

牛茄子叶绿体全基因组特征及系统发育分析

龚秋怡^{1,2}, 董姝洁^{1,2}, 许琴^{1,2}, 葛宇清^{1,2*}

1. 浙江中医药大学附属第一医院 中心实验室, 浙江 杭州 310005

2. 浙江中医药大学药学院, 浙江 杭州 310053

摘要: **目的** 以牛茄子 *Solanum capsicoides* 新鲜叶片为材料, 对叶绿体基因组序列进行组装、注释和分析, 探究牛茄子的序列特征和同属其他物种的系统发育关系。 **方法** 基于 Illumina HiSeq 高通量测序, 获得牛茄子完整叶绿体基因组序列, 利用生物信息学方法进一步对其功能及特征、密码子偏好性、IR 区域边界、核苷酸多样性、系统发育关系进行分析。 **结果** 牛茄子叶绿体基因组全长为 155 465 bp, 为典型的四分体结构, 总 GC 含量为 37.85%, 其中大单拷贝区、小单拷贝区和反向重复序列的长度分别为 88 265、18 462、24 369 bp; 共编码 131 个基因, 包括 86 个蛋白编码基因、37 个 tRNA 基因、8 个 rRNA 基因。简单重复序列检测表明, 叶绿体基因组包括 51 个 SSR 位点, 其中以 A 和 T 组成的单核苷酸重复次数最多, 占 45.1%。密码子偏好性分析表明, 亮氨酸是使用频率最高的氨基酸, AUU 是使用次数最多的密码子, 密码子偏向使用 A/U 结尾。属内叶绿体基因组比较分析显示, IR 边界区域相对保守, 但仍存在部分基因扩张和收缩等现象; 核苷酸多样性分析获得了 4 个高突变热点: *trnQ-psbK*、*psbF-psbE*、*clpP*、*yef1*, 可作为物种鉴定和遗传多样性研究的潜在特征。系统发育分析显示牛茄子和喀西茄的亲缘关系更近, 存在姐妹关系。 **结论** 牛茄子叶绿体基因组的组装和系统发育分析, 为茄属植物的分类和群体间遗传多样性的研究提供理论依据。

关键词: 牛茄子; 茄属; 叶绿体基因组; 序列特征; 密码子偏好性; 系统发育分析

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2024)12-4150-09

DOI: 10.7501/j.issn.0253-2670.2024.12.022

Complete chloroplast genome and phylogenetic analysis of *Solanum capsicoides*

GONG Qiuyi^{1,2}, DONG Shujie^{1,2}, XU Qin^{1,2}, GE Yuqing^{1,2}

1. The First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou 310005, China

2. School of Pharmaceutical Sciences, Zhejiang Chinese Medical University, Hangzhou 310053, China

Abstract: Objective The chloroplast genome sequence was assembled, annotated and analyzed using fresh leaves of *Solanum capsicoides* to explore the sequence characteristics of the *S. capsicoides* and the phylogenetic relationships with other species in the genus. **Methods** Based on Illumina HiSeq high-throughput sequencing, the complete chloroplast (cp) genome sequence of *S. capsicoides* was obtained, and its function and characteristics, codon preference, IR region boundary, nucleotide diversity and phylogenetic relationship were further analyzed by bioinformatics method. **Results** The cp genome of *S. capsicoides* exhibited a typical quadripartite structure with a length of 155 465 bp, and the total GC content was 37.85%. The lengths of large single copy region, small single copy region, and reverse repeat sequence were 88 265 bp, 18 462 bp and 24 369 bp, respectively. A total of 131 genes were encoded, including 86 protein-coding genes, 37 tRNA genes, eight rRNA genes. Simple repeat sequence analysis showed that the chloroplast genome contained 51 SSR loci, with the single nucleotide repeats consisting of A and T being the most numerous at 45.1%. Codon preference analysis showed that leucine was the most frequently used amino acid and AUU was the most frequently used codon. Besides, codon preference was ending by using A/U. Comparative analysis of chloroplast genomes within the genus showed that the IR boundary region was relatively conserved, while expansion and contraction still existed in some genes. Nucleotide diversity analysis obtained four mutation hotspots: *trnQ-psbK*, *psbF-psbE*, *clpP*, and *yef1*, which could be used as a potential characterization for species identification and genetic diversity study. Phylogenetic analysis showed that *S. capsicoides* and *S.*

收稿日期: 2023-11-06

基金项目: 国家自然科学基金面上项目 (82074071); 浙江中医药大学校级科研项目 (2021JKGJY004)

作者简介: 龚秋怡 (1999—), 女, 硕士研究生, 研究方向为中药分析与鉴定。E-mail: iamgongqiuyi@163.com

*通信作者: 葛宇清, 副研究员, 从事中药分析与鉴定研究。E-mail: geyuqing@zcmu.edu.cn

aculeatissimum were more closely related and exhibited a sister relationship. **Conclusion** The assembly and phylogenetic analysis of the chloroplast genome of *S. capsicoides* provide theoretical basis for the classification of *Solanum* plants and genetic diversity among populations.

Key words: *Solanum capsicoides* All.; solanum genus; chloroplast genome; sequence features; codon preference; phylogenetic analysis

牛茄子 *Solanum capsicoides* All. 是茄科 (*Solanaceae*) 茄属 *Solanum* L. 植物, 分布于我国云南、四川、贵州、广西等地区, 喜生于荒地或灌木丛中。其根或全株入药, 味苦, 辛, 温, 有毒^[1]。牛茄子的地上部分含有甾体生物碱皂苷类、甾体皂苷类、苯丙类成分^[2-3], 具有显著的抗炎、镇痛和抗氧化作用, 民间多用于跌打损伤、慢性咳嗽痰喘、胃脘痛和瘰疬等的治疗。除药用价值外, 牛茄子因其鲜艳的色彩, 也被视为观赏植物。目前已经鉴定了 2 000 多种茄属植物, 它们广泛分布在全世界的热带和亚热带地区。近年来对牛茄子的研究主要集中于其化学成分与药理作用。研究发现牛茄子的甲醇提取物通过影响外周血管扩张, 使心脏肥大减少, 尿量增加和血管功能改善, 并对高血压发展有潜在减弱作用^[4], 且对不同肿瘤细胞系表现出抗增殖活性的同时不表现毒性作用^[5]。由于茄属植物在植株形态等方面存在相似性, 通过形态鉴定茄属植物存在一定困难。此外, 在中药材市场上, 茄属植物品种繁多, 由于牛茄子生长环境因素等的改变, 野生资源逐渐减少, 经常被潜在的伪品所替代, 例如同属植物喀西茄 *Solanum aculeatissimum* Jacquin. 因其植株外观和药用部位与牛茄子极其相似, 常被掺杂于混伪品中^[6], 这可能损害其临床安全性并影响其药理作用。关于茄属植物的系统发育关系和物种分类方面的研究较少, 因此, 通过分子生物学手段鉴别牛茄子对确保用药安全、有效、可控具有十分重要的意义。

叶绿体是自然界高等植物通过光合作用和氧气释放过程将光能转化为化学能的重要细胞器, 在植物的碳水化合物、蛋白质和脂质的生物合成中发挥着重要作用^[7]。最常见的完整植物叶绿体基因组结构包括 1 个大单拷贝区 (large single copy, LSC)、1 个小单拷贝区 (small single copy, SSC) 以及 2 个反向重复区 (inverted repeats, IRa and IRb), 呈双链环状结构, 长度一般为 120~160 kbp^[8]。叶绿体基因组 (cpDNAs) 在大多数植物中由母系遗传^[9], 在基因内容、含量和顺序中高度保守^[10]。尽管叶绿体基因组具有高度保守性, 但部分仍存在差异, 例如

核苷酸位点的插入和缺失, IR 区域的收缩和扩张, 基因的重排和易位^[11-12]。有研究显示, 对茄属植物叶绿体基因组序列的研究有助于增强对生物抗病性、耐寒性等生物抗性的理解^[13]。与核基因组相比, 叶绿体基因组的突变率更低, 遗传方式更稳定, 因此适合作为植物系统发育和物种鉴别的媒介^[14]。随着高通量测序技术不断完善, 极大地丰富了对植物生物学和进化的理解, 扩大了可用于物种鉴定的分子工具, 并促进了有价值的药用植物的栽培。叶绿体全基因组丰富了对植物生物学和进化的理解, 为物种鉴定提供准确有效的分子工具^[15]。

本研究以茄属植物牛茄子叶片为材料, 基于 Illumina 高通量测序技术, 获得牛茄子的叶绿体基因组全系列, 利用生物信息学软件, 对其结构特征和组成、简单重复序列、长重复序列、密码子偏好性、IR 区域的收缩与扩张、核苷酸多样性、系统发育关系进行分析。为丰富茄属植物的研究做出一定贡献, 也为后续茄属植物的起源、亲缘关系、系统发育研究提供理论依据, 更有助于对茄属优良种质的合理保护开发提供技术支持。

1 材料与仪器

1.1 样品来源

牛茄子样品采自浙江省富阳地区 (30°05'N, 119°53'E) 的浙江中医药大学药用植物植物园, 经浙江中医药大学葛宇清副研究员鉴定为茄科植物牛茄子 *S. capsicoides* All.。选择健康无病虫害的新鲜嫩叶, 放入液氮速冻 30 min, 保存于 -80 °C 冰箱备用。叶标本保存在浙江中医药大学药用植物标本馆。

1.2 仪器

Nanodrop 2000 分光光度计 (Thermo Fisher Scientific, 美国); Illumina HiSeq 2500 测序平台 (美国)。

2 方法

2.1 总 DNA 提取

使用改良 CTAB 法从牛茄子的新鲜和健康叶片中提取总基因组 DNA^[16]。通过 1.0% 琼脂糖凝胶电泳和分光光度计检测 DNA 完整性和浓度。

2.2 叶绿体全基因组测序、组装及注释

使用 Illumina HiSeq 2500 平台进行测序, 用 FastQC 评估成对末端 Illumina 读数的质量, 用 Fastp 去除低质量读数。使用 metaSPAdes 软件, 以 NCBI 已发表的蒜芥茄完整叶绿体基因组 (MN218090.1) 作为参考, 重新组装过滤后的读数, 并通过 GeSeq 注释工具对蛋白质编码基因、mRNA 基因、tRNA 基因进行注释^[17]。使用 CPGAVAS2 软件注释蛋白质编码基因^[18]。使用 BLAST 用于进一步人工校正叶绿体基因组的注释。使用 OGDRAW 工具绘制牛茄子叶绿体基因组圆形图谱^[19]。最后, 将完全注释的叶绿体基因组信息上传至 NCBI 数据库, 登录号为 ON203960.1。

2.3 叶绿体基因组数据分析

使用 MISA 工具对牛茄子叶绿体基因组的简单重复序列 (simple sequence repeats, SSRs) 进行检测, 单核苷酸至六核苷酸重复次数分别不少于 10、5、4、3、3、3 个。使用 REPuter 工具对牛茄子叶绿体基因组的长重复序列 (长度 ≥ 30 bp) 进行分析, 正向重复 (forward repeats, F)、反向重复 (reverse repeats, R)、回文重复 (palindromic repeats, P)、互补重复 (complement repeats, C); 汉明距离、最大计算重复次数、最小重复长度分别设为 3、5 000、30。使用 Codon W v1.4.2 软件对牛茄子叶绿体基因组相对同义密码子使用度 (relative synonymous codon usage, RSCU) 进行计算和分析, 参数参考系统默认值^[20]。当 RSCU

值 > 1 时, 表明该密码子使用频率较高, 具有密码子偏好性; 当 RSCU 值 < 1 时, 表明该密码子使用频率较低; 当 RSCU 值 = 1 时, 说明该密码子无偏好性。使用 IR SCOPE 工具生成 8 个茄属物种 IR 边界图进行比较分析。使用 DnaSP v6 软件对 8 个茄属植物进行核酸变异性进行分析, 窗口大小、步移尺寸分别设为 600 bp 和 200 bp^[21]。

2.4 叶绿体基因组系统进化分析

基于本研究获得的牛茄子叶绿体基因组序列数据, 再加上从 NCBI 数据库下载已公布的 11 个茄亚科物种 (7 个茄属物种、2 个辣椒属物种、2 个紫铃花属物种) 的叶绿体基因组序列数据, 以矮牵牛亚科植物 *Petunia exserta* Stehm. (NC_050782.1) 和 *Calibrachoa hybridcultivar* Cerv. (NC_050169.1) 作为外类群。使用 MEGA X 软件构建系统发育树模型 (最大似然法, bootstrap = 1 000) ^[22]。

3 结果与分析

3.1 叶绿体基因组的结构特征基因组成

牛茄子的叶绿体基因组图谱见图 1。牛茄子完整的叶绿体基因组 (GenBank 登录号 ON203960) 显示出一个闭合环状双链结构, 呈典型的 4 段式结构, 全长为 155 465 bp, 包括 1 个大单拷贝区 (large single copy, LSC) 和 1 个小单拷贝区 (small single copy, SSC), 长度分别为 88 265、18 462 bp, 2 个反向重复区 IRa、IRb 长度为 24 369 bp, 见表 1。LSC、SSC 和 IR 区的鸟嘌呤-胞嘧啶 (GC) 含量分

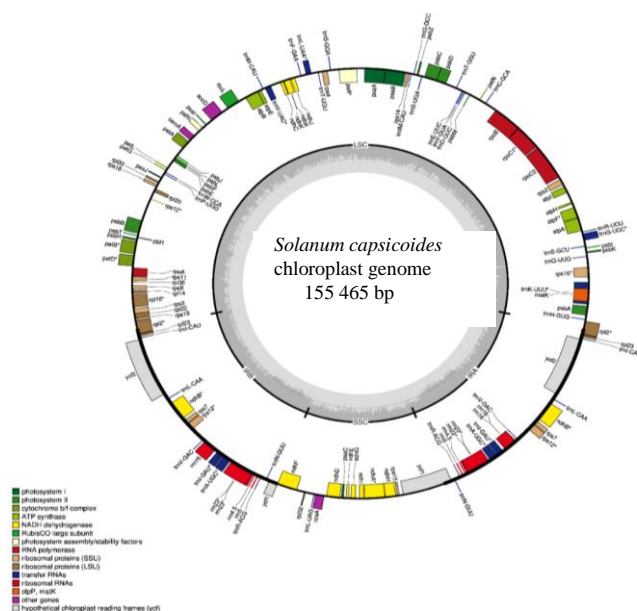


图 1 牛茄子的叶绿体基因组图谱

Fig. 1 Map of *S. capsicoides* chloroplast genome

表 1 牛茄子叶绿体基因组基因功能注释与分类

Table 1 Gene functional annotation and classification of *S. capsicoides* chloroplast genome

基因功能分类	基因种类	基因名称	基因个数	
光合作用相关基因	光系统I亚基	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	5	
	光系统II亚基	<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>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>	15	
	细胞色素亚基	<i>petA</i> , <i>petB</i> ¹⁾ , <i>petD</i> ¹⁾ , <i>petG</i> , <i>petL</i> , <i>petN</i>	6	
	ATP合成酶亚基	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>	6	
	NADH脱氢酶基因	<i>ndhA</i> , <i>ndhB</i> (×2) ¹⁾ , <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	
	Rubisco 大亚基	<i>rbcL</i>	1	
	光系统组装/稳定性因子	<i>psaI</i>	1	
	转录翻译相关基因	RNA聚合酶亚基	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC</i> ¹⁾ , <i>rpoC2</i>	4
		核糖体小亚基	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> (×2) ²⁾ , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> ¹⁾ , <i>rps18</i> , <i>rps19</i>	14
		核糖体大亚基	<i>rpl2</i> (×2) ¹⁾ , <i>rpl14</i> , <i>rpl16</i> ¹⁾ , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	11
转运RNA基因		<i>trnA</i> -UGC (×2), <i>trnN</i> -GUU (×2), <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnM</i> -CAU, <i>trnG</i> -GCC (×2), <i>trnH</i> -GUG, <i>trnI</i> -CAU (×2), <i>trnI</i> -GAU (×2), <i>trnK</i> -UUU ¹⁾ , <i>trnL</i> -UAA, <i>trnL</i> -UAG, <i>trnL</i> -CAA (×2), <i>trnM</i> -CAU, <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG (×2), <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> -UAC, <i>trnV</i> -GAC (×2), <i>trnW</i> -CCA, <i>trnY</i> -GUA	37	
核糖体RNA基因		<i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)	8	
其他基因	成熟酶基因	<i>matK</i>	1	
	依赖ATP蛋白酶P基因	<i>clpP</i> ¹⁾	1	
	C型细胞色素合成基因	<i>ccsA</i> ,	1	
	囊膜蛋白基因	<i>cemA</i>	1	
	乙酰辅酶A羧化酶亚基	<i>accD</i>	1	
	翻译起始因子	<i>infA</i>	1	
未知基因	保守假定的叶绿体开放阅读框	<i>yef1</i> (×2), <i>yef2</i> (×2), <i>yef4</i>	5	

×2: 拷贝数为2; ¹⁾: 1个内含子; ²⁾: 2个内含子。

× 2: two copies; ¹⁾: one intron; ²⁾: two introns.

别为 36.13%、32.07%和 43.18%。牛茄子的叶绿体基因组共编码 131 个基因，包括 86 个蛋白编码基因、37 个 tRNA 基因、8 个 rRNA 基因。在所有的 131 个基因中，有 9 个基因具有 1 个内含子，分别是 *petB*、*petD*、*ndhB*、*rpoC1*、*rps16*、*rpl2*、*rpl16*、*trnK*-UUU、*clpP*；有 1 个基因包含 2 个内含子，是 *rps12*；其余的基因均不含内含子，见表 2。*rps12* 基因是一个特殊的反式剪切基因，3'端位于 IR 区域、5'端位于 LSC 区域。

3.2 重复序列分析

牛茄子叶绿体基因组中共检测到 51 个 SSR 位点，包括单核苷酸重复序列 23 个，二核苷酸重复序列 12 个，三核苷酸重复序列 4 个，四核苷酸重复序列 10 个，六核苷酸重复序列 2 个，未发现五核苷酸重复序列，见表 2。SSRs 的类型主要以 A/T 为主，占总重复序列的 45.10%；其次为 AT/AT，占总重复序列的 23.53%。牛茄子叶绿体基因组中可以检测到

表 2 牛茄子叶绿体基因组 SSRs 位点类型及数量

Table 2 Type and number of SSRs of *S. capsicoides* chloroplast genome

核苷酸类型	重复序列	重复次数	总数	比例/%	合计占比/%
单核苷酸	A/T	23	23	45.10	45.10
二核苷酸	AT/AT	12	12	23.53	23.53
三核苷酸	AAG/CTT	1	4	1.96	7.84
	AAT/ATT	3		5.88	
四核苷酸	AAAC/GTTT	3	10	5.88	19.61
	AAAT/ATTT	5		9.80	
	AGAT/ATCT	2		3.92	
六核苷酸	AATCCT/AGGATT	2	2	3.92	3.92

31 个长重复序列，包括 15 个正向重复序列，1 个回文重复序列，15 个互补重复序列，未检测到反向重复序列。

3.3 密码子偏好性分析

牛茄子叶绿体基因组中共有密码子 28 769 个，

编码 20 种氨基酸。其中，编码亮氨酸 (Leu) 的密码子频率最高，同义密码子为 CUA、CUC、CUG、CUU、UUA、UUG，总数为 3 047 个，占总密码子数量的 10.59%；编码半胱氨酸 (Cys) 的密码子频率最低，同义密码子为 UGC、UGU，总数为 341 个，占总密码子数量的 1.18%。相对同义密码子使用度 (RSCU) 大于 1 的有 28 个，除 UUG 以外，27 个以

A (11) 和 U (16) 结尾，占比 96.42%；相对同义密码子使用度小于 1 的有 30 个，其中 27 个以 C (15) 和 G (12) 结尾，占比 90%；相对同义密码子使用度等于 1 的有 2 个，均以 G 结尾，见图 2。使用频率最高的密码子是 AUU，使用频次为 1180 次，RSCU 值为 1.46；使用频率最低的密码子是 UGA，使用频次为 97 次，RSCU 值为 0.57。

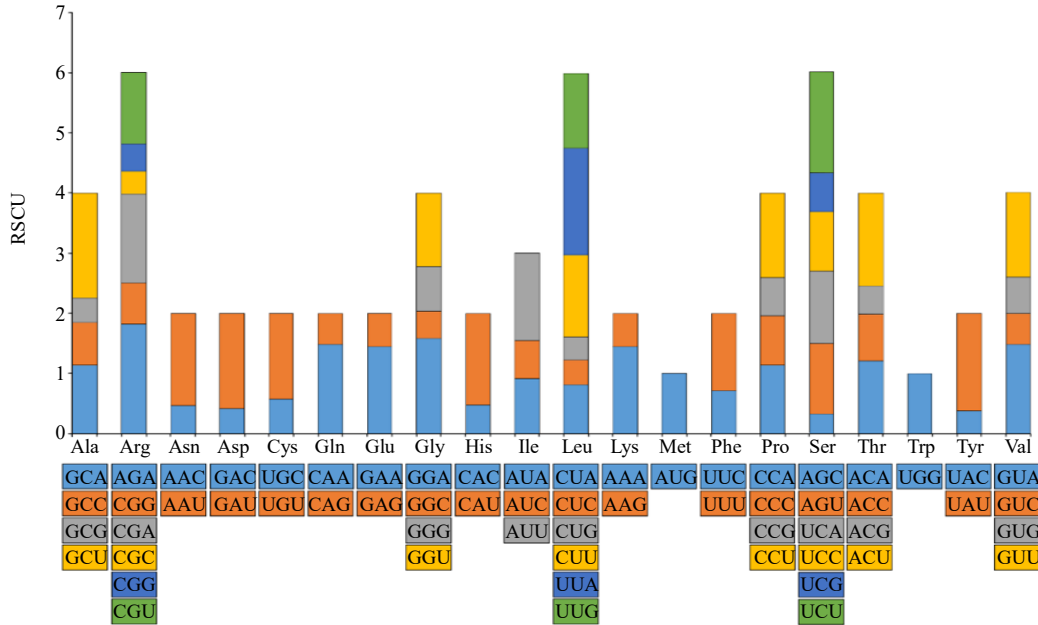


图 2 牛茄子叶绿体基因组相对同义密码子使用度

Fig. 2 Relative synonymous codon usage of *S. capsicoides* chloroplast genome

3.4 牛茄子及茄属部分物种叶绿体基因组特征比较

8 个茄属植物叶绿体基因组特征比较分析见表 3。牛茄子及 7 个茄属物种的叶绿体基因组长度范围在 155 187~155 820 bp，基因组 GC 含量为 37.7%~37.9%，LSC 区 GC 含量为 35.8%~36.1%，SSC 区 GC 含量为 31.9%~32.1%，IR 区 GC 含量

为 43.0%~43.2%。8 个茄属物种的 GC 含量均为 IRs 区>LSC 区>SSC 区。

3.5 IR 区域边界分析

通过在线网站 IR scope 比较分析了茄属 8 个物种的叶绿体基因组 LSC、SSC 和 IR 区域的边界区域收缩与扩张，见图 3。结果表明这 8 个物种的叶

表 3 茄属植物叶绿体基因组基本特征

Table 3 Basic characteristics of chloroplast genome in *Solanum* plants

植物	长度bp				GC%			
	基因组	LSC 区	SSC 区	IRs 区	基因组	LSC 区	SSC 区	IRs 区
牛茄子 (ON203960)	155 465	88 265	18 462	24 369	37.9	36.1	32.1	43.2
龙葵 (NC_039413)	155 541	86 173	18 514	25 427	37.7	35.8	31.9	43.0
富茄 NC_039610)	155 187	86 113	18 244	25 415	37.8	35.9	32.1	43.1
<i>Solanum aureitomentosum</i> (NC_039412)	155 303	85 950	18 495	25 429	37.7	35.8	31.9	43.1
光滑龙葵 (NC_039603)	155 368	86 081	18 517	25 385	37.7	35.8	31.9	43.1
地衣茄 (NC_039598)	155 574	86 217	18 515	25 421	37.7	35.8	31.9	43.1
黄花刺茄 (NC_057245)	155 559	86 281	18 442	25 418	37.8	35.9	31.9	43.1
喀西茄 (NC_061388)	155 820	86 488	18 494	25 419	37.8	35.9	31.9	43.1

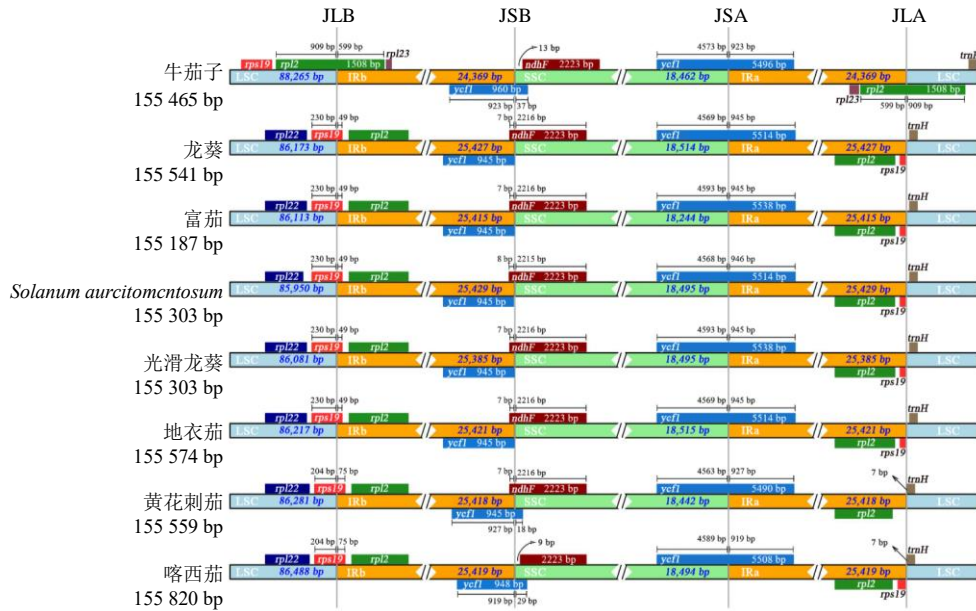


图3 茄属植物叶绿体基因组边界分析

Fig. 3 Boundary analysis of chloroplast genome in *Solanum* plants

绿体基因组长度差异较小，分布范围为 155 187~155 820 bp，序列整体上相似度较高，但在边界区有收缩和扩张现象。在 LSC/IRb (JLB) 边界，除牛茄子外，其他 7 个茄属植物的 JLB 边界均落在 *rps19* 基因，且 *rps19* 基因分布在 IRb 中的长度大多都是 49 bp 和 75 bp，牛茄子的 *rps19* 基因分布在 LSC 中；牛茄子的 JLB 边界落在 *rpl2* 基因上，其余 7 个物种的 *rpl2* 基因均位于 IRb 中。在 IRb/SSC (JSB) 边界，8 个物种均存在 *ycf1* 基因，其中大多 *ycf1* 基因长度均为 945 bp，仅牛茄子和喀西茄的 *ycf1* 基因长度为 960 和 948 bp，且分别朝 SSC 区进入 37 和 29 bp；除牛茄子和喀西茄外，其余 6 个茄属植物的 JSB 边界均落在 *ndhF* 基因上，牛茄子和喀西茄的 *ndhF*

基因距边界分别为 13 和 9 bp。在 SSC/IRa (JSA) 边界，8 个物种的 *ycf1* 基因朝 IRa 区进入了 919~946 bp，长度为 5 490~5 538 bp。在 IRa/LSC (JLA) 边界，牛茄子和壶萼刺茄无 *rps19* 基因且牛茄子的 *rpl2* 基因横跨 JLA 边界，壶萼刺茄和喀西茄的 JLA 边界落在 *trnH* 基因上，均朝 IRa 区进入 7 bp，其余 5 个物种的 JLA 边界全部位于 *rps19* 基因的末端。

3.6 核苷酸多态性分析

使用 DnaSP 软件进行分析，经过 MAFFT 比对后的 8 个茄属植物叶绿体基因组的核苷酸多态性，见图 4。结果表明，LSC 区的变异性整体较高，IR 区变异性最低，表明 IR 区为叶绿体基因组中最为保守的区域。通过核苷酸多样性 (nucleic

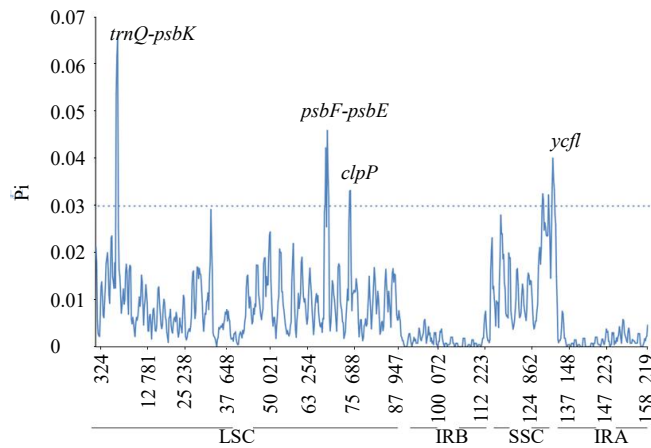


图4 茄属植物叶绿体基因组核苷酸多态性分析

Fig. 4 Nucleotide diversity analysis of chloroplast genome in *Solanum* plants

acid diversity, Pi) (≥ 0.03) 筛选出 4 个高变异区域, 其中 3 个位于 LSC 区, 分别为 *trnQ-psbK*、*psbF-psbE* 和 *clpP*; 1 个位于 SSC 区, 为 *ycf1*。变异度最高的位点为 *trnQ-psbK* ($Pi=0.0656$)。

3.7 系统发育关系分析

为了揭示牛茄子的系统发育位置, 使用 MEGA X, 以矮牵牛亚科植物 *P. exserta* Stehm. 和 *C. hybrid cultivar Cerv.* 作为外类群, 对牛茄子以及其他 11 个有代表性的茄亚科物种进行聚类分析, 包括 7 个茄

属 *Solanum* L. 物种、2 个辣椒属 *Capsicum* L. 物种和 2 个紫铃花属 *Ichroma* L. 物种, 构建系统发育树。分析表明, 12 个茄亚科植物分为 3 个主要进化枝, 各节点的支持率均为 100, 见图 5。进化枝第 1 支是由茄属共 8 个物种组成, 其中牛茄子和喀西茄单独为一组, 第 2 支是由辣椒属共 2 个物种组成, 第 3 支是由紫铃花属共 2 个物种组成。茄属植物组合在一起形成一个单系群, 与辣椒属和紫铃花属物种的分支表现出姐妹关系。

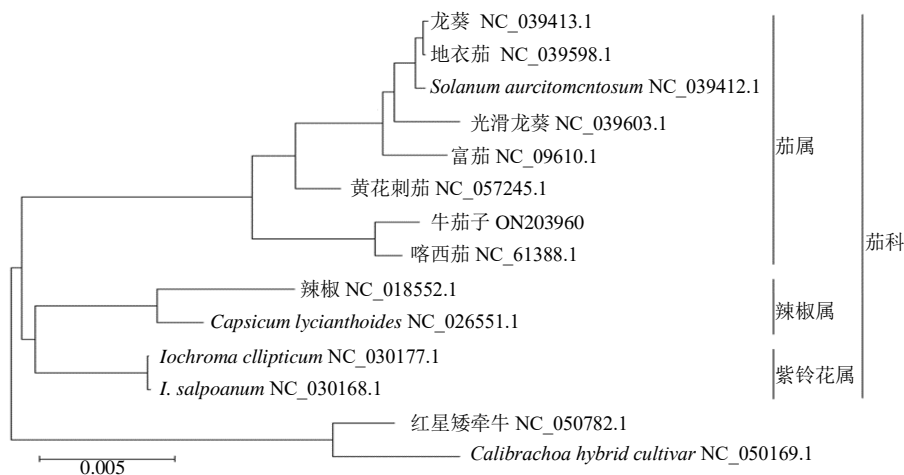


图 5 基于叶绿体基因组构建的茄亚科系统发育树

Fig. 5 Phylogenetic tree of Solanaceae based on chloroplast genome

4 讨论

叶绿体基因组序列是植物鉴定和建立物种间遗传关系的一种通用工具^[23]。本研究基于 Illumina 高通量测序获得茄科茄属植物牛茄子的叶绿体基因组全序列, 并通过生物信息学手段首次对其进行分析研究。牛茄子叶绿体基因组全长为 155 465 bp, 为典型的四分体结构, GC 含量为 37.9%, 共编码 131 个基因。与本研究中选择的 7 个茄属物种相似, 这 8 个茄属物种的叶绿体基因组长度均在 155 000 bp 左右, 基因组 GC 含量在 37.7%~37.9%, IRs 区域 GC 含量最高, 在 43.0%~43.2%, 其原因可能是 IR 区域含有更多的 tRNA 和 rRNA 基因^[24]。本研究将牛茄子叶绿体全基因组序列信息提交至 NCBI 数据库, 丰富了数据库中茄属植物的叶绿体基因组数据。

SSR 是一种高效的分子标记, 具有数量多、重现性高、显性遗传、单亲遗传和相对保守等优点, 是在种群和个体水平上进行物种鉴定和遗传变异评估的最佳选择^[25]。牛茄子叶绿体基因组中共检测到 31 个长重复序列, 其中 F 为 15 个, R 为 0 个, P 为 1 个, C 为 15 个; 共检测到 51 个 SSR 位点, 其中

单核苷酸重复序列 A/T 的重复次数最多, 占 45.10%, 其次是二核苷酸重复序列 AT/AT, 占 23.53%, 其余 SSRs 重复序列也多由 A 和 T 碱基组成, 这与大花茄、水茄、蒜芥茄等茄属植物叶绿体基因的特征相似^[26]。这些发现与先前报道的叶绿体基因组通常由聚腺嘌呤 (poly A) 或胸腺嘧啶 (poly T) 重复组成并很少包含串联鸟嘌呤 (G) 或胞嘧啶 (C) 重复的观点一致^[27]。本研究填补了茄属植物牛茄子 SSR 位点研究的空白, 为牛茄子分子标记的开发和茄属物种鉴定奠定了一定基础。

密码子使用偏好性 (codon usage bias, CUB) 反映物种或基因的起源、突变模式和进化。对基因组中密码子偏好性的研究可以揭示生物之间的系统发育关系、水平基因转移、基因的分子进化, 并确定驱动基因进化的选择性力量^[28]。密码子的使用不仅在生物体之间, 而且在基因组内和单个基因内也有很大差异^[29]。在牛茄子叶绿体基因组的 28 769 个密码子中, 编码亮氨酸的密码子使用频率最高, 编码半胱氨酸的密码子使用频率最低, 并且牛茄子对 A/U 结尾的密码子较为偏好, 这与多数茄属植物叶

叶绿体基因组一致^[30]。

叶绿体基因组的 IR/SC 区域的扩张和收缩会导致各种植物谱系之间的基因组长度和结构差异^[31]。本研究将牛茄子与其他 7 个茄属植物进行对比,发现 8 个植物的叶绿体基因组序列全长都在 155 000 bp 左右,各区域的长度也差异甚微。所有 8 个茄属植物在 JSB、JSA 边界的基因组成方面完全一致,与其他 7 个茄属植物相比,牛茄子在 JLB 边界的 *rps19*、*rpl2* 等基因存在扩张与收缩现象,在 JLA 边界的假基因 *rps19* 存在缺失现象。尽管大多茄属植物在 LSC 和 IRa 区域的边界上拥有 2 个 *rps19* 基因,但也有部分茄属植物如壶萼刺茄^[32]、大花茄、水茄^[26]等存在 IRa 区域的假基因 *rps19* 缺失现象,类似现象也发生在姜科的距药姜属^[33]、豆科的木豆属^[34]、蔷薇科的李属^[35]中。*rps19* 的变化可能是由于部分基因突变和 IR 区域的收缩、扩张^[36]。正是由于这种基因特异性导致不同物种间叶绿体基因组大小的区别,为茄属的起源和进化提供理论基础。已有研究证明,AT 碱基偏好性与叶绿体基因组各区域的变异程度相关,GC 含量较高的 IR 区可使叶绿体基因组减少受到结构重排的影响,维持叶绿体基因组的稳定^[37]。通过对 8 种茄属植物的核苷酸多态性结构进行分析,可以发现这 8 个物种间存在较高相似性,其中 GC 含量较高的 IR 区变异程度较低,而 GC 含量较低的 LSC 区和 SSC 区变异程度较高,即基偏好性可能与序列的变异程度呈正相关。叶绿体基因组高变异的区域可以开发为有效的 DNA 条形码,并广泛用于物种鉴定和大规模系统发育分析^[38]。已有研究证实, *ycf1* 的 DNA 条形码在贝母属物种的鉴定^[39]和报春花属物种的系统发育重建^[40]方面具有优异的能力, *clpP* 序列片段可很好的对龙血树属植物作出鉴别。本研究中, *trnQ-psbK*、*psbF-psbE*、*clpP* 和 *ycf1* 4 个具有丰富可变点的高变异区域可以开发为潜在的 DNA 条形码,尤其 *ycf1* 和 *clpP* 条形码在茄属植物的应用鉴别方面具有巨大潜力,值得进一步研究。以上发现有助于对茄属物种的种间关系进行深入研究。

为进一步探讨牛茄子在进化发育中的地位,从 NCBI 数据库公开的植物中选取了包括牛茄子在内的 8 个茄属物种,选择茄亚科的最近共同祖先辣椒属^[41]和近些年被发现并正式收录的新分支紫铃花属作为代表物种,以矮牵牛亚科植物 *P. exserta* 和 *C. hybrid cultivar* 作为外类群,使用最大似然法进行聚

类分析,显示牛茄子与喀西茄聚为一支,分支支持率达到 100%,成为亲缘关系最近的 2 个物种,确定了牛茄子与茄属植物的亲缘关系。在《中国植物志》分类系统位置中,牛茄子和喀西茄均属与单毛组,与本研究构建的系统发育树结果一致。本研究为茄属植物的起源、亲缘关系及遗传育种提供了进一步的理论依据,更有助于茄属植物物种后续的系统发育的研究及开发该属的植物资源提供理论基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 国家中医药管理局.《中华本草》委员会.中华本草[M].上海:上海科学技术出版社,1999:7315-7316.
- [2] Lu Y Y, Luo J G, Kong L Y. Steroidal alkaloid saponins and steroidal saponins from *Solanum surattense* [J]. *Phytochemistry*, 2011, 72(7): 668-673.
- [3] Wang H, Liu Y, Jang Y K, et al. Phenylpropanoids from *Solanum capsicoides* and their anti-inflammatory activity [J]. *J Asian Nat Prod Res*, 2023, 25(2): 118-124.
- [4] Simões L O, Conceição-Filho G, Ribeiro T S, et al. Evidences of antihypertensive potential of extract from *Solanum capsicoides* All. in spontaneously hypertensive rats [J]. *Phytomedicine*, 2016, 23(5): 498-508.
- [5] Petreanu M, Guimarães Á A A, Broering M F, et al. Antiproliferative and toxicological properties of methanolic extract obtained from *Solanum capsicoides* All. seeds and carpesterol [J]. *Naunyn Schmiedebergs Arch Pharmacol*, 2016, 389(10): 1123-1131.
- [6] 马莉,房志坚.野颠茄及其混淆品喀西茄的性状与显微鉴别[J].*中药材*,2012,35(2):216-220.
- [7] Daniell H, Lin C S, Yu M, et al. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 2016, 17(1): 134.
- [8] Wang Y H, Wang S, Liu Y L, et al. Chloroplast genome variation and phylogenetic relationships of *Atractylodes* species [J]. *BMC Genomics*, 2021, 22(1): 103.
- [9] Zhao L, Jiang X W, Zuo Y J, et al. Multiple events of allopolyploidy in the evolution of the racemose lineages in *Prunus* (Rosaceae) based on integrated evidence from nuclear and plastid data [J]. *PLoS One*, 2016, 11(6): e0157123.
- [10] Li H Y, Chen M D, Wang Z B, et al. Characterization of the complete chloroplast genome and phylogenetic implications of *Euonymus microcarpus* (Oliv.) sprague [J]. *Genes*, 2022, 13(12): 2352.
- [11] Sloan D B, Triant D A, Forrester N J, et al. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae) [J]. *Mol Phylogenet Evol*, 2014, 72: 82-89.
- [12] Ahmed I, Biggs P J, Matthews P J, et al. Mutational dynamics of aroid chloroplast genomes [J]. *Genome Biol Evol*, 2012, 4(12): 1316-1323.

- [13] Song B, Song Y, Fu Y, *et al.* Draft genome sequence of *Solanum aethiopicum* provides insights into disease resistance, drought tolerance, and the evolution of the genome [J]. *Gigascience*, 2019, 8(10): giz115.
- [14] Liu F X, Movahedi A, Yang W G, *et al.* The complete chloroplast genome and characteristics analysis of *Musa basjoo* Siebold [J]. *Mol Biol Rep*, 2021, 48(11): 7113-7125.
- [15] Daniell H, Lin C S, Yu M, *et al.* Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 2016, 17(1): 134.
- [16] Doyle JJ. A rapid DNA isolation procedure for small quantities of fresh leaf tissue [J]. *Phytochem Bull*, 1987, 19: 11-5.
- [17] Tillich M, Lehwark P, Pellizzer T, *et al.* GeSeq- versatile and accurate annotation of organelle genomes [J]. *Nucleic Acids Res*, 2017, 45(W1): W6-W11.
- [18] Shi L C, Chen H M, Jiang M, *et al.* CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. *Nucleic Acids Res*, 2019, 47(W1): W65-W73.
- [19] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes [J]. *Nucleic Acids Res*, 2019, 47(W1): W59-W64.
- [20] 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 Res*, 1987, 15(3): 1281-1295.
- [21] Rozas J, Ferrer-Mata A, Sánchez-DelBarrio J C, *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets [J]. *Mol Biol Evol*, 2017, 34(12): 3299-3302.
- [22] Kumar S, Stecher G, Li M, *et al.* MEGA X: Molecular evolutionary genetics analysis across computing platforms [J]. *Mol Biol Evol*, 2018, 35(6): 1547-1549.
- [23] Nock C J, Waters D L E, Edwards M A, *et al.* Chloroplast genome sequences from total DNA for plant identification [J]. *Plant Biotechnol J*, 2011, 9(3): 328-333.
- [24] Yang Y, Dang Y Y, Li Q, *et al.* Complete chloroplast genome sequence of poisonous and medicinal plant *Datura stramonium*: Organizations and implications for genetic engineering [J]. *PLoS One*, 2014, 9(11): e110656.
- [25] Yang H S, Li X P, Liu D J, *et al.* Genetic diversity and population structure of the endangered medicinal plant *Phellodendron amurense* in China revealed by SSR markers [J]. *Biochem Syst Ecol*, 2016, 66: 286-292.
- [26] Yang Q H, Li Y, Cai L Y, *et al.* Characteristics, comparative analysis, and phylogenetic relationships of chloroplast genomes of cultivars and wild relatives of eggplant (*Solanum melongena*) [J]. *Curr Issues Mol Biol*, 2023, 45(4): 2832-2846.
- [27] Wang L, Wuyun T N, Du H Y, *et al.* Complete chloroplast genome sequences of *Eucommia ulmoides*: Genome structure and evolution [J]. *Tree Genet Genomes*, 2016, 12(1): 12.
- [28] Parvathy S T, Udayasuriyan V, Bhadana V. Codon usage bias [J]. *Mol Biol Rep*, 2022, 49(1): 539-565.
- [29] Plotkin J B, Kudla G. Synonymous but not the same: The causes and consequences of codon bias [J]. *Nat Rev Genet*, 2011, 12(1): 32-42.
- [30] Zhang R Z, Zhang L, Wang W, *et al.* Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild *Solanum* species [J]. *Int J Mol Sci*, 2018, 19(10): 3142.
- [31] Wang C X, Liu J J, Su Y, *et al.* Complete chloroplast genome sequence of *Sonchus brachyotus* helps to elucidate evolutionary relationships with related species of Asteraceae [J]. *Biomed Res Int*, 2021, 2021: 9410496.
- [32] Amiryousefi A, Hyvönen J, Poczai P. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae [J]. *PLoS One*, 2018, 13(4): e0196069.
- [33] Yang Q, Fu G F, Wu Z Q, *et al.* Chloroplast genome evolution in four montane Zingiberaceae taxa in China [J]. *Front Plant Sci*, 2022, 12: 774482.
- [34] Wang W, Yang T, Wang H L, *et al.* Comparative and phylogenetic analyses of the complete chloroplast genomes of six almond species (*Prunus* spp. L.) [J]. *Sci Rep*, 2020, 10(1): 10137.
- [35] Jung J, Kim C, Kim J H. Insights into phylogenetic relationships and genome evolution of subfamily Commelinoideae (Commelinaceae Mirb.) inferred from complete chloroplast genomes [J]. *BMC Genomics*, 2021, 22(1): 231.
- [36] Zhao X Q, Yan M, Ding Y, *et al.* Characterization and comparative analysis of the complete chloroplast genome sequence from *Prunus avium* 'Summit' [J]. *Peer J*, 2019, 7: e8210.
- [37] Guisinger M M, Kuehl J V, Boore J L, *et al.* Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage [J]. *Mol Biol Evol*, 2011, 28(1): 583-600.
- [38] Islam S U, Dar T U H, Khuroo A A, *et al.* DNA barcoding aids in identification of adulterants of *Trillium govanianum* Wall. ex D. Don [J]. *J Appl Res Med Aromat Plants*, 2021, 23: 100305.
- [39] Chen Q, Wu X B, Zhang D Q. Comparison of the abilities of universal, super, and specific DNA barcodes to discriminate among the original species of *Fritillariae cirrhosae bulbis* and its adulterants [J]. *PLoS One*, 2020, 15(2): e0229181.
- [40] Xu W B, Xia B S, Li X W. The complete chloroplast genome sequences of five pinnate-leaved *Primula* species and phylogenetic analyses [J]. *Sci Rep*, 2020, 10(1): 20782.
- [41] Wu F, Tanksley S D. Chromosomal evolution in the plant family Solanaceae [J]. *BMC Genomics*, 2010, 11: 182-189.

[责任编辑 时圣明]