部分组成。‘怀玉山’高山马铃薯及其 10 个近缘种 *rpl22*、*rps19*、*rpl2*、*ycf1*、*ndhF*、*trnH* 和 *psbA* 位置基本一致, 但收缩和扩张的长度存在一些差异 (图 2)。

#### 2.6 叶绿体基因组比对分析

‘怀玉山’高山马铃薯及其 10 个近缘种叶绿体基因组核苷酸多样性的变化范围为 0~0.13927, 高变区主要分布在 LSC 和 SSC。LSC 的 *trnL*-UAA-*trnF*-GAA、*cemA*、*rps12*-*exon1*-*clpP1*、*clpP1* 基因变异率最高; SSC 的 *rpl32*-*trnL*-UAG、*ycf1* 基因变异率最高。

#### 2.7 叶绿体基因组密码子使用偏性分析

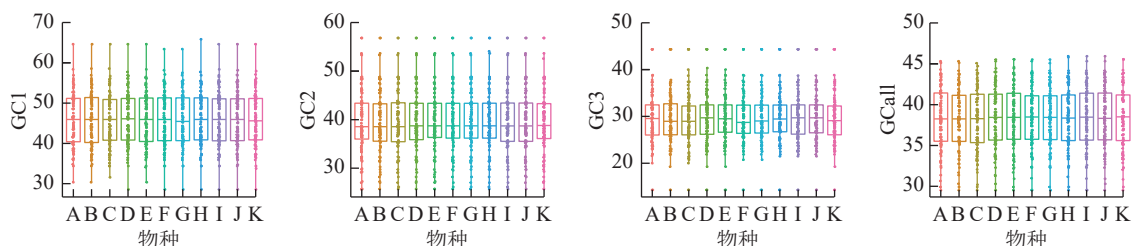
2.7.1 同义密码子的偏性分析 叶绿体基因组 87 个 CDS 基因密码子 3 个位置 GC 比例的平均值为 38.38%, GC1、GC2、GC3 分别为 45.98%、39.55%、29.60%, 这说明 GC 在密码子 3 个位点上的分布存在显著差异, 只有 GC2 与平均 GC 大致接近 (图 3)。ENC 是密码子偏性分析的重要指标, 通常将 35 作为区分值来评估密码子偏倚的强度。叶绿体基因组 87 个 CDS 基因的平均 ENC 为 47.29, ENC>45 的基因有 60 个, ENC>35 的基因有 83 个, 有 4 个基因的 ENC<35, 这表明叶绿体基因组的密码子偏性较弱。通过 SPSS 20.0 进行相关性分析, 结果表明: 密码子总 GC 比例 (GCall) 与 GC1、GC2 在 0.01 水平上均存在极显著的正相关, GCall 与 GC3 在 0.05 水平上显著相关; GC1 与 GC2 在 0.05 水平上存在显著



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。

图 2 ‘怀玉山’高山马铃薯及其 10 个近缘种叶绿体基因组大单拷贝区、小单拷贝区和反向重复区边界位置的比较

Figure 2 Comparison on the boundary locations of large single copy region, small single copy region and inverted repeat region in chloroplast genomes of *S. tuberosum* var. *cormosum* ‘Huaiyushan’ and its 10 related species



GC1、GC2、GC3 分别表示密码子第 1、2、3 位的 GC 比例；GCAll 表示密码子总 GC 比例；字母 A-K 分别代表 *S. cochoae*、多毛番茄、潘那利番茄、*S. bukasovii*、*S. boliviense* NC\_062870、‘怀玉山’高山马铃薯、*S. trisectum*、*S. salamancae*、*S. clivorum*、*S. mertonii*、*S. insanum*。

图 3 ‘怀玉山’高山马铃薯及其 10 个近缘种叶绿体基因组密码子组成成分分析

Figure 3 Composition analysis of chloroplast genome codons of *S. tuberosum* var. *cormosum* ‘Huaiyushan’ and its 10 related species

正相关，但两者均与 GC3 不相关。这表明叶绿体基因组密码子前 2 位的碱基组成相似，而与第 3 位不相似。ENC 与 GC1、GC2、GC3 均不相关，说明密码子上第 1 位、第 2 位和第 3 位的碱基组成对 ENC 没有显著影响。叶绿体基因组 87 个 CDS 基因序列共有 31 个 RSCU>1 的密码子。在这 31 个密码

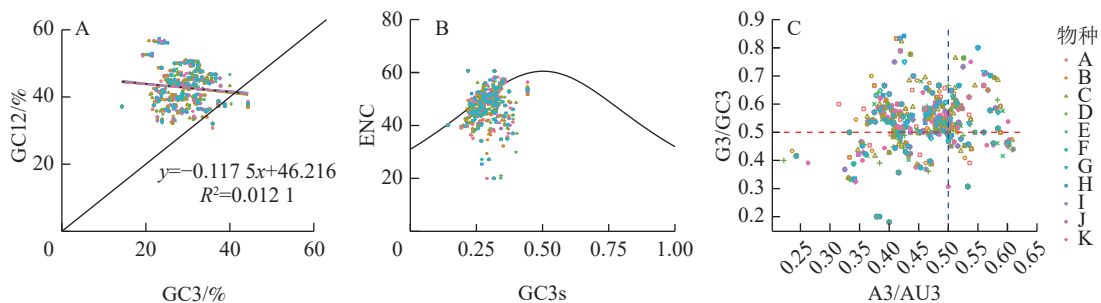
子中，除 AUG、UUG 外，其余都以 A、U 结尾，表明 A、U 碱基在密码子最后位点上出现的频率最高。‘怀玉山’高山马铃薯叶绿体基因组密码子偏好以 A、U 结尾(表 2)。

表 2 ‘怀玉山’高山马铃薯叶绿体基因组同义密码子的使用频率

Table 2 Relative synonymous codon usage (RSCU) of chloroplast genome of *S. tuberosum* var. *cormosus* ‘Huaiyushan’

密码子	氨基酸	相对同义密码子使用频率	数量/个	密码子	氨基酸	相对同义密码子使用频率	数量/个	密码子	氨基酸	相对同义密码子使用频率	数量/个
GCA	Ala	1.134 37	401	GGG	Gly	0.731 07	333	CCU	Pro	1.521 13	432
GCC	Ala	0.718 53	254	GGU	Gly	1.242 59	566	AGC	Ser	0.342 61	119
GCG	Ala	0.390 38	138	CAC	His	0.479 87	149	AGU	Ser	1.191 94	414
GCU	Ala	1.756 72	621	CAU	His	1.520 13	472	UCA	Ser	1.197 70	416
AGA	Arg	1.829 81	491	AUA	Ile	0.909 09	680	UCC	Ser	0.955 85	332
AGG	Arg	0.633 54	170	AUC	Ile	0.609 63	456	UCG	Ser	0.595 97	207
CGA	Arg	1.453 42	390	AUU	Ile	1.481 28	1 108	UCU	Ser	1.715 93	596
CGC	Arg	0.368 94	99	CUA	Leu	0.821 14	391	UAA	Ter	1.655 17	48
CGG	Arg	0.424 85	114	CUC	Leu	0.434 72	207	UAG	Ter	0.758 62	22
CGU	Arg	1.289 44	346	CUG	Leu	0.403 22	192	UGA	Ter	0.586 21	17
AAC	Asn	0.485 74	315	CUU	Leu	1.297 86	618	ACA	Thr	1.221 17	421
AAU	Asn	1.514 26	982	UUA	Leu	1.812 39	863	ACC	Thr	0.771 57	266
GAC	Asp	0.408 80	223	UUG	Leu	1.230 66	586	ACG	Thr	0.446 70	154
GAU	Asp	1.591 20	868	AAA	Lys	1.462 87	1 054	ACU	Thr	1.560 55	538
UGC	Cys	0.556 29	84	AAG	Lys	0.537 13	387	UGG	Trp	1.000 00	490
UGU	Cys	1.443 71	218	AUG	Met	1.987 34	628	UAC	Tyr	0.397 12	193
CAA	Gln	1.491 10	712	GUG	Met	0.012 66	4	UAU	Tyr	1.602 88	779
CAG	Gln	0.508 90	243	UUC	Phe	0.722 19	542	GUA	Val	1.502 75	547
GAA	Glu	1.477 52	1 035	UUU	Phe	1.277 81	959	GUC	Val	0.524 73	191
GAG	Glu	0.522 48	366	CCA	Pro	1.193 66	339	GUG	Val	0.524 73	191
GGA	Gly	1.571 90	716	CCC	Pro	0.742 96	211	GUU	Val	1.447 80	527
GGC	Gly	0.454 45	207	CCG	Pro	0.542 25	154				

2.7.2 GC3-GC12 分析 分析发现：‘怀玉山’高山马铃薯及其 10 个近缘种植物叶绿体基因的 GC3 比例分布为 0.142 9~0.443 2，GC12 比例分布为 0.285 7~0.658 5，两者大多沿对角线上方分布。两者的相关系数 ( $r$ ) 为 0.110 1 ( $R^2=0.012 1$ )，相关不显著 ( $P>0.05$ )，回归斜率为 0.117 5，说明 GC12 与 GC3 不相关(图 4A)。表明‘怀玉山’高山马铃薯叶绿体基因组密码子使用偏性很大程度上受自然选择的影响，而受突变压力的影响小。



GC12. 密码子第 1 和 2 位的 GC 百分含量; GC3. 密码子第 3 位的 GC 百分含量; ENC. 有效密码子数; GC3s. 同义密码子第 3 位的 GC 含量; G3/GC3. 密码子第 3 位的 G 百分含量与密码子第 3 位的 GC 百分含量的比值; A3/AU3. 密码子第 3 位的 A 百分含量与密码子第 3 位的 AU 百分含量的比值; 字母 A-K 分别代表 *S. cochoae*、多毛番茄、潘那利番茄、*S. bukasovii*、*S. boliviense* NC\_062870、‘怀玉山’高山马铃薯、*S. trisectum*、*S. salamancae*、*S. clivorum*、*S. mertonii*、*S. insanum*。

图 4 ‘怀玉山’高山马铃薯及其 10 个近缘种叶绿体基因组密码子 GC3-GC12 分析 (A)、ENC-plot 分析 (B) 和 PR2-plot 分析 (C)

Figure 4 GC3-GC12 analysis (A), ENC-plot analysis (B) and PR2-plot analysis (C) of chloroplast genome codons of *S. tuberosum* var. *cormosus* ‘Huaiyushan’ and its 10 related species

2.7.3 ENC-plot 分析 分析表明: 分布在期望曲线上或曲线附近的基因较少, 分布在期望曲线下方且远离曲线的基因较多, 说明大部分基因的实际 ENC (ENCobs) 与理论 ENC (ENCexp) 存在差异。为了解实际 ENC 和理论 ENC 的差异度, 计算了 ‘怀玉山’高山马铃薯 ENC 比值频数, 即  $(ENC_{exp}-ENC_{obs})/ENC_{exp}$ 。结果表明: ‘怀玉山’高山马铃薯叶绿体基因组基因中, 有 16.47%(14 个) 的基因分布在 0~0.1 区间, 分布于期望曲线上或曲线附近, 即 ENCobs 接近于 ENCexp 值, 有 83.53% 的基因分布在 0~0.1 区间外, 远离期望曲线分布, 即 ENCexp 和 ENCobs 相差较大, 表明自然选择是影响 ‘怀玉山’高山马铃薯叶绿体基因组密码子使用偏性的主要因素, 而突变压力的作用较小 (图 4B)。

2.7.4 PR2-plot 分析 分析表明: A3/AU3 轴、G3/GC3 轴均以 0.5 为界限, 发现 4 个平面内基因分布不均衡。从 G3/GC3 轴看, 多数基因位于上方 ( $>0.5$ ), 少数基因位于下方 ( $<0.5$ ); 从 A3/AU3 轴看, 多数基因位于左侧 ( $<0.5$ ), 少数基因位于右侧 ( $>0.5$ )。这表明 4 种碱基在同义密码子第 3 位上存在 C>G、T>A 现象 (图 4C)。当密码子使用存在偏性完全受突变压力影响时, C 和 G 以及 A 和 T 同义密码子在第 3 位上的分布应相等。因此, ‘怀玉山’高山马铃薯叶绿体基因组密码子使用偏性主要受自然选择等因素影响。

2.7.5 最优密码子确定 RSCU 分析可知: 同时满足  $RSCU > 1$  和  $\Delta RSCU \geq 0.08$  的密码子共 10 个, 即 CGU、AAA、CUU、GUU、GGA、GUA、GGU、UCA、GCU、CCU, 这些密码子都以 A、U 结尾, 被确定为 ‘怀玉山’高山马铃薯叶绿体基因组的最优密码子 (表 3)。

表 3 ‘怀玉山’高山马铃薯叶绿体基因组最优密码子筛选

Table 3 Optimal codon screening of chloroplast genome of *S. tuberosum* var. *cormosus* ‘Huaiyushan’

密码子	氨基酸	相对同义密码子使用率	密码子高表达相对概率	密码子低表达相对概率	$\Delta RSCU$	密码子	氨基酸	相对同义密码子使用率	密码子高表达相对概率	密码子低表达相对概率	$\Delta RSCU$
CGU*	Arg	1.289 44	0.631 58	1.428 57	0.796 99	UUU	Phe	1.277 81	1.315 79	1.200 00	-0.115 79
AAA*	Lys	1.462 87	1.000 00	1.750 00	0.750 00	UUG	Leu	1.230 66	0.750 00	0.600 00	-0.150 00
CUU*	Leu	1.297 86	1.022 73	1.500 00	0.477 27	AUU	Ile	1.481 28	1.428 57	1.263 16	-0.165 41
GUU*	Val	1.447 80	0.800 00	1.250 00	0.450 00	CAU	His	1.520 13	1.333 33	1.000 00	-0.333 33
GGA*	Gly	1.571 90	1.302 33	1.750 00	0.447 67	GAA	Glu	1.477 52	1.469 39	1.000 00	-0.469 39
GUA*	Val	1.502 75	1.400 00	1.750 00	0.350 00	ACA	Thr	1.221 17	1.000 00	0.444 44	-0.555 56
GGU*	Gly	1.242 59	1.302 33	1.500 00	0.197 67	CCA	Pro	1.193 66	1.142 86	0.571 43	-0.571 43
UCA*	Ser	1.197 70	1.000 00	1.153 85	0.153 85	GAU	Asp	1.591 20	1.600 00	1.000 00	-0.600 00
GCU*	Ala	1.756 72	2.105 26	2.250 00	0.144 74	AAU	Asn	1.514 26	1.600 00	1.000 00	-0.600 00
CCU*	Pro	1.521 13	1.571 43	1.714 29	0.142 86	CAA	Gln	1.491 10	1.615 38	1.000 00	-0.615 38
AUG	Met	1.987 34	1.000 00	1.000 00	0.000 00	AGU	Ser	1.191 94	1.444 44	0.769 23	-0.675 21
UAA	Ter	1.655 17	1.200 00	1.200 00	0.000 00	CGA	Arg	1.453 42	1.421 05	0.714 29	-0.706 76
UGU	Cys	1.443 71	1.000 00	1.000 00	0.000 00	UUA	Leu	1.812 39	1.704 55	0.900 00	-0.804 55
ACU	Thr	1.560 55	1.600 00	1.555 56	-0.044 44	UCU	Ser	1.715 93	1.888 89	0.769 23	-1.119 66
GCA	Ala	1.134 37	0.421 05	0.375 00	-0.046 05	AGA	Arg	1.829 81	2.368 42	1.071 43	-1.296 99
UAU	Tyr	1.602 88	1.548 39	1.500 00	-0.048 39						

说明: 标注\*的密码子为最优密码子。

## 2.8 系统发育分析

基于 ‘怀玉山’高山马铃薯和 18 个近缘种以及烟草属 2 个外类群物种叶绿体基因组构建的系统发育树分析可知: 茄属聚为一大类, 烟草属聚为另一大类。在茄属中, MLS 与 *S. tuberosum* NC\_008096 (‘Ddeiree’) 聚为一小分支。说明 ‘怀玉山’高山马铃薯与 *S. tuberosum* ‘Ddeiree’ 亲缘关系较近, 两者同源 (图 5)。

## 3 讨论与结论

叶绿体基因组结构保守、独立母系遗传, 是被子植物基因组的重要组成部分, 广泛用于被子植物的



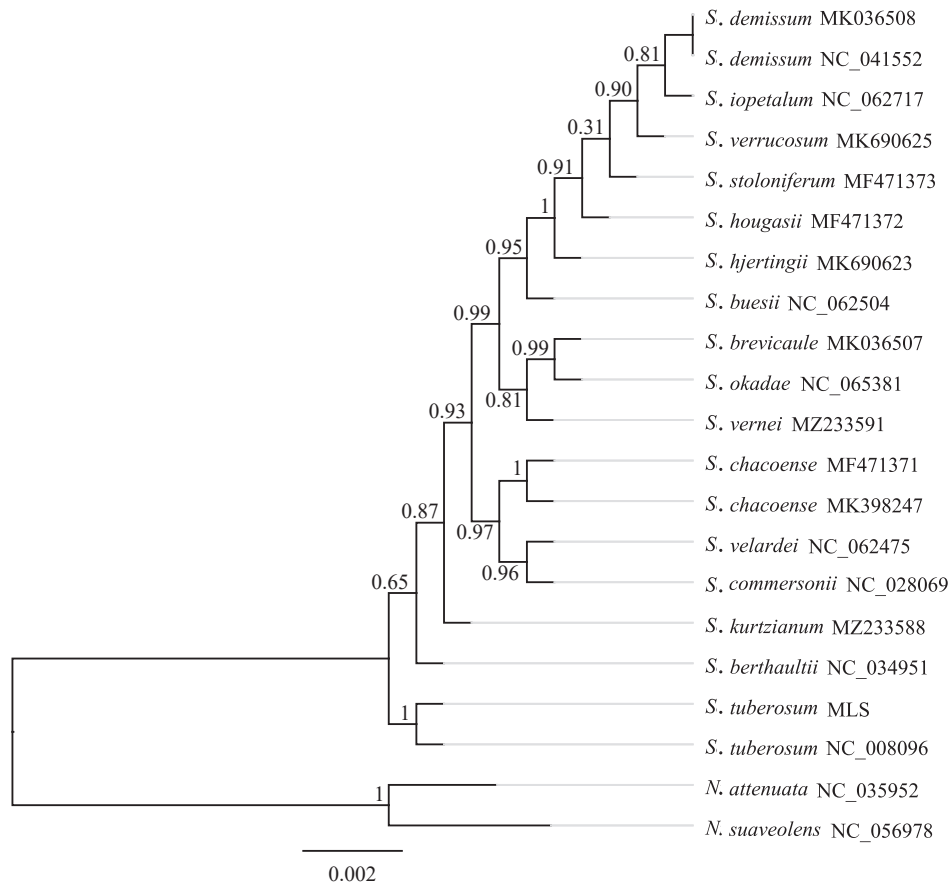


图5 基于叶绿体基因组的‘怀玉山’高山马铃薯及其18个近缘种的系统发育树  
Figure 5 Phylogenetic tree of *S. tuberosum* var. *cormosus* ‘Huaiyushan’ and its 18 related species based on chloroplast genome

生长发育、类群分析和进化分析<sup>[22]</sup>。被子植物叶绿体基因组大小一般为120~180 kb, IR大小一般为20~30 kb<sup>[23-24]</sup>。在本研究中,‘怀玉山’高山马铃薯叶绿体基因组长度和IR长度分别为155 296和25 593 bp,与*S. tuberosum* ‘Shepody’<sup>[16]</sup>叶绿体基因组长度和IR长度一致,与其他马铃薯品种<sup>[13-15, 17-18]</sup>相比,叶绿体基因组长度和IR长度不超过500 bp,说明马铃薯各个品种的叶绿体基因组较为保守。

叶绿体的SSR不仅与核基因组SSR一样,具有高多态性、多等位性、共显性<sup>[25]</sup>,也具有单亲遗传模式,结构简单、相对保守<sup>[26]</sup>,因此,叶绿体的SSR有较好的种间、种内遗传变异区分能力,已成为区分物种的重要分子标记而被广泛应用<sup>[27]</sup>。关惜今等<sup>[13]</sup>研究表明:*S. fernandezianum*与其野生近缘种(*S. phureja*、*S. palustre*、*S. etuberosum*)叶绿体基因组中共检测到36、36、42、40个SSR,SSR类型比较单一,只有单核苷酸和二核苷酸等2种类型,单核苷酸为A和T等2种类型,二核苷酸包括TA和AT等2种类型,其数目比较少。本研究结果与此一致。在本研究中,在‘怀玉山’高山马铃薯叶绿体基因组中共检测到38个SSR位点,其中,单碱基重复有36个,双碱基重复有2个,较少的SSR位点存在表明‘怀玉山’高山马铃薯叶绿体基因组可能不易发生重排。

IR和单拷贝区(SC)边界的膨胀和收缩被认为是被子植物叶绿体全基因组大小变化的主要机制<sup>[28]</sup>,同一属不同品种叶绿体基因组IR/SC边界位置变化也不同<sup>[29]</sup>。关惜今等<sup>[13]</sup>研究表明:*S. fernandezianum*与其野生近缘种(*S. phureja*、*S. palustre*、*S. etuberosum*)叶绿体基因组*rps19*基因均横跨JLB,*S. phureja*的*ndhF*基因横跨JSB,*S. fernandezianum*、*S. palustre*、*S. etuberosum*的*ndhF*基因均右移,分布在SSC,*S. fernandezianum*、*S. phureja*、*S. palustre*、*S. etuberosum*的*ycf1*基因总长度为5 664 bp,均横跨SSC和IRa区域。在本研究中,对‘怀玉山’高山马铃薯及其10个近缘种叶绿体基因组IR/SC边界区域的分析结果表明:这些叶绿体基因组的IR都存在扩张或收缩的现象。‘怀玉山’高山马铃薯的*rps19*基因横跨JLB,横跨JLB的左边和右边长度分别为209和69 bp,在JSA,‘怀玉山’高山马铃薯*ycf1*基因为5 663 bp,左边和右边长度分别为4 541和1 122 bp。

许多植物存在密码子偏好性 (CUB), 即某一或几种特定密码子频率超过其他同义密码子。密码子偏好性可用来评估基因组中蛋白质编码区 (CDS) 的密码子使用情况<sup>[30]</sup>。植物密码子偏好性是物种不断适应外界环境进化所导致的结果, 生物获得特定的密码子使用模式以适应起源、进化、自然选择和突变压力等多种因素<sup>[31]</sup>。影响不同物种中密码子偏好性差异的因素主要有碱基突变、基因表达水平、自然选择等, 自然选择和突变压力被认为是 2 个最重要的因素<sup>[32-34]</sup>。密码子第 3 个碱基的同义突变不能改变氨基酸的类型, 但被认为是决定氨基酸类型的重要特征, 因此 GC3 经常被用作密码子偏向的重要指标<sup>[35-36]</sup>。本研究发现‘怀玉山’高山马铃薯叶绿体基因组的平均 GC 比例为 38.38%, GC3 为 29.60%, 更倾向于使用 A/T 密码子。RSCU 分析结果也证实了这一点。‘怀玉山’高山马铃薯叶绿体基因组中存在 A/T 密码子使用偏向, 这与大多数高等植物的模式一致<sup>[37]</sup>。‘怀玉山’高山马铃薯叶绿体基因组平均 ENC 为 47.29, ENC>35 的基因有 83 个, 有 4 个基因的 ENC<35, 表明‘怀玉山’高山马铃薯叶绿体基因组的密码子偏性较弱。当密码子的使用受到自然选择的影响时, GC3 值往往分布在一个较小的范围内, GC12 和 GC3 之间没有显著的相关性<sup>[38]</sup>。密码子偏好性可以通过调节基因翻译的准确性和效率影响基因表达, 基因表达水平越高, 密码子偏好性越强<sup>[39-40]</sup>。通过建立的高低基因表达库, 本研究挖掘到‘怀玉山’高山马铃薯叶绿体基因组 10 个最优密码子, 即 CGU、AAA、CUU、GUU、GGA、GUA、GGU、UCA、GCU、CCU, 说明‘怀玉山’高山马铃薯叶绿体基因组密码子更偏好于以 A/U 结尾。筛选到的最优密码子可以用于设计叶绿体基因表达载体, 以提高叶绿体基因组中基因的表达水平, 也可以利用已知密码子的使用偏好来推测和预测未知基因的表达和功能, 可为今后从遗传水平上进行‘怀玉山’高山马铃薯育种改良提供参考。

含有足够信息位点的叶绿体基因组已被证明可有效判断系统发育关系, 甚至是在较低的分类学水平下植物之间也有较强的分类学意义, 为物种间系统发育的研究提供了新的思路<sup>[41]</sup>。在本研究中, 在茄属中‘怀玉山’高山马铃薯与 *S. tuberosum* ‘Desiree’ 单独聚为一分支。说明‘怀玉山’高山马铃薯与 *S. tuberosum* ‘Desiree’ 亲缘关系较近, 表明两者同源, 推测‘怀玉山’高山马铃薯可能是 *S. tuberosum* ‘Desiree’ 从美国引种的。

综上所述, 本研究测序组装了‘怀玉山’高山马铃薯叶绿体基因组全序列, 分析了其编码蛋白基因的密码子使用特点, 从高表达优越密码子和高频密码子中选出两者共有的密码子, 最终筛选得到了 10 个叶绿体蛋白编码基因的最优密码子。‘怀玉山’高山马铃薯密码子的偏好性受到突变、选择及其他多方面因素的共同影响, 但自然选择的影响更大, 这为用基因工程手段改造外源基因密码子, 提高其在‘怀玉山’高山马铃薯叶绿体中的表达量提供了参考, 也为在分子水平上研究茄科茄属植物的系统进化提供参考。

## 4 参考文献

- [1] 尹明华, 谭鑫, 郑亚娇, 等. 高海拔生境下‘怀玉山’高山马铃薯和怀玉山本土农家薯块茎的转录组分析[J]. 核农学报, 2019, 33(7): 1330-1339.  
YIN Minghua, TAN Xin, ZHENG Yajiao, et al. The transcriptome analysis of alpine potato tuber and local farm potato tuber in Huaiyushan under high altitude habitats [J]. *Journal of Nuclear Agricultural Sciences*, 2019, 33(7): 1330-1339.
- [2] 尹明华, 刘燕, 郁雪婷, 等. ‘怀玉山’高山马铃薯茎尖再生苗 6 种病毒的 DAS-ELISA 检测与分析[J]. 浙江农业学报, 2017, 29(10): 1699-1705.  
YIN Minghua, LIU Yan, YU Xueting, et al. DAS-ELISA detection and analysis of six kinds of viruses in plantlets regenerated from Huaiyushan high mountain potato shoot-tips [J]. *Acta Agriculturae Zhejiangensis*, 2017, 29(10): 1699-1705.
- [3] 尹明华, 王钦, 张红蕾, 等. 高海拔生境下‘怀玉山’高山马铃薯和本土农家薯的全基因组重测序分析[J]. 基因组学与应用生物学, 2020, 39(3): 1198-1207.  
YIN Minghua, WANG Qin, ZHANG Honglei, et al. Whole genome re-sequencing analysis of alpine potato and local farm potato in Huaiyu Mountain under high altitude habitats [J]. *Genomics and Applied Biology*, 2020, 39(3): 1198-1207.
- [4] 洪森荣, 张铭心, 叶思雨, 等. 高山马铃薯种质资源遗传多样性的同工酶分析[J]. 浙江农业学报, 2018, 30(9): 1445-

- 1453.
- HONG Senrong, ZHANG Mingxin, YE Siyu, *et al.* Genetic diversity analysis of alpine potato germplasm resources by isozyme [J]. *Acta Agriculturae Zhejiangensis*, 2018, **30**(9): 1445 – 1453.
- [5] MEHMETOGLU E, KAYMAZ Y, ATES D, *et al.* The complete chloroplast genome sequence of *Cicer echinospermum*, genome organization and comparison with related species [J/OL]. *Scientia Horticulturae*, 2022, **296**: 110912[2023-01-20]. doi: 10.1016/j.scienta.2022.110912.
- [6] SHENG Jiajing, YAN Mi, WANG Jia, *et al.* The complete chloroplast genome sequences of five *Miscanthus* species, and comparative analyses with other grass plastomes [J/OL]. *Industrial Crops and Products*, 2021, **162**: 113248[2023-01-20]. doi: 10.1016/j.indcrop.2021.113248.
- [7] TANG Danfeng, WEI Fan, KASHIF M H, *et al.* Analysis of chloroplast differences in leaves of rice isonuclear alloplasmic lines [J]. *Protoplasma*, 2018, **255**(3): 863 – 871.
- [8] LI Guoling, PAN Zonglian, GAO Shichen, *et al.* Analysis of synonymous codon usage of chloroplast genome in *Porphyra umbilicalis* [J]. *Genes & Genomics*, 2019, **41**(10): 1173 – 1181.
- [9] DURET L. Evolution of synonymous codon usage in metazoans [J]. *Current Opinion in Genetics & Development*, 2002, **12**(6): 640 – 649.
- [10] HERSHBERG R, PETROV D A. Selection on codon bias [J]. *Annual Review of Genetics*, 2008, **42**: 287 – 299.
- [11] WANG Hongjun, MENG Tao, WEI Wenqiang. Analysis of synonymous codon usage bias in helicase gene from *Autographa californicamultiple ucleopolyhedrovirus* [J]. *Genes & Genomics*, 2018, **40**(7): 767 – 780.
- [12] LONG Shiyu, YAO Huipeng, WU Qi, *et al.* Analysis of compositional bias and codon usage pattern of the coding sequence in Banna virus genome [J]. *Virus Research*, 2018, **258**: 68 – 72.
- [13] 关惜今, 朱智国, 郑昊吉, 等. 马铃薯与其野生近缘种叶绿体基因组差异分析[J]. 云南师范大学学报(自然科学版), 2021, **41**(4): 33 – 40.
- GUAN Xijin, ZHU Zhiguo, ZHENG Haoji, *et al.* Comparative analysis of plastid chloroplast genomes between potato and its wild relatives [J]. *Journal of Yunnan Normal University (Natural Sciences Edition)*, 2021, **41**(4): 33 – 40.
- [14] ZHAO Chunbo, CHEN Shanshan, SUN Kai, *et al.* Sequencing and characterization the complete chloroplast genome of the potato, *Solanum tuberosum* L. [J]. *Mitochondrial DNA Part B*, 2019, **4**(1): 953 – 954.
- [15] CHUNG H J, JUNG J D, PARK H W, *et al.* The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence [J]. *Plant Cell Reports*, 2006, **25**: 1369 – 1379.
- [16] CHEN Shanshan, ZHAO Yanfei, ZHANG Jingying, *et al.* Characterization of the complete chloroplast genome of the *Solanum tuberosum* L. cv. Shepody (Solanaceae) [J]. *Mitochondrial DNA Part B*, 2021, **6**(8): 2342 – 2344.
- [17] PARK T H. Complete chloroplast genome sequence of the wild diploid potato relative, *Solanum acaule* [J]. *Mitochondrial DNA Part B*, 2021, **6**(3): 1189 – 1191.
- [18] PARK T H. Complete chloroplast genome sequence of the wild diploid potato relative, *Solanum brevicaulle* [J]. *Mitochondrial DNA Part B*, 2019, **4**(2): 4159 – 4160.
- [19] KHAN A R, PARK C E, PARK G S, *et al.* The whole chloroplast genome sequence of black nightshade plant (*Solanum nigrum*) [J]. *Mitochondrial DNA Part A*, 2015, **28**(2): 169 – 170.
- [20] 尹明华, 卢咏琪, 罗怿文, 等. 怀玉山高山马铃薯脱落酸和环境胁迫诱导蛋白基因的克隆和序列分析[J]. 西南农业学报, 2021, **34**(6): 1181 – 1187.
- YIN Minghua, LU Yongqi, LUO Yiwen, *et al.* Cloning and sequence analysis of abscisic acid and environmental stress inducible protein gene in alpine potato in Huaiyushan [J]. *Southwest China Journal of Agricultural Sciences*, 2021, **34**(6): 1181 – 1187.
- [21] 尹明华, 叶思雨, 宁本松, 等. 高山马铃薯脱毒苗DNA甲基化的MSAP分析[J]. 核农学报, 2019, **33**(6): 1079 – 1087.
- YING Minghua, YE Siyu, NING Bensong, *et al.* MSAP analysis of genomic DNA methylation in virus-free plantlets of alpine potato [J]. *Journal of Nuclear Agricultural Sciences*, 2019, **33**(6): 1079 – 1087.
- [22] FENG Shangguo, ZHENG Kaixin, JIAO Kaili, *et al.* Complete chloroplast genomes of four *Physalis* species (Solanaceae): lights into genome structure, comparative analysis, and phylogenetic relationships [J/OL]. *BMC Plant Biology*, 2020, **20**(1):

- 242 [2023-01-20]. doi: 10.1186/s12870-020-02429-w.
- [23] ZHANG Tongwu, FANG Yongjun, WANG Xumin, *et al.* The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes [J/OL]. *PLoS One*, 2012, **7**(1): e30531 [2023-01-20]. doi: 10.1371/journal.pone.0030531.
- [24] 李连星, 彭劲谕, 王大玮, 等. 长爪柃[木衣]叶绿体基因组特征系统发育及密码子偏好性分析[J]. *生物工程学报*, 2022, **38**(1): 328 – 342.
- LI Lianxing, PENG Jinyu, WANG Dawei, *et al.* Chloroplast genome phylogeny and codon preference of *Docynia longiunguis* [J]. *Chinese Journal of Biotechnology*, 2022, **38**(1): 328 – 342.
- [25] PROVAN J, POWELL W, HOLLINGSWORTH P M. Chloroplast microsatellites: new tools for studies in plant ecology and evolution [J]. *Trends in Ecology & Evolution*, 2001, **16**(3): 142 – 147.
- [26] KAUNDUN S S, MATSUNOTO S. Heterologous nuclear and chloroplast microsatellite amplification and variation in tea, *Camellia sinensis* [J]. *Genome*, 2002, **45**: 1041 – 1048.
- [27] ZHANG Xinye, SHIRAISHI S, HUANG Minren. Analysis of genetic structure in population of *Larix kaempferi* by chloroplast SSR markers [J]. *Hereditas*, 2004, **26**(4): 486 – 490.
- [28] ZHAO Yuhui, LU Dengxue, HAN Rongbing, *et al.* The complete chloroplast genome sequence of the shrubby cinquefoil *Dasiphora fruticosa* (Rosales: Rosaceae) [J]. *Conservation Genetics Resources*, 2018, **10**(4): 675 – 678.
- [29] WANG Wenbin, YU Huan, WANG Jiahui, *et al.* The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae) [J/OL]. *International Journal of Molecular Sciences*, 2017, **18**(11): 2288 [2023-01-20]. doi: 10.3390/ijms18112288.
- [30] KURLAND C, GALLANT J. Errors of heterologous protein expression [J]. *Current Opinion in Biotechnology*, 1996, **7**(5): 489 – 493.
- [31] QIAN Wenfeng, ZHANG Jianzhi. Codon usage bias and nuclear mRNA concentration: correlation vs. causation [J/OL]. *Proceedings of the National Academy of Sciences*, 2021, **118**: e2104714118 [2023-01-20]. doi: 10.1073/pnas.2104714118.
- [32] GUSTAFSSON C, GOVINDARAJAN S, MINSHULL J. Codon bias and heterologous protein expression [J]. *Trends in Biotechnology*, 2004, **22**(7): 346 – 353.
- [33] BUTT A M, NASRULLAH I, TONG Yigang. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses [J/OL]. *PLoS One*, 2014, **9**(3): e90905 [2023-01-20]. doi: 10.1371/journal.pone.0090905.
- [34] BHATTACHARYYA D, UDDIN A, DAS S, *et al.* Mutation pressure and natural selection on codon usage in chloroplast genes of two species in *Pisum* L. (Fabaceae: Faboideae) [J]. *Mitochondrial DNA Part A*, 2019, **30**(4): 664 – 673.
- [35] GU Wanjun, ZHOU Tong, MA Jianmin, *et al.* The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens* [J]. *Biosystems*, 2004, **73**(2): 89 – 97.
- [36] INGVARSSON P K. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula* [J]. *Molecular Biology and Evolution*, 2006, **24**(3): 836 – 844.
- [37] 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.
- SHANG Mingzhao, LIU Fang, HUA Jinping, *et al.* Analysis on codon usage of chloroplast genome of *Gossypium hirsutum* [J]. *Scientia Agricultura Sinica*, 2011, **44**(2): 245 – 253.
- [38] LIU Qingpo, XUE Qingzhong. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species [J]. *Journal of Genetics*, 2005, **84**(1): 55 – 62.
- [39] PLOTKIN J B, KUDLA G. Synonymous but not the same: the causes and consequences of codon bias [J]. *Nature Reviews Genetics*, 2011, **12**(1): 32 – 42.
- [40] ZHOU Zhipeng, DANG Yunkun, ZHOU Mian, *et al.* Codon usage is an important determinant of gene expression levels largely through its effects on transcription [J/OL]. *Proceedings of the National Academy of Sciences*, 2016, **113**(41): E6117 – E6125 [2023-01-20]. doi: 10.1073/pnas.1606724113.
- [41] ZHOU Tao, ZHU Honghong, WANG Jian, *et al.* Complete chloroplast genome sequence determination of *Rheum* species and comparative chloroplast genomics for the members of Rumiceae [J]. *Plant Cell Reports*, 2020, **39**(6): 811 – 824.