

## • 药材与资源 •

## 红花变豆菜叶绿体基因组组装与序列特征分析研究

王震<sup>1</sup>, 柳驰<sup>2#</sup>, 任伟超<sup>1</sup>, 张美琦<sup>1</sup>, 刘秀波<sup>1\*</sup>, 马伟<sup>1,3,4\*</sup>

1. 黑龙江中医药大学药学院, 黑龙江 哈尔滨 150040

2. 开姆尼茨工业大学信息学院, 德国 开姆尼茨 09111

3. 江苏康缘药业股份有限公司, 江苏 连云港 222001

4. 中药制药过程新技术国家重点实验室, 江苏 连云港 222001

**摘要:**目的 以红花变豆菜 *Sanicula rubriflora* 新鲜叶片为试验材料, 通过高通量测序手段, 对其叶绿体基因组进行组装和序列特征分析, 为进一步开展变豆菜属植物的进化和遗传发育提供指导方法。方法 实验流程按照 Illumina 公司提供的标准操作流程执行, 包括样品质量检测、文库构建、文库质量检测 and 文库测序等流程, 使用已报道的变豆菜 *Sanicula chinensis* 的叶绿体基因组作为参考序列, 得到红花变豆菜完整的叶绿体基因组序列并对其进行组装、特征序列分析和进化分析。结果 红花变豆菜叶绿体基因组具有典型的环状四分体结构, 长度 155 700 bp, 包括 1 个大的单拷贝区 (large single copy, LSC, 85 979 bp), 1 个小的单拷贝区 (small single copy, SSC, 17 053 bp), 1 对相反的序列 (inverted repeats sequence, IR, 26 333 bp); 共注释到 130 个基因, 其中蛋白编码的基因 86 个, tRNA 基因 36 个, rRNA 基因 8 个, AT 含量占叶绿体基因组的 61.83%; 共检测到 1 个正向长重复序列 (forward repeat sequence), 3 个回文长重复序列 (palindromic repeat sequence); 此外, 还检测到了 168 个简单重复序列 (simple sequence repeats, SSR) 位点; 与蛋白编码基因对应的密码子偏好使用 A/T 碱基, 编码异亮氨酸的密码子 (I) 使用次数最高, 编码半胱氨酸的密码子 (C) 使用次数最低; 最后与 7 种已报道的伞形科植物叶绿体基因组和 2 个外类群物种通过最大似然法 (maximum likelihood, ML) 构建系统发育树, 发现红花变豆菜与变豆菜和直刺变豆菜亲缘关系最为密切。结论 红花变豆菜的叶绿体基因组数据为伞形科植物研究提供了更为详细完善的资料, 为该物种叶绿体基因工程和系统进化分析提供参考依据。

**关键词:** 红花变豆菜; 叶绿体基因组; 高通量测序; 系统进化; 最大似然法

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2022)22-7183-08

DOI: 10.7501/j.issn.0253-2670.2022.22.022

Assembly and sequence analysis of chloroplast genome of *Sanicula rubriflora*WANG Zhen<sup>1</sup>, LIU Chi<sup>2</sup>, REN Wei-chao<sup>1</sup>, ZHANG Mei-qi<sup>1</sup>, LIU Xiu-bo<sup>1</sup>, MA Wei<sup>1,3,4</sup>

1. College of Pharmaceutical Sciences, Heilongjiang University of Chinese Medicine, Harbin 150040, China

2. School of Information, Technical University of Kempten, Kempten 09111, Germany

3. Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang 222001, China

4. State Key Laboratory of New-tech for Chinese Medicine Pharmaceutical Process, Lianyungang 222001, China

**Abstract: Objective** The chloroplast genome was assembled and the sequence characteristics analyzed with fresh leaves of *Sanicula rubriflora* as the experimental materials through high-throughput sequencing, so as to provide guidance for the further development of the evolution and genetic development of *S. rubriflora*. **Methods** The experimental procedures were carried out according to Illumina's standard operating procedures, including the sample quality test, the construction of library, the quality test of the sample library and the sequencing of the library, the reported chloroplast genome of *Sanicula chinensis* was used as a reference

收稿日期: 2022-02-02

基金项目: 中央本级重大增减支项目: 名贵中药资源可持续利用能力建设项目 (2060302); 中药制药过程新技术国家重点实验室开放基金项目 (SKL2020M0302)

作者简介: 王震 (1995—), 男, 硕士, 研究方向为药用植物工程。Tel: (0451)87266827 E-mail: 1021707041@qq.com

\*通信作者: 马伟, 博士, 研究员, 博士生导师, 研究方向为药用植物生物工程与中药资源。Tel: (0451)87266827 E-mail: mawei@hljucm.net

刘秀波 (1975—), 教授, 研究方向为药用植物生物工程。Tel: 13796353268 E-mail: 358270831@qq.com

#共同第一作者: 柳驰 (1991—), 男, 硕士, 研究方向为信息分析。Tel: (0451)87266827

sequence, and the complete chloroplast genome of *S. rubriflora* was obtained and assembled, and the assembly characteristic sequence analysis and evolutionary analysis were performed. **Results** The chloroplast genome of *S. rubriflora* had a typical circular quadripartite structure with a size of 155 700 bp in length. In which a large single-copy (LSC) of 85 979 bp and a pair of inverted repeats (IRs) of 26 333 bp were disconnected by a small single-copy (SSC) of 17 053bp. A total of 130 genes were annotated, including 86 protein-coding genes (PCGs), 36 transfer RNA genes (tRNAs) and eight ribosomal RNA genes (rRNAs). AT content accounted for 61.83% of the chloroplast genome. One forward repeat sequence and three palindromic repeat sequences were detected. In addition, 168 simple sequence repeats (SSR) sites were detected. The codon corresponding to the protein coding gene preferred to use A/T base. The codon (I) encoding isoleucine had the highest frequency of use, while the codon (C) encoding cysteine had the lowest frequency. A phylogenetic tree with maximum likelihood (ML) was constructed based on the chloroplast genomes of seven reported Apiaceae species and two other species. It was found that *S. rubriflora* were most closely related with *S. chinensis* and *S. orthacantha*. **Conclusion** The chloroplast genome of *S. rubriflora* provides more detailed and complete data for the study of Apiaceae plants, and provides a reference basis for chloroplast genetic engineering and systematic evolution analysis of this species.

**Key words:** *Sanicula rubriflora* Fr. Schmidt; complete chloroplast genome; high-throughput sequencing; phylogeny; maximum likelihood

红花变豆菜 *Sanicula rubriflora* Fr. Schmidt 是伞形科 (Apiaceae) 变豆菜属多年生草本植物, 是一种可食用的山野菜, 具有一定的药用价值, 变豆菜属全世界约 40 余种<sup>[1]</sup>。其生长在山间林下, 阴湿及腐殖质较多的地方, 海拔 200~470 m, 主要产地在我国的东北地区, 另外, 蒙古、朝鲜、日本北部也有分布<sup>[2]</sup>。变豆菜属植物在中国约有 18 种, 具有一定的经济价值和药用价值<sup>[3]</sup>, 薄片变豆菜 *S. lamelligera* Hance 又叫血经草, 具有治疗跌打损伤和风寒感冒等功效<sup>[4]</sup>, 天蓝变豆菜 *S. caerulescens* Franch.、直刺变豆菜 *S. orthacantha* S. Moor 和川滇变豆菜 *S. astrantiifolia* Wolff ex Kretsch 均为有名的地方药材, 具有活血化瘀和止咳化痰的药效, 同属的红花变豆菜也可能具有相应的药用价值<sup>[5]</sup>。变豆菜属植物主要基于传统的形态学进行鉴定, 如从花粉形态<sup>[6]</sup>和果实微形态出发进行分类, 但是部分品种只从形态学鉴定很难区分, 导致其传统分类发展缓慢, 因此需要寻找新的方法来对其鉴定手段进行补充。

叶绿体普遍存在于陆地植物、藻类和部分原生生物当中, 是绿色植物进行光合作用的细胞器<sup>[7]</sup>, 功能包括产生色素、合成糖和某些氨基酸等<sup>[8]</sup>。被子植物中, 叶绿体基因组相对保守, 绝大多数为双链环状结构, 包括 1 个小的单拷贝区 (small single copy, SSC)、1 个大的单拷贝区 (large single copy, LSC) 以及 2 个编码相同, 方向相反的序列 (inverted repeats sequence, IR) IRa 和 IRb<sup>[9]</sup>。相对于核基因组, 叶绿体基因组具有结构保守、碱基变异速率适中、易于测序等多种优势, 已广泛应用于各植物类群的系统进化研究<sup>[10]</sup>。基于本课题组前期对红花变豆菜叶绿体基因组的研究<sup>[11]</sup>, 对其序

列数据进一步挖掘分析, 对组装后的叶绿体基因组结构、长重复序列分析、简单重复序列 (simple sequence repeat, SSR) 位点、密码子偏好性以及与其他伞形科植物使用最大似然法 (maximum likelihood, ML) 构建系统发育树, 分析近缘物种亲缘关系, 为红花变豆菜的鉴定、系统发育研究提供新的方法。

## 1 材料与方法

### 1.1 材料

新鲜的红花变豆菜样本于 2020 年 8 月 20 日采集自中国黑龙江省伊春市 (N47°81'08", E128°90'97")。样品保存在黑龙江中医药大学, 样品标本号为 YCL20190507007, 由黑龙江中医药大学马伟研究员鉴定为伞形科变豆菜属植物红花变豆菜 *S. rubriflora* Fr. Schmidt。将采集的红花变豆菜新鲜幼嫩的叶片, 通过液氮速冻, 并进行研磨, 于 -80 °C 冰箱保存备用。

### 1.2 方法

**1.2.1 DNA 提取** 使用 CTAB 法提取红花变豆菜的全基因组 DNA。

**1.2.2 测序** 总 DNA 样品经武汉贝纳科技服务有限公司检测合格后, 用机械打断的方法 (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, 使用琼脂糖凝胶电泳进行片段大小选择, 进行 PCR 扩增形成测序文库 (NEBNext® Ultra™ DNA Library Prep Kit for Illumina®), 建好的文库先进行文库质检, 质检合格的文库用 Illumina NovaSeq 进行测序, 对测序得到的序列, 通过 SOAPnuke (version: 1.3.0) 软件进行低质量数据过滤, 过滤

标准为去除 N 碱基含量超过 5% 的 reads、去除低质量 (质量值 ≤ 5) 碱基数目达到 50% 的 reads 和去除有 adapter 污染的 reads, 最终得到有效数据 (clean reads)。

**1.2.3 叶绿体基因组组装** 使用 SPAdes (version: 3.13.0; 参数: -k 127) [12] 软件进行基因组拼接, 选用 NCBI 数据库中近缘参考植物变豆菜叶绿体基因组 (MK208987.1) 与拼接结果进 blastn (version: BLAST 2.2.30+; 参数: -evalue 1 × 10<sup>-5</sup>) 比对, 后续使用 Gapcloser (Version: 1.12) 对 gap (含 N 序列) 进行补洞。

**1.2.4 叶绿体基因组特征分析** 利用专门针对叶绿体的注释软件 CPGAVAS2 (<http://47.96.249.172:16019/analyzer/annotate>) [13] 进行基因注释并绘图红花变豆菜叶绿体基因组物理图; 使用 VMATCH (<http://www.vmatch.de/>) 软件 (参数: minimal repeat size 30 bp) 查找叶绿体基因组中的散在长重复序列片段; 采用 MISA 软件 [版本: 1.0; 默认参数; 对应的各个重复单元 (unit size) 的最少重复次数分别为 1-8、2-4、3-4、4-3、5-3、6-3] 对叶绿体进行 SSR 检测; 使用 CodonW (Version: 1.4.4) 对密码子偏好性进行分析, 统计估算相对同

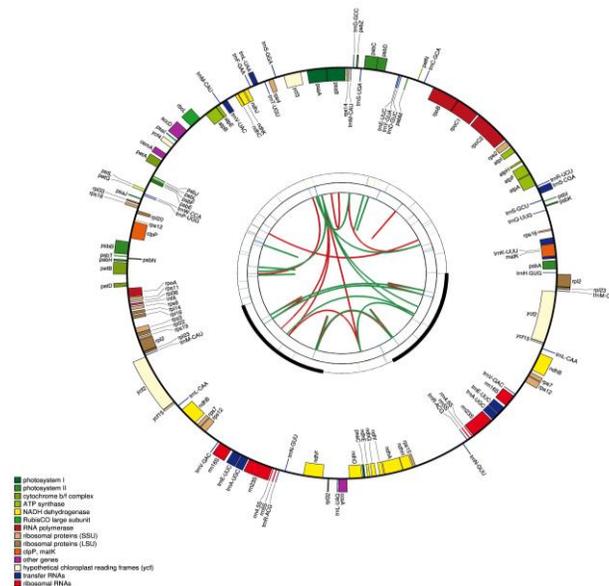
义密码子的使用频率; 通过 NCBI 数据库下载 6 种伞形科植物, 包括变豆菜 *Sanicula chinensis* L. (MK208987)、明日叶 *Angelica keiskei* L. (MW125613)、隔山香 *Ostericum citriodorum* L. (MT501096)、辽藁本 *Ligusticum jeholense* L. (MN652885)、防风 *Saposhnikovia divaricata* L. (MN857472)、白芷 *Angelica dahurica* L. (KT963037) 以及 2 种非伞形科植物玉米 *Zea mays* L. (NC001666) 和拟南芥 *Arabidopsis thaliana* L. (NC000932) 的叶绿体基因组序列, 使用 MAFFT [14] 软件进行多序列比对, 最后使用 MEGA X [14] 软件基于 ML 法构建系统发育树。

## 2 结果与分析

### 2.1 红花变豆菜叶绿体基因组测序及结构解析

基于课题组已报道的研究发现, 红花变豆菜的叶绿体基因组大小为 155 700 bp, 具有典型的环状四分体结构: 包括 1 个 LCS 区 (85 979 bp), 1 个 SCC 区 (17 053 bp) 和 2 个反向互补的 IRs 区 (26 333 bp), GC 含量为 61.83% [14]。

后续通过对红花变豆菜叶绿体基因组物理图谱的绘制 (图 1) 及进一步统计分析发现, 测序有效数据 Q20 值为 97.41%、Q30 值为 92.84%, 根据比



叶绿体基因组图谱上 4 个环。从中心向外, 第 1 圈显示红弧和绿弧分别连接的正向重复和反向重复; 第 2 圈显示用短杠标记的串联重复; 第 3 圈为简单重复序列 SSR; 第 4 圈显示质体上的基因结构; 基因的颜色是根据他们的功能分类来区分

There are four rings on the chloroplast genome map. From the center to the outside, the first circle shows the forward repeat and the reverse repeat connected by the red arc and the green arc respectively; the second circle shows the tandem repeat marked with a short bar; the third circle is the simple sequence repeats; the fourth circle shows the gene structure on the plastid; the color of genes is distinguished according to their functional classification

图 1 红花变豆菜叶绿体基因组环状图

Fig. 1 Ring map of chloroplast genome of *S. rubriflora*

对覆盖度发现变豆菜 (MK208987.1) 为最优参考序列, 其覆盖度为 96.44%, 不含有 gap。IRs、LSC 和 SSC 4 个区域的 GC 值都存在一定的差异, IRs 区域的 CG 值最高, 为 42.92%; SSC 区域的 CG 值最低, 为 32.56%, LCS 区域的 CG 值介于二者之间, 为 36.39% (表 1)。

### 2.2 叶绿体基因组结构基本特征分析

基于课题组已报道的研究发现, 红花变豆菜叶绿体基因组共成功注释了 130 个基因, 其中蛋白编码的基因 86 个, tRNA 基因 36 个, rRNA 基因 8 个<sup>[1]</sup>。

通过对以上结果进一步的统计分析发现, 根据其基因编码功能可以分为 4 类 (表 2), 第 1 类是与自我复制有关的基因, 分为 5 个组, 分别是 tRNA 基因、核糖体小亚基、核糖体大亚基、RNA 聚合酶亚基和核糖体 RNA 基因, 成员数量分别为 26、12、9、4 和 4 个; 第 2 类为与光合作用相关的基因, 细分为 7 组, 分别是光系统 I、光系统 II、ATP

合成亚基、细胞色素亚基、二磷酸核酮糖羧化酶大亚基、NADPH 乳酸脱氢酶亚基和依赖 ATP 蛋白酶 P 基因亚基, 成员数量分别为 5、15、6、6、1、11 和 1 个; 第 3 类为其他功能基因, 包括成熟酶、外膜蛋白基因、C 型细胞色素合成基因和乙酰辅酶 A 羧化酶亚基各 1 个; 第 4 类为 4 个未知功能的基因, 还需要进一步研究, 以确定其功能。

### 2.3 长重复序列分析

所有的长重复序列包括 3 种类型, 包括正向重复序列 (forward repeat sequence)、回文重复序列 (palindromic repeat sequence) 和串联重复序列 (tandem repeat sequence), 这些长重复序列可能具有促进叶绿体基因组重排的功能, 并且可以增加其居群遗传多样性<sup>[16]</sup>。在红花变豆菜叶绿体基因组中共发现了 1 个正向重复序列, 3 个回文重复序列, 并未发现串联重复序列。其中 1 个正向重复序列和 2 个回文重复序列非常短, 分别为 34、34、30 bp, 1 个回文重复序列非常长, 为 26 333 bp,

表 1 红变豆菜叶绿体基因组碱基组成

Table 1 Base composition of chloroplast genome in *S. rubriflora*

区域	起始位点	结束位点	长度	A/%	T/%	C/%	G/%	AT/%	GC/%
总长度	1	155 700	155 700	30.61	31.22	19.40	18.78	61.82	38.18
LSC	1	85 979	85 979	31.16	32.45	18.63	17.76	63.61	36.39
IRs	85 980	112 312	26 333	28.72	28.36	20.74	22.18	57.08	42.92
SSC	112 313	129 365	17 053	34.19	33.26	16.96	15.59	67.44	32.56

表 2 红花变豆菜叶绿体基因组注释基因

Table 2 List of annotated genes in chloroplast genome of *S. rubriflora*

基因功能分类	基因分组	基因名称
自我复制相关	tRNA 基因	<i>trnL-UAG</i> 、 <i>trnR-UCU</i> 、 <i>trnL-UAA</i> 、 <i>trnS-CGA</i> 、 <i>trnS-GCU</i> 、 <i>trnN-GUU</i> 、 <i>trnV-UAC</i> 、 <i>trnY-GUA</i> 、 <i>trnS-UGA</i> 、 <i>trnW-CCA</i> 、 <i>trnE-UUC</i> 、 <i>trnQ-UUG</i> 、 <i>trnF-GAA</i> 、 <i>trnK-UUU</i> 、 <i>trnG-GCC</i> 、 <i>trnV-GAC</i> 、 <i>trnR-ACG</i> 、 <i>trnP-UGG</i> 、 <i>trnH-GUG</i> 、 <i>trnL-CAA</i> 、 <i>trnA-UGC</i> 、 <i>trnS-GGA</i> 、 <i>trnC-GCA</i> 、 <i>trnD-GUC</i> 、 <i>trnM-CAU</i> 、 <i>trnT-UGU</i>
	核糖体小亚基	<i>rps4</i> 、 <i>rps18</i> 、 <i>rps2</i> 、 <i>rps14</i> 、 <i>rps12</i> 、 <i>rps11</i> 、 <i>rps7</i> 、 <i>rps19</i> 、 <i>rps3</i> 、 <i>rps16</i> 、 <i>rps8</i> 、 <i>rps15</i>
	核糖体大亚基	<i>rpl32</i> 、 <i>rpl14</i> 、 <i>rpl33</i> 、 <i>rpl22</i> 、 <i>rpl2</i> 、 <i>rpl36</i> 、 <i>rpl20</i> 、 <i>rpl23</i> 、 <i>rpl16</i>
	RNA 聚合酶亚基	<i>rpoC2</i> 、 <i>rpoB</i> 、 <i>rpoA</i> 、 <i>rpoC1</i>
	核糖体 RNA 基因	<i>rrn4.5S</i> 、 <i>rrn23S</i> 、 <i>rrn5S</i> 、 <i>rrn16S</i>
光合作用相关	光系统 I	<i>psaC</i> 、 <i>psaA</i> 、 <i>psaB</i> 、 <i>psaJ</i> 、 <i>psaI</i>
	光系统 II	<i>psbL</i> 、 <i>psbZ</i> 、 <i>psbN</i> 、 <i>psbK</i> 、 <i>psbC</i> 、 <i>psbM</i> 、 <i>psbA</i> 、 <i>psbJ</i> 、 <i>psbH</i> 、 <i>psbB</i> 、 <i>psbT</i> 、 <i>psbD</i> 、 <i>psbI</i> 、 <i>psbE</i> 、 <i>psbF</i>
	ATP 合成亚基	<i>atpE</i> 、 <i>atpB</i> 、 <i>atpH</i> 、 <i>atpA</i> 、 <i>atpI</i> 、 <i>atpF</i>
	细胞色素亚基	<i>petA</i> 、 <i>petL</i> 、 <i>petD</i> 、 <i>petG</i> 、 <i>petB</i> 、 <i>petN</i>
	二磷酸核酮糖羧化酶大亚基	<i>rbcL</i>
	NADPH 乳酸脱氢酶亚基	<i>ndhB</i> 、 <i>ndhF</i> 、 <i>ndhD</i> 、 <i>ndhJ</i> 、 <i>ndhA</i> 、 <i>ndhC</i> 、 <i>ndhH</i> 、 <i>ndhG</i> 、 <i>ndhI</i> 、 <i>ndhK</i> 、 <i>ndhE</i>
	依赖 ATP 蛋白酶 P 基因亚基	<i>clpP</i>
其他基因	成熟酶	<i>matK</i>
	C 型细胞色素合成基因	<i>ccsA</i>
	乙酰辅酶 A 羧化酶亚基	<i>accD</i>
	外膜蛋白基因	<i>cemA</i>
未知功能基因	保守开放阅读框	<i>ycf4</i> 、 <i>ycf15</i> 、 <i>ycf3</i> 、 <i>ycf2</i>

是叶绿体基因组的 IR 区域。利用这些重复序列，可以为以后开发种群进化标记研究提供基础。

### 2.4 SSR 位点分析

SSR 又称短串联重复标记、微卫星序列标记<sup>[17-18]</sup>。SSR 表示由基因组中 1~6 个核苷酸组成的基本单位重复多次构成的一段 DNA 序列，广泛存在于基因组的各个区域，且侧翼序列通常都是保守性较强的单一序列，可用于个体或物种之间的多态性研究<sup>[19]</sup>。通过对红花变豆菜叶绿体基因组的分析，共发现了 168 个 SSR 位点，分为 4 个类型，包括复杂重复类型 112 个，其数量最多；3 个碱基重复类型 2 个，其数量最少；2 个碱基重复类型 47 个；4 个碱基重复类型 7 个，没有单碱基重复类型。97.60% 的 SSR 位点都含有 A/T 的碱基，仅有 4 个 SSR 位点由 G/C 碱基组成，说明 A/T 碱基具有碱基偏好性，这可能跟红花变豆菜叶绿体基因组中 A/T 碱基含量占比高（61.82%）有关联，造成这种偏好性的原因可能与解链难易程度有关。

### 2.5 密码子偏好性分析

红花变豆菜叶绿体基因组中 86 个蛋白由 21 963 个密码子共同编码（表 3）。由 AUU 编码的异亮氨酸数量最多，为 876 个；由 UGC 编码的半胱氨酸数量最少，仅有 60 个。3 种终止密码子 UAA、UAG 和 UGA 在红花变豆菜的叶绿体基因组中使用，其在同义密码子中的占比为 50%、25% 和 25%。通过对 86 个蛋白编码基因序列分析，得出了同义密码子相对使用度（relative synonymous codon usage, RSCU）<sup>[20]</sup>，RSCU 是对同义密码子使用偏好的评估，通过统计发现，由 UUA 编码的亮氨酸使用最频繁，偏好性最大；由 AGC 编码的丝氨酸使用频率最低，偏好性最小；由 UGG 编码的色氨酸和由 AUG 编码的甲硫氨酸无偏好性。

### 2.6 红花变豆菜与其他 7 种伞形科植物叶绿体基因组的比较分析

本研究基于近些年报道的直刺变豆菜<sup>[21]</sup>、变豆菜<sup>[22]</sup>、明日叶<sup>[23]</sup>、隔山香<sup>[24]</sup>、辽藁本<sup>[25]</sup>、防风<sup>[26]</sup>和白芷<sup>[27]</sup>共 7 种伞形科植物与红花变豆菜的叶绿体序列长度、编码基因数量及组成结构进行比较（表 4）。经过数据统计分析，伞形科内 8 种植物叶绿体基因组的长度为 146 918~155 919 bp，并且都是由典型的四分体结构组成；其大拷贝区和小拷贝区的序列长度差异不大；除了辽藁本的反向重复序列长度为明日叶和防风的 2 倍，其他 7

表 3 红花变豆菜叶绿体基因组密码子使用情况

Table 3 Codon usage in chloroplast genome of *S. rubriflora*

氨基酸	密码子	密码子在同义密码子中的占比/%	密码子个数	同义密码子相对使用度
苯丙氨酸	UUU	61.87	743	1.24
	UCG	38.13	458	0.76
异亮氨酸	AUU	47.51	876	1.43
	AUA	30.69	566	0.92
缬氨酸	AUC	21.80	402	0.65
	UAU	81.29	656	1.63
苏氨酸	UAC	18.71	151	0.37
	ACU	41.31	468	1.65
半胱氨酸	ACC	19.06	216	0.76
	ACG	10.86	123	0.43
	ACA	28.77	326	1.15
组氨酸	UGU	75.00	180	1.50
	UGC	25.00	60	0.50
丝氨酸	CAU	74.32	411	1.50
	CAC	25.68	142	0.50
天冬酰胺	UCU	29.26	498	1.76
	UCC	16.57	282	0.99
	UCA	18.57	316	1.11
	UCG	10.40	177	0.62
天冬氨酸	AGU	19.57	333	1.17
	AGC	5.64	96	0.34
	AAU	75.52	756	1.51
	AAC	24.48	245	0.49
赖氨酸	GAU	80.56	750	1.61
	GAC	19.44	181	0.39
缬氨酸	AAA	73.45	783	1.47
	AAG	26.55	283	0.53
	GUU	36.37	439	1.45
	GUC	11.76	142	0.47
亮氨酸	GUA	36.21	437	1.45
	GUG	15.66	189	0.63
	UUA	29.73	684	1.78
	UUG	21.43	493	1.29
脯氨酸	CUU	21.95	505	1.32
	CUC	6.74	155	0.40
	CUA	14.08	324	0.84
	CUG	6.08	140	0.37
	GCU	43.99	553	1.76
	GCC	18.06	227	0.72
	GCA	27.21	342	1.09
甘氨酸	GCG	10.74	135	0.43
	CCU	38.67	367	1.55
	CCC	19.18	182	0.77
	CCA	27.29	259	1.09
	CCG	14.86	141	0.59
	GGU	32.18	511	1.29
	GGC	11.84	188	0.47
精氨酸	GGA	37.47	595	1.50
	GGG	18.51	294	0.74
	CGU	23.58	316	1.41
	CGC	6.12	82	0.37
	CGA	22.99	308	1.38
	CGG	8.43	113	0.51
	AGA	28.06	376	1.68
谷氨酰胺	AGG	10.82	145	0.65
	CAA	73.61	597	1.47
谷氨酸	CAG	26.39	214	0.53
	GAA	73.11	832	1.46
甲硫氨酸	GAG	26.89	306	0.54
	UGG	100.00	392	1.00
终止	AUG	100.00	502	1.00
	UAA	50.00	28	1.50
	UAG	25.00	14	0.75
	UGA	25.00	14	0.75

表 4 伞形科内 8 种植物叶绿体基因组特征比较

Table 4 Comparison of chloroplast genome characteristics of eight plants in Apiaceae

基因组特征	基因登录号	序列全长/bp	基因总数	编码蛋白基因数	大拷贝区长度/bp	小拷贝区长度/bp	反向重复序列长度/bp
红花变豆菜	MW690208	155 700	130	86	85 979	17 053	26 333
直刺变豆菜	MK293943	157 163	129	84	87 547	17 122	26 247
变豆菜	MK208987	154 650	129	84	85 263	16 927	26 230
明日叶	MW125613	147 007	128	84	92 415	17 576	18 508
隔山香	MT501096	155 919	127	81	85 393	19 760	25 383
辽藁本	MN652885	148 493	127	83	93 932	19 629	36 932
防风	MN857472	147 834	115	81	93 202	17 342	18 654
白芷	KT963037	146 918	129	85	93 605	17 669	17 822

种反向重复序列长度差异也不大；从编码的蛋白质数量来看，红花变豆菜编码的蛋白质数量最多，可能其叶绿体在生长过程中能行使更多的功能。红花变豆菜相比与变豆菜与直刺变豆菜相比，编码的蛋白最多和 SSC 区序列最长。除此之外，这 8 种伞形科植物的叶绿体基因组基因顺序和结构与大多数已报道的被子植物叶绿体基因组相似，这说明在植物进化过程中，叶绿体基因序列具有高度的保守性<sup>[28]</sup>。

### 2.7 叶绿体基因组全序列的聚类分析

叶绿体基因组的聚类分析对植物发育进化研究具有重要意义<sup>[28]</sup>。通过对红花变豆菜与其他 7 种伞形科植物叶绿体基因组序列长度，基因结构以及编码蛋白分析，发现其并未有太大差异，于是通过 ML 法对其构建系统发育树，并加入玉米和拟南芥的叶绿体基因组序列作为外类群，对红花变豆菜进行进化分析，以确定其在伞形科植物中的进化位置（图 2）。本课题组发现不同科的植物分为不同的进化分支，8 种伞形科植物聚在了 1 个分支上；说明植物科间关系明确；在这个大的分支基础上，红花变豆菜与变豆菜和直刺变豆菜聚到了 1 个小分支上，并且自展值为 100，说明这 3 个物种在漫长的进化过程中亲缘关系最为接近。

### 3 讨论

本研究采用高通量测序技术对红花变豆菜叶绿体基因组进行重测序，并以已报道植物变豆菜叶绿体基因组为参考，成功组装出其完整的叶绿体基因组。相对于传统意义上叶绿体基因组序列的获取，如紫荆泽兰 *Ageratina adenophora* L.<sup>[29]</sup> 和菝葜 *Smilax china* L.<sup>[30]</sup> 都是采用先从植物样本中分离出叶绿体，然后对其叶绿体进行 DNA 提取，最终再通过测序技术，实现叶绿体基因组序列的获取，该方法复杂繁琐、不易操作，不利于大范围使

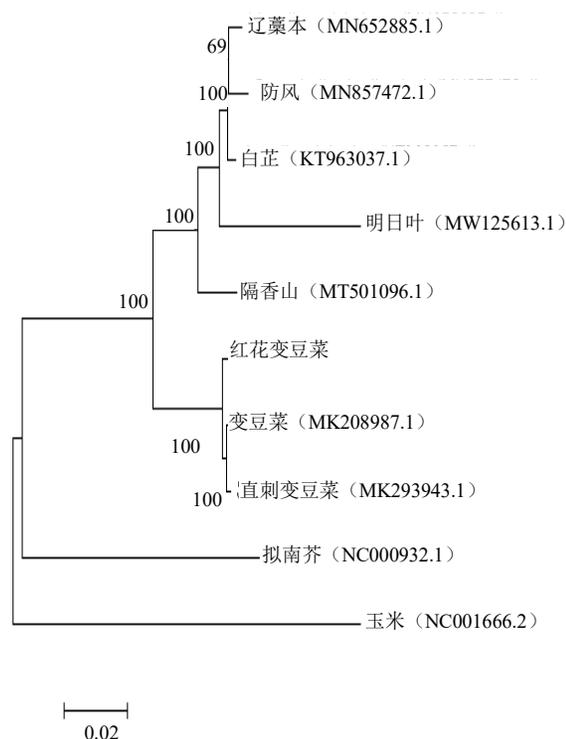


图 2 基于叶绿体全基因组 10 个物种的 ML 系统发育树结果

Fig. 2 Cluster analysis of 10 species using complete chloroplast genome sequence by ML method

用。利用高通量测序技术对红花变豆菜全基因组 DNA 进行重测序，省略了先分离植物叶绿体再提取叶绿体 DNA 的复杂操作过程<sup>[31]</sup>，只需提取红花变豆菜全基因组 DNA，进行高通量测序，选取已报道植物变豆菜叶绿体基因组序列作为参考基因组，将其所测得的全基因组序列与参考叶绿体基因组序列进行 BLASTN 比对，提取出关联的叶绿体 raw reads，使用过滤软件 SOAPnuke 对 reads 进行低质量序列过滤得到 clean reads，再利用 SPAdes 软件对这些序列进行组装及优化，最后使用

GapCloser 软件对组装结果进行补洞, 最终得到完整的红花变豆菜叶绿体基因组序列。本研究使用的方法相较于传统方法, 大大简化了试验步骤、降低了试验所需的时间、减少了实验成本, 并且该方法实用性广, 限制条件少, 为大量植物测序叶绿体基因组提供了可能。

变豆菜属 *Sanicula* L. 在克朗奎斯特系统 (Cronquist system)、哈钦松系统、恩格勒系统 (Engler system)、被子植物发育系统 (angiosperm phylogeny group, APG) 以及《中国植物志》等各个国家的植物分类体系中都被分为一个自然的类群, 归属于伞形科 (Apiaceae)、变豆菜亚科 (Saniculoideae Drude) [32]。其中一些变豆菜属植物具有一定的药用价值, 如变豆菜具有治疗风湿、咳嗽和激活血液循环的功效, 直刺变豆菜具有祛风止咳、活血通络和清热解毒的功效, 也是一味著名的民族药, 收录于《四川省中草药标准 (试行稿)》。中国变豆菜属植物约有 18 种, 属于特有的为 12 种, 但是鉴定困难, 种间属间易混淆, 极其不利于此类植物的发展与利用。针对其分类已经有相关的研究, 包括通过形态学研究和分子系统学研究 [34]。本研究通过对红花变豆菜叶绿体基本组进行测序及数据挖掘, 分析了红花变豆菜的长重复序列、SSR 位点和密码子偏好性, 为变豆菜属植物研究提供了更为详细、完善的资料, 为该物种叶绿体基因工程和系统进化分析提供参考依据。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] Wyk B E V, Tilney P M, Magee A R. African Apiaceae: A Synopsis of the Apiaceae/umbelliferae of sub-Saharan Africa and Madagascar [M]. Pretoria: Briza Academic Books, 2013: 288.
- [2] Vandellook F, Assche J V. Deep complex morphophysiological dormancy in *Sanicula europaea* (Apiaceae) fits a recurring pattern of dormancy types in genera with an Arcto-Tertiary distribution [J]. *Botany*, 2008, 86: 1370-1377.
- [3] 谢文远, 马丹丹, 陈锋, 等. 黄花变豆菜: 浙江变豆菜属(伞形科)一新种 [J]. 杭州师范大学学报: 自然科学版, 2019, 18(1): 9-12.
- [4] 四川省中草药标准 (试行稿) [S]. 1979: 10.
- [5] 周小江. 肺经草的品质研究 [D]. 成都: 成都中医药大学, 2000.
- [6] 姚雪莹, 陈志祥, 王奇志. 变豆菜属 11 种植物的叶表皮微形态特征研究 [J]. 植物研究, 2019, 39(5): 683-691.
- [7] 马孟莉, 张薇, 孟衡玲, 等. 豆蔻属药用植物叶绿体基因组密码子偏性分析 [J]. 中草药, 2021, 52(12): 3661-3670.
- [8] 邢少辰, Clarke J L. 叶绿体基因组研究进展 [J]. 生物化学与生物物理进展, 2008, 35(1): 21-28.
- [9] Zhang T W, Fang Y J, Wang X M, et al. The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: Insights into the evolution of plant organellar genomes [J]. *PLoS One*, 2012, 7(1): e30531.
- [10] 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.
- [11] Wang Z, Ren W C, Yan S, et al. Characterization of the complete chloroplast genome of *Sanicula rubriflora* F. Schmidt ex Maxim [J]. *Mitochondrial DNA B Resour*, 2021, 6(7): 1999-2000.
- [12] Bankevich A, Nurk S, Antipov D, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing [J]. *J Comput Biol*, 2012, 19(5): 455-477.
- [13] 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.
- [14] Rozewicki J, Li S L, Amada K M, et al. MAFFT-DASH: Integrated protein sequence and structural alignment [J]. *Nucleic Acids Res*, 2019, 47(W1): W5-W10.
- [15] 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.
- [16] Tourasse N J, Shtaida N, Khozin-Goldberg I, et al. The complete mitochondrial genome sequence of the green microalga *Lobosphaera (Parietochloris) incisa* reveals a new type of palindromic repetitive repeat [J]. *BMC Genomics*, 2015, 16(1): 580.
- [17] Farhangian-Kashani S, Azadi A, Khaghani S, et al. Association analysis and evaluation of genetic diversity in wheat genotypes using SSR markers [J]. *Biol Futur*, 2021, 72(4): 441-452.
- [18] Shioi S, Shimamoto A, Nakagami Y, et al. Precision length determination and *in silico* simulation in PCR of microsatellite repeat sequences [J]. *Electrophoresis*, 2021, 42(12/13): 1323-1332.
- [19] 陈中苏直, 田波, 蔡传涛. 基于 SSR 分子标记的滇重楼遗传多样性研究 [J]. 中草药, 2017, 48(9): 1834-1838.
- [20] 周茜, 陈芸, 王玉州, 等. 独行菜 LaBBX 基因低温表达响应与密码子偏性分析 [J]. 核农学报, 2021, 35(6): 1253-1262.
- [21] Chen Z X, Yao X Y, Downie S R, et al. Assembling and

- analysis of *Sanicula orthacantha* chloroplast genome [J]. *Biodivers Sci*, 2019, 27(4): 366-372.
- [22] Chen Z X, Yao X Y, Wang Q Z. The complete chloroplast genome of *Sanicula chinensis* [J]. *Mitochondrial DNA Part B*, 2019, 4(1): 734-735.
- [23] Yan W J, Yang T G, Qin R, et al. Complete chloroplast genome of *Angelica keiskei* (Umbelliferae) [J]. *Mitochondrial DNA B Resour*, 2021, 6(2): 536-537.
- [24] Liao C Y, Gao Q, Wu F. The chloroplast genome of *Ostericum citriodorum* (Apiaceae), an endemic medicinal plant to China [J]. *Mitochondrial DNA B Resour*, 2020, 5(3): 3782-3783.
- [25] Wang S J, Zhang T T, Xu L, et al. The complete chloroplast genome of *Ligusticum jeholense* (Umbelliferae, *Ligusticum* L.) [J]. *Mitochondrial DNA B Resour*, 2020, 5(3): 3025-3027.
- [26] Li L L, Geng M L, Li Y S, et al. Characterization of the complete plastome of *Saposhnikovia divaricata* (Turcz.) schischk [J]. *Mitochond DNA B Resour*, 2020, 5(1): 786-787.
- [27] Zhang R, Xu B H, Cao T Y. Characterization of the complete chloroplast genome of *Angelica dahurica* (Apiaceae) as an herb in China [J]. *Mitochondrial DNA B Resour*, 2020, 5(1): 678-679.
- [28] 朱婷婷, 张磊, 陈万生, 等. 1342 个植物叶绿体基因组分析 [J]. 基因组学与应用生物学, 2017, 36(10): 4323-4333.
- [29] Nie X J, Lv S Z, Zhang Y X, et al. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*) [J]. *PLoS One*, 2012, 7(5): e36869.
- [30] Liu J, Qi Z C, Zhao Y P, et al. Complete cpDNA genome sequence of *Smilax china* and phylogenetic placement of Liliales: Influences of gene partitions and taxon sampling [J]. *Mol Phylogenet Evol*, 2012, 64(3): 545-562.
- [31] 王虹雨, 李锦萍, 张瑞峰, 等. 基于高通量技术的椭圆叶花锚叶绿体全基因组测序及系统发育分析 [J/OL]. 分子植物育种: [2022-10-21]. <http://kns.cnki.net/kcms/detail/46.1068.S.20210611.1420.012.html>.
- [32] 姚雪莹. 直刺变豆菜的谱系地理学和遗传多样性研究 [D]. 泉州: 华侨大学, 2020.
- [33] 陈志祥, 姚雪莹, Stephen R D, 等. 直刺变豆菜叶绿体全基因组及其特征 [J]. 生物多样性, 2019, 27(4): 366-372.
- [34] 杨晨, 陈志祥, 姚雪莹, 等. 中国 15 种变豆菜属植物的花粉形态及系统学分析 [J]. 植物研究, 2020, 40(6): 805-812.

[责任编辑 时圣明]