



草地学报  
Acta Agrestia Sinica  
ISSN 1007-0435, CN 11-3362/S

## 《草地学报》网络首发论文

题目： 苜蓿藜叶绿体基因组的解析与系统发育分析  
作者： 孔卫青, 嵇苏, 杨金宏  
收稿日期： 2024-07-23  
网络首发日期： 2024-12-10  
引用格式： 孔卫青, 嵇苏, 杨金宏. 苜蓿藜叶绿体基因组的解析与系统发育分析[J/OL]. 草地学报. <https://link.cnki.net/urlid/11.3362.S.20241210.1541.006>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188, CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

# 苋色藜叶绿体基因组的解析与系统发育分析

孔卫青, 赧苏, 杨金宏\*

(安康学院现代农业与生物科技学院/陕西省蚕桑重点实验室, 陕西 安康 725000)

**摘要:** 本研究对苋色藜 (*Chenopodium amaranticolor*) 的叶绿体基因组进行了高通量测序、组装和注释分析, 比较了藜属植物叶绿体基因组的特征和系统发育关系, 发现苋色藜叶绿体基因组长 152 193 bp, GC 含量 37.25%, 编码 131 个基因, 其中蛋白质编码基因 86 个, 偏好使用 A/U 结尾的密码子。共线性和滑动窗口分析藜属植物叶绿体基因组具有较高序列相似性, 核苷酸多态性主要在 LSC 和 SSC 区的基因间隔区以及 *ndhF* 和 *ycf1* 基因内部; SSR 以单碱基重复为主, 主要组成碱基为 A、T; 散在重复序列主要为正向重复和回文重复; 系统发育分析苋色藜与台湾藜 (*Chenopodium formosanum*) 的亲缘关系较近; 计算基于核基因的苋色藜与藜麦 (*Chenopodium quinoa*) 的遗传距离大于叶绿体基因。本研究可为藜属植物遗传结构与多样性、特异基因挖掘以及系统发育等研究提供参考。

**关键词:** 苋色藜; 叶绿体基因组; 高通量测序; 系统发育

中图分类号: Q949.4 文献标识码: A

## Chloroplast Genome Characteristics and Phylogenetic

### Analysis of *Chenopodium amaranticolor*

KONG Wei-qing, ZHUO Su, YANG Jin-hong

(School of Modern Agriculture & Biotechnology / Shaanxi Key Laboratory of Sericulture, Ankang University, Ankang, Shaanxi 725000, China)

**Abstract:** This study conducted high-throughput sequencing, assembly, and annotation analysis of the chloroplast genome of *Chenopodium amaranticolor*. The analysis of features and phylogeny of *Chenopodium* chloroplast genome discovered that the chloroplast genome of *C. amaranticolor* was 152 193 bp in size with GC content of 37.25%, encoding a total of 131 genes, and 86 of them were protein coding genes, which preferred using codons ending in A/U. The collinearity and sliding window analysis showed that the chloroplast genomes of *Chenopodium* genus plants exhibited high sequence similarity, and the nucleotide polymorphism sites were mainly found in the intergenic regions of LSC and SSC regions, as well as within the *ndhF* and *ycf1* genes. The SSRs type in *Chenopodium* plants chloroplast genomes were mainly single nucleotide repeats, and the main component base were A/T. The types of repetitive sequences were mostly forward repeats and palindrome repeats. The phylogenetic analysis showed that *C. amaranticolor* was closely related to *C. formosanum*. The genetic distance between *C. amaranticolor* and *C. quinoa* based on the nuclear genes is greater than that of chloroplast genes. The results of this study could provide references for further study on genetic structure and diversity, specific gene mining, and phylogenetic studies of *C. spp.*

**Key words:** *Chenopodium amaranticolor*; chloroplast genome; high-throughput sequencing; phylogeny

收稿日期: 2024-07-23; 修回日期: 2024-09-09

基金项目: 陕西省重点研发计划项目 (2020NY-138); 陕西省教育厅重点科研项目 (20JS003) 资助

作者简介: 孔卫青 (1980-), 女, 汉族, 山东菏泽人, 博士, 研究员, 主要从事植物病害与资源开发利用研究, E-mail: 275455929@qq.com; \*通信作者 Author for correspondence, E-mail: yangjinhong@aku.edu.cn

藜属 (*Chenopodium*) 植物为一年生草本植物, 全世界约有 250 种。藜属植物富含蛋白质、矿物质、碳水化合物等营养物质, 籽粒可用作粮食, 其中藜麦 (*C. quinoa*) 曾被南美洲土著居民作为主粮, 且目前仍被作为一种全营养食物在全世界范围内种植<sup>[1]</sup>。藜属植物具有良好的抗逆性<sup>[2]</sup>, 是主要的田间杂草<sup>[3-4]</sup>, 如尖头叶藜 (*C. acuminatum*)、藜 (*C. album*)、小藜 (*C. ficifolium*)、麻籽藜 (*C. berlandieri*), 或作为饲用资源开发利用<sup>[5-6]</sup>。苋色藜 (*Chenopodium amaranticolor*) 是藜属植物中分布较少的一个种, 其天然红色素具有一定的稳定性, 是化工和食品原料, 在部分地区也作为蔬菜食用和观赏<sup>[7]</sup>。苋色藜还具有广谱的病毒抗性, 是植物病毒学研究中广泛应用的一种枯斑寄主指示植物, 能够繁殖双生病毒科、豇豆花叶病毒科、杆状病毒科等多种植物病毒<sup>[8]</sup>。

叶绿体具有独立于细胞核的基因组, 可翻译合成自身的蛋白质, 是一种半自主细胞器。叶绿体基因组通常为 120 000~170 000 bp 的环状 DNA 分子, 自 1986 年烟草<sup>[9]</sup>和地钱<sup>[10]</sup>的叶绿体 DNA 被测序以来, 大量植物和藻类的叶绿体 DNA 被测序分析, 其结构具有较高的保守性, 一般由大单拷贝区域 (Large single-copy, LSC), 小单拷贝区域 (Small single-copy, SSC), 以及两个反向重复区域 (Inverted repeat a/b, IR a/b) 等 4 个区域构成, 广泛应用于植物物种鉴定、系统发育和物种起源的研究<sup>[11]</sup>。截止 2024 年 6 月, NCBI 公布的藜属植物叶绿体基因组有藜、藜麦、台湾藜 (*C. formosanum*)、麻籽藜、尖头叶藜、菱叶藜 (*Chenopodium bryoniaefolium*)、小藜、苍白茎藜 (*Chenopodium pallidicaule*)、瑞典藜 (*Chenopodium suecicum*) 和 *Chenopodium petiolare* 在内的 10 个物种。而苋色藜作为基础性研究中广泛应用的种质资源, 其在系统发育中的位置及其与近缘物种的关系尚不明晰。

从叶绿体基因组角度探究以上问题, 本研究对苋色藜进行了二代测序, 组装分析其叶绿体基因组及特征, 并比较分析了藜属植物叶绿体基因组的结构、GC 含量、基因组成、重复序列、叶绿体基因组序列变异, 构建了系统发育树, 为进一步研究藜属植物的遗传结构、遗传多样性及亲缘性等奠定理论基础。

## 1 材料与方法

### 1.1 试验材料

实验使用苋色藜材料保存于安康学院陕西省蚕桑重点实验室。叶片经纯水清洗擦干后用锡纸包裹放入液氮速冻, 之后于-80℃冰箱保存。

### 1.2 试验方法

#### 1.2.1 苋色藜基因组 DNA 提取与高通量测序

使用植物基因组 DNA 提取试剂盒 (DP305, TIANGEN 公司, 北京) 提取苋色藜总 DNA, 琼脂糖凝胶电泳和分光光度法检测 DNA 的质量。合格样品送生工生物工程(上海)股份有限公司进行建库和 MGISEQ-200RS 高通量测序平台双端双向 150 bp 测序。

#### 1.2.2 叶绿体基因组组装和注释

fastp 软件对测定获得的原始数据进行过滤<sup>[12]</sup>, 得到去除了接头和低质量的干净数据。以 NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide>) 下载的藜属植物叶绿体基因组 NC 序列做参考 (表 1), bowtie2 软件 very-sensitive 模式筛选干净数据<sup>[13]</sup>, spade3.15 和 GeSeq 对筛选所得序列进行组装和注释<sup>[14-15]</sup>, 运用细胞器基因组绘图软件 OGDRAW 绘制物理图谱<sup>[16]</sup>, 同时提交至 GeneBank 数据库, 登录号 PP727303。

表 1 文中使用序列及特征

Table 1 The sequences and characteristics used in the article

种 Species	基因组 Genome		大单拷贝区 LSC		小单拷贝区 SSC		反向重复区 IR		登录号
	长度	GC/%	长度	GC/%	长度	GC/%	长度	GC/%	Accession
	Length/ bp		Length/ bp		Length/ bp		Length/ bp		No.
苋色藜 <i>C. amaranticolor</i>	152 193	37.25	83 675	35.31	18 132	31.01	25 193	42.72	PP727303
台湾藜 <i>C. formosanum</i>	152 195	37.25	83 677	35.30	18 130	31.01	25 195	42.72	NC_086915
<i>C. petiolare</i>	152 064	37.24	83 520	35.28	18 108	30.92	25 218	42.76	NC_081912
麻籽藜 <i>C. berlandieri</i>	152 194	37.25	83 676	35.30	18 132	31.02	25 193	42.72	NC_085145
菱叶藜 <i>C. bryoniifolium</i>	152 166	37.25	83 675	35.30	18 105	30.99	25 193	42.72	NC_085144
藜麦 <i>C. quinoa</i>	152 099	37.24	83 582	35.29	18 107	30.96	25 205	42.74	NC_034949
尖头叶藜 <i>C. acuminatum</i>	152 200	37.25	83 683	35.30	18 131	31.01	25 193	42.72	NC_054154
小藜 <i>C. ficifolium</i>	151 823	37.26	83 668	35.31	17 937	31.00	25 109	42.75	NC_041200
藜 <i>C. album</i>	152 167	37.26	83 676	35.30	18 105	30.99	25 193	42.73	NC_034950
苍白茎藜 <i>C. pallidicaule</i>	151 293	37.30	82 948	35.32	17 967	31.13	25 189	42.75	OR654979
瑞典藜 <i>C. suecicum</i>	151 865	37.25	83704	35.29	17 943	31.01	25 109	42.75	OR654989
灰绿藜 <i>Oxybasis glauca</i>	151 655	36.87	83 579	34.73	17 914	30.35	25 081	42.77	NC_047226

### 1.2.3 叶绿体基因组特征分析

MegaX 软件统计藜属植物叶绿体基因组密码子的数量、使用频率及相对同义密码子使用度 (relative synonymous codon usage, RSCU) [17], MISA 软件分析其中的简单重复序列 (simple sequence repeat, SSR) 位点 [18], 设置单碱基至六碱基的最少重复性数目为 10, 5, 4, 3, 3, 3, 两个 SSR 之间的距离  $\geq 100$  bp。使用在线 REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) 分析其中的散在重复序列。

### 1.2.4 共线性与多态性分析

Mauve 软件比对藜属植物叶绿体基因组的 GB 注释文件并分析其共线性 [19], mafft 进行全基因组序列比对 [20], DnaSP 6 软件进行序列多态性 Pi 滑动窗口比较分析 [21], 设置窗口长度 600 bp, 步长 200 bp。

### 1.2.5 系统发育分析

下载灰绿藜 (*Oxybasis glauca*) 的叶绿体基因组序列做外群, 分析藜属植物的系统发育关系。mafft 比对灰绿藜核藜属植物叶绿体基因组 [20], modeltest 获得最佳碱基替换模型为 GTR+I+G4 [22], 使用该模型和 Raxml-ng 构建最大似然 (ML) 进化树 [23], 同时使用 MrBayes 构建 BI 系统发育树 [24]。Bootstrap 法检验各分支的支持率, 重复 1 000 次。

### 1.2.6 苋色藜与藜麦单拷贝核基因和叶绿体基因的遗传距离分析

单拷贝核基因是一种分子系统学研究的分子标记, 在生命树的主干及主干和末梢之间分枝构建中非常有价值。以藜麦的单拷贝核基因集 (<https://treeoflife.kew.org/>) 作参考, easy353 组装苋色藜的单拷贝核基因, 参数设置最低  $8 \times$  覆盖度 [25], mafft 对苋色藜与藜麦的同源单拷贝核基因进行两两比对。同时从 genbank 文件中提取 CDS 序列, mafft 对二者的同源序列进行两两比对。基于 JC69 模型, 分别计算 2 个物种单拷贝核基因和叶绿体基因的遗传距离。

## 2 结果与分析

### 2.1 藜属植物叶绿体基因组结构特征

高通量测序共获得苋色藜基因组原始数据 16 788 751 条, 质控去除接头和低质量序列后, 得到干净数据约 4.82G, 16 762 445 条。以叶绿体基因组 NC 序列为参考进行筛选, 得到苋色藜的潜在叶绿体基因组来源序列 1 180 783 条, 组装覆盖度  $1 104.71 \times$ , 所得苋色藜叶绿体基因组长 152 193 bp, GC 含量 37.25%, 与其他藜属植物叶绿体基因组 151 293 (苍白茎藜) ~ 152 195 bp (台湾藜)、GC 含量 37.24% ~ 37.30% 相符 (图 1, 表 1)。11 个藜属

植物的叶绿体基因组均为典型的环状双链四分体结构，LSC 长度在 82 948 ~ 83 704 bp，GC 含量 35.28% ~ 35.32%，SSC 和 IR 的长度分别为 17 937 ~ 18 132 bp 和 25 109 ~ 25 218 bp，GC 含量 30.92% ~ 31.13% 和 42.72% ~ 42.76%。苋色藜叶绿体基因组及四分体的长度和 GC 含量均与台湾藜和麻籽藜比较接近。

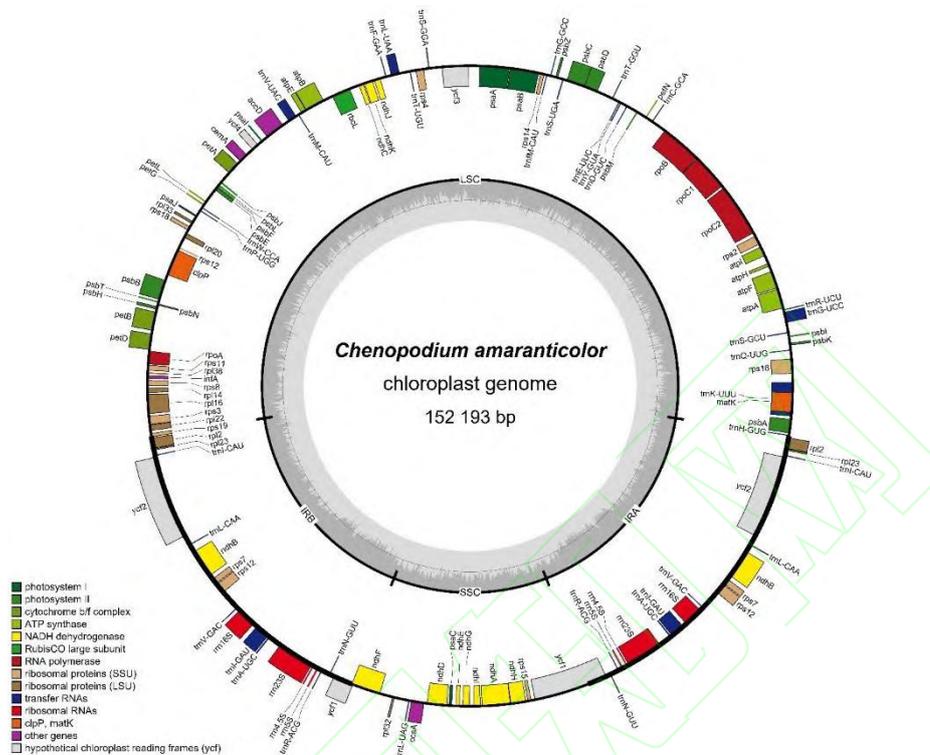


图 1 苋色藜叶绿体基因组图谱

Fig. 1 The gene map of *C. amaranticolor* chloroplast genome

## 2.2 苋色藜叶绿体基因组的注释及功能分类

分析苋色藜叶绿体基因组的编码基因，共编码 131 个基因，其中蛋白质编码基因 (Protein-coding genes, PCGs) 86 个，核糖体 RNA (Ribosomal RNA, rRNA) 8 个，转运 RNA (Transfer RNA, tRNA) 37 个 (表 2)，与其他藜属植物一致。18 个基因有 2 个拷贝，包括 7 个 PCGs (*rpl2*、*rpl23*、*rps7*、*rps12*、*ndhB*、*ycf1* 和 *ycf2*)，7 个 tRNA (*trnA-UGC*、*trnI-CAU*、*trnI-GAU*、*trnL-CAA*、*trnN-GUU*、*trnR-ACG* 和 *trnV-GAC*)，以及 4 个 rRNA (*rrn16S*、*rrn23S*、*rrn4.5S* 和 *rrn5S*)，其他基因为单拷贝。苋色藜叶绿体基因组的大部分基因没有内含子，除了 8 个 PCGs (*atpF*、*petB*、*petD*、*ndhA*、*ndhB*、*rpl16*、*rps16* 和 *rpoC1*) 和 6 个 tRNA (*trnA-UGC*、*trnG-UCC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA* 和 *trnV-UAC*) 有 1 个内含子，3 个 PCGs (*ycf3*、*clpP* 和 *rps12*) 有 2 个内含子。*matK* 基因位于 *trnK-UUU* 的内含子中。*rps12* 基因为反式剪接基因，5'端和重复 3'端分别位于 LSC 和 IR 区域 (图 1)。

表 2 苋色藜叶绿体基因组注释基因信息

Table 2 Genes annotated in *C. amaranticolor* chloroplast genome

基因类别 Gene Category	基因分组 Gene group	基因名称 Gene name
RNA 基因 RNA gene	核糖体 RNA Ribosomal RNA	<i>rrn16S<sup>#</sup></i> , <i>rrn23S<sup>#</sup></i> , <i>rrn4.5S<sup>#</sup></i> , <i>rrn5S<sup>#</sup></i>
	转运 RNA Transfer RNA	<i>trnA-UGC<sup>#</sup></i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-GCC</i> , <i>trnG-UCC<sup>#</sup></i> , <i>trnH-GUG</i> , <i>trnI-GAU<sup>#</sup></i> , <i>trnI-CAU<sup>#</sup></i> , <i>trnK-UUU<sup>#</sup></i> , <i>trnL-UAG</i> , <i>trnL-UAA<sup>#</sup></i> , <i>trnL-CAA<sup>#</sup></i> , <i>trnM-CAU</i> , <i>trnN-GUU<sup>#</sup></i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-UCU</i> , <i>trnR-ACG<sup>#</sup></i> , <i>trnS-GGA</i> , <i>trnS-GCU</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> ,

自我复制基因	核糖体小亚基	<i>trnV-UAC<sup>*</sup>, trnV-GAC<sup>z</sup>, trnW-CCA, trnY-GUA, trnM-CAU</i>	
Gene for self-replication	Small subunit of ribosome	<i>rps11, rps12<sup>**</sup>, rps14, rps15, rps16<sup>*</sup>, rps18, rps19, rps2, rps3, rps4, rps7<sup>z</sup>, rps8</i>	
	核糖体大亚基	<i>rpl14, rpl16<sup>*</sup>, rpl2<sup>z</sup>, rpl20, rpl22, rpl23<sup>z</sup>, rpl32, rpl33, rpl36</i>	
光合作用基因	RNA 聚合酶 RNA polymerase	<i>rpoA, rpoB, rpoC1<sup>*</sup>, rpoC2</i>	
	光系统 I Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>	
	Gene for Photosynthesis	光系统 II Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
		细胞色素复合物 b/f 亚基	<i>petA, petB<sup>*</sup>, petD<sup>*</sup>, petG, petL, petN</i>
	Subunit of cytochrome b/f complex		
	ATP 合酶亚基	<i>atpA, atpB, atpE, atpF<sup>*</sup>, atpH, atpI</i>	
	Subunit of ATP synthase		
	Rubisco 亚基	<i>rbcl</i>	
	Large subunit of Rubisco		
	NADH 脱氢酶亚基	<i>ndhA<sup>*</sup>, ndhB<sup>**</sup>, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	
Subunit of NADH dehydrogenase			
其他基因	成熟酶 Maturase	<i>matK</i>	
	Other gene	被膜蛋白 Envelop membrane protein	<i>cema</i>
		乙酰辅酶 A 羧化酶亚基	<i>accD</i>
	Subunit of acetyl-CoA carboxylase		
	c 型细胞色素合成基因	<i>ccsA</i>	
	c-type cytochrome synthesis gene		
	蛋白酶	<i>clpP<sup>**</sup></i>	
	Protease		
	翻译起始因子	<i>infA</i>	
	Translational initiation factor		
未知基因	保守开放阅读框	<i>ycf1<sup>z</sup>, ycf2<sup>z</sup>, ycf3<sup>**</sup>, ycf4</i>	
Unknown gene	Conservative open reading box		

注: #代表拷贝数为 2; \*表示有一个内含子; \*\*表示有 2 个内含子

Note: # represents 2 copy number; \* indicates one intron inclusive and \*\* indicates two introns inclusive

### 2.3 密码子偏好分析

分析藜属植物叶绿体 CDS 的 GC 含量在 25.93% ~ 50.00%，平均 38.02%。密码子不同位置 GC 含量差异较大，GC1 含量 31.03% ~ 65.62%，平均 45.96%，GC2 和 GC3 含量平均 39.78%和 28.34%，GC3 含量最低（图 2），表明藜属植物叶绿体基因组偏好以 A 和 U 碱基结尾。RSCU 分析显示，值大于 1 的密码子有 30 种，除 UUG（Leu）外，其它 29 个均以 A/U 碱基结尾，说明高频密码子的第三个碱基偏向 A/U（图 3）。

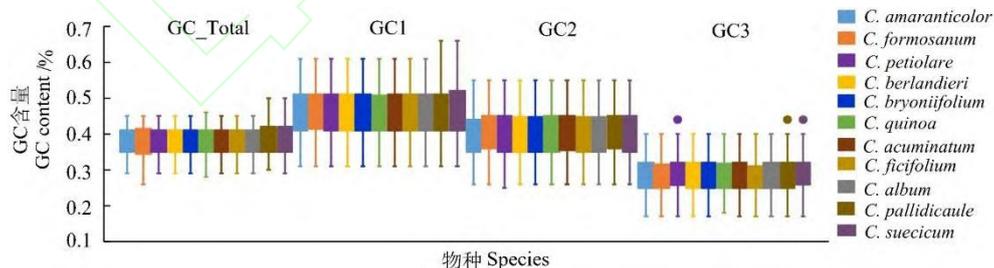


图 2 苋色藜及 10 个同属种叶绿体基因组密码子组成成分分析

Fig. 2 Composition analysis of chloroplast genome codons of *C. amaranticolor* and its 10 congeneric species

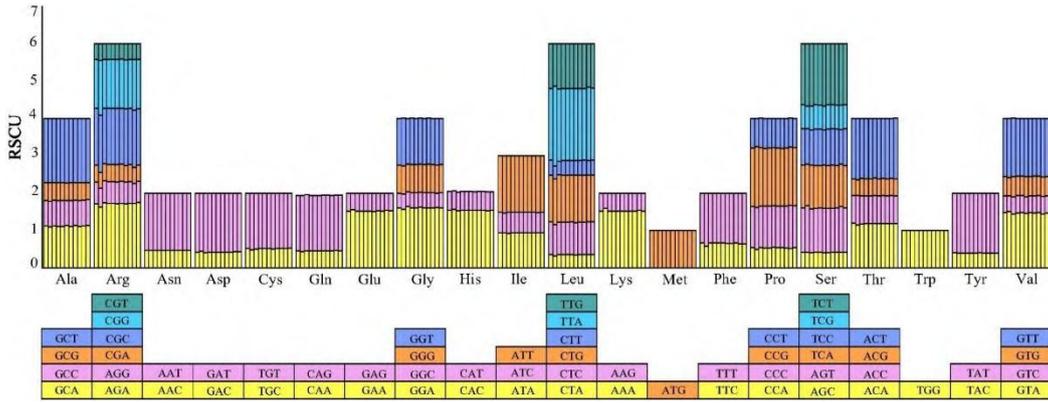


图3 苋色藜及10个同属种叶绿体基因组密码子RSCU分析

Fig.3 The RSCU of chloroplast genome codons of *C. amaranticolor* and its 10 congeneric species

### 2.4 重复序列分析

MISA分析藜属植物叶绿体基因组中的SSR，结果显示所有种中均只有单碱基至五碱基重复，数量从44~54不等（图4）。单碱基重复的数目最多，占比最低的为小藜57.45%。其次为二碱基重复7~12个以及四碱基重复5~9个。三碱基和五碱基重复较少，共1~3个。从碱基组成看，除苍白茎藜有(C)<sub>11</sub>和(G)<sub>11</sub>各1个外，其他的单碱基重复和所有的三碱基以及五碱基重复全部由A、T碱基组成。二碱基和四碱基重复中，所有物种都含有2个(CG)<sub>5</sub>和各1个(AGGT)<sub>3</sub>、(CTAC)<sub>3</sub>，位于IR区的两个*rrn23S*基因中，其他重复以AT碱基为主。分析SSRs在苋色藜叶绿体基因组四分体结构的分布，LSC有36个，SSC有9个，IRa和IRb各4个。从基因的编码区和非编码区的角度来看，32个位于基因间隔区，7个位于内含子，14个位于外显子。

Reputer分析其中的散在重复，数量在50~72个之间，几乎全部为正向重复（Forward repeat, F）和回文重复（Palindrome repeat, P）。重复的长度以30~39 bp为主，其次为40~49 bp长度，绝大部分物种中二者占比多于60%（图5）。

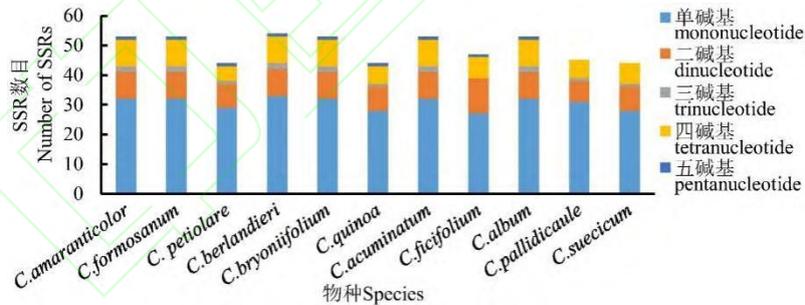


图4 苋色藜及10个同属种叶绿体基因组SSR分析

Fig.4 SSR in chloroplast genome of *C. amaranticolor* and its 10 congeneric species

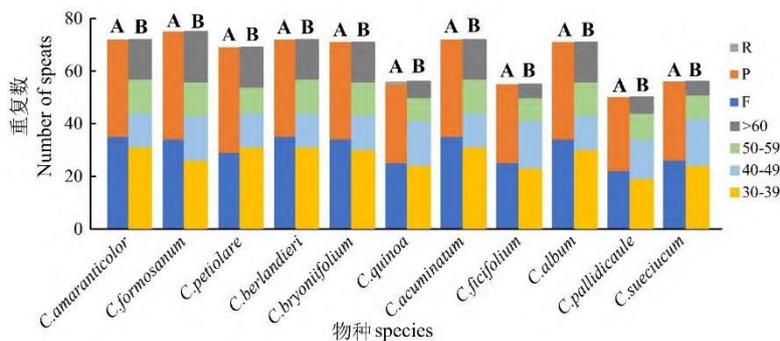


图5 苋色藜及10个同属种叶绿体基因组重复序列的类型(A)和长度(B)分析

Fig.5 The type (A) and length (B) of repeat sequences in chloroplast genome of *C. amaranticolor* and its 10 congeneric species

## 2.5 藜属植物叶绿体基因组的共线性与核苷酸多样性分析

使用 Mauve 软件进行多重基因组比对分析 11 个藜属植物和市藜属植物灰绿藜的叶绿体基因组中的重排和共线性。结果显示，藜属植物的叶绿体基因组有 5 个局部共线区域，序列具有较高的相似性，而灰绿藜的 LSC 区 *rbcl-trnV-UAC* 间的序列发生了倒置（图 6）。从叶绿体基因组的 4 个组分看，IR 区序列的变异较低，LSC 和 SSC 区的变异程度较高。

滑动窗口分析 11 个藜属植物叶绿体基因组的核苷酸多样性指数 Pi 值在 0 ~ 0.0244，平均 0.0057，从四分体结构看，SSC 区 0.0093 > LSC 区 0.0062 > IR 区 0.0012。Pi 最高值 0.0244 在 *trnT-UGU-trnF-GAA* 处，大于 0.015 的区域还有 LSC 区的基因间隔区 *matk-rps16*、*psbI-trnS-GCU*、*psaA-ycf3*、*ycf3-trnS-GGA*、*rps4-trnT-UGU*、*trnT-UGU-trnF-GAA*、*psbE-petL*、*psaJ-rpl33* 和 *rpl33-rps18* 和 SSC 区的基因间隔区 *ndhF-rpl32* 和 *rpl32-trnL-UAG*，以及基因 *ndhF* 和 *ycf1* 内部（图 7）。



图 6 12 个物种叶绿体基因组的共线性分析

Fig.6 Collinearity analysis of the chloroplast genome of 12 species

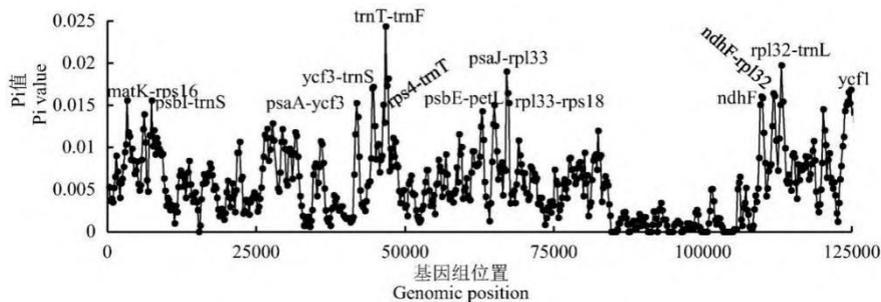


图 7 苋色藜及 10 个同属种叶绿体基因组 Pi 的多样性指数

Fig.7 Pi diversity index analysis of chloroplast genomes of *C. amaranticolor* and its 10 congeneric species

## 2.6 系统发育

以灰绿藜为外群分析藜属植物的进化关系，分别构建最大似然法（ML）和贝叶斯方法（BI）系统发育树，结果两个进化树拓扑结构基本一致（图 8），且 ML 进化树分支上支持值均大于 50%，BI 系统发育树除苋色藜和台湾藜的聚合位点外，也都大于 50%，说明分析结果较为可靠。两个进化树中，苋色藜均与台湾藜聚合，与麻籽藜和尖头叶藜、藜和菱叶藜聚合的分支互为姊妹群。小藜和瑞典藜聚合在一个分支，藜属重要的粮食作物藜麦和苍白茎藜以及 *C. petiolare* 一起位于进化树的基部，与传统的进化分类一致。

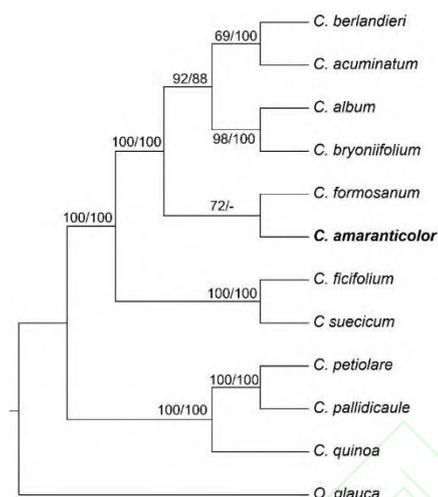


图 8 藜属植物的系统发育树，斜线前后数值分别为 ML 和 BI 方法支持值

Fig.8 Phylogenetic tree of *Chenopodium* plants. The numbers before and after the diagonal indicate the probability of ML and BI methods, respectively.

## 2.7 藜麦和苋色藜单拷贝核基因和叶绿体基因的遗传距离

以藜麦为参照，easy353 组装获得符合条件的苋色藜单拷贝核基因 64 个，分别计算二者所有单拷贝核基因的遗传距离在 0 ~ 0.738 之间，其中小于 0.1 和 0.1 ~ 0.2 的基因数分别 27 和 22 个，二者占比 76.56%（图 9）。同时计算二者叶绿体基因的遗传距离最大值 0.023，大部分基因（59 个，占比 76%）的遗传距离小于 0.01，说明核基因的遗传距离大于叶绿体基因。

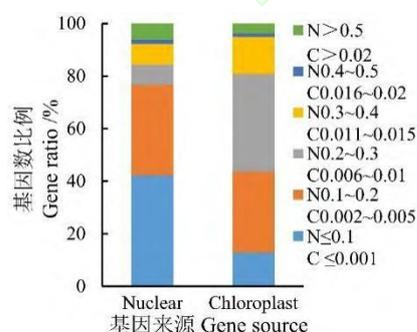


图 9 苋色藜和藜麦的单拷贝核基因（N）和叶绿体基因（C）的遗传距离

Fig.9 The genetic distances between single copy nuclear genes (N) and chloroplast genes (C) of *C. amaranticolor* and *C. quinoa*

### 3 讨论

苋色藜在野外的种群密度非常低,常被用作病毒指示植物,其遗传信息十分匮乏,也缺乏有效的基因标签。本研究通过 MGISEQ-200RS 高通量测序技术获得了苋色藜叶绿体基因组为共价闭环环状双链分子,由 LSC、SSC 和两个 IR 区组成,序列长 152 193 bp, GC 含量 37.25%,注释蛋白编码基因 84 个,与目前公布的同属植物叶绿体基因组特征较为一致。

碱基突变和自然选择是影响密码子偏好的重要因素,其中,植物线粒体和叶绿体基因组密码子的偏好主要受自然选择的影响<sup>[26]</sup>。本研究分析藜属不同种植物的叶绿体基因组密码子不同位置的 GC 含量都是 GC1>GC2>GC3,说明该属植物密码子第 3 位偏好使用 A、U 结尾,即为 NNA 或 NNU 型,这一特征与其它绿色植物叶绿体基因组的特征一致。RSCU 分析藜属植物叶绿体 20 种氨基酸均有密码子显示出使用偏好 (RSCU>1),说明开展藜属植物叶绿体基因工程研究时,最优密码子应选择 A/U 碱基结尾,能够提高外源基因表达水平。

基因组中的重复序列,一般具有很好的重复性和丰富的多态性,在种属遗传多样性研究和种质资源鉴定等方面具有重要应用。本研究藜属植物的 SSR 分布在 LSC、SSC 和 IR 区,单核苷酸重复占比最高,IR 区较为保守,LSC 区和 SSC 区的具有较高的多态性,可开发为藜属内物种分子鉴定的标签。共线性分析获得藜属植物叶绿体基因组结构较保守,而市藜属灰绿藜 LSC 区 *rbcl-trnV-UAC* 间的序列倒置,这种情况也发生在泽泻目植物的部分科<sup>[27]</sup>、豆科苜蓿<sup>[28]</sup>等中。滑动窗口分析核苷酸多样性 Pi 指数获得具有高多态性的区域,多态性位点主要在基因间隔区和少量基因的内部,可作为藜属植物种质资源鉴定研究的候选。系统发育关系研究苋色藜与台湾藜的亲缘关系最近,同时二者也具有相似的长度和碱基组成,这些均为利用叶绿体鉴定其种质提供了数据支持。

大多数被子植物都经历了全基因组复制或加倍化,甚至大规模的重复序列,这导致植物基因组的特征复杂,也为基于核 DNA 序列的遗传分析和种质鉴定带来一定的困难<sup>[29]</sup>。同时基因组中也存在拷贝数较少,只有 1 个或几个但获取比较困难的单拷贝核基因。该类基因大多为组成性表达持家基因,能更准确进行系统发育的重建和估算。目前藜麦的单拷贝核基因已被鉴定,并被上传被子植物单拷贝核基因网站。我们以其为参考,鉴定获得了 64 个苋色藜的核单拷贝核基因,这为从核的角度鉴定苋色藜提供了依据。同时我们计算了苋色藜和藜麦核单拷贝核基因的遗传距离最高 0.738,大部分为 0~0.2,而大部分叶绿体基因的遗传距离小于 0.01,这与植物叶绿体基因进化速度快于线粒体,而慢于核基因的结论一致<sup>[30]</sup>。

### 4 结论

苋色藜叶绿体基因组全长 152 193 bp,为典型的四分体结构,GC 含量 37.25%,共编码 131 个基因,其中蛋白质编码基因 84 个,密码子偏好以 A/U 结尾。SSR 位点分布不均衡,单核苷酸重复为主要组成。共线性分析苋色藜没有基因重排现象,与其他物种的差异主要在基因间隔区。在系统发育上,苋色藜与台湾藜的亲缘关系最近,与藜属内其他植物的遗传距离也较近,与传统分类学一致。基于苋色藜与藜麦叶绿体基因的遗传距离小于核基因,这说明叶绿体基因在研究种属的分类中具有优势。

### 参考文献

- [1] LI L K, LIETZ G, SEAL C J. Phenolic, apparent antioxidant and nutritional composition of quinoa (*Chenopodium quinoa* Willd.) seeds[J]. International Journal of Food Science and Technology,2021,56(7):3245-3254
- [2] 王昶,杨发荣,李敏权,等. 藜麦种质资源对霜霉病的抗性鉴定与评价[J]. 草地学报,2022,30(10):2626-2634
- [3] 郭家驹,胡梦琪,张琳晗,等. 中国北方地区四种熟型藜属杂草始花期的模拟[J]. 草地学报,2024,32(6):1872-1881

- [4] 刘勇,张淼,魏亚东,等. 我国外来杂草新纪录种—麻籽藜[J]. 植物检疫,2023,37(3):62-64
- [5] 王一博,陈奕彤,刘馨迪,等. 饲用藜麦在黄河三角洲盐碱地的适应性研究与评价[J]. 草地学报,2023,31(6):1810-1817
- [6] 路平乐,魏玉明,杨发荣,等. 体外产气法评价全株藜麦与全株玉米混合青贮饲料饲用价值[J]. 草地学报,2021,29(11):2391-2398
- [7] 赵忠,俞剑燊,孙微,等. 苜蓿细胞培养产生的天然红色素性质及其稳定性[J]. 无锡轻工大学学报,2000,19(1):38-40
- [8] 龚前园. 苜蓿 NDR1 基因的克隆与抗病毒功能研究[D]. 北京:中国农业科学院,2014:1
- [9] SHINOZAKI K, OHME M, TANAKA M, et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression[J]. EMBO Journal, 1986,5(9):2043-2049
- [10] OHYAMA K, FUKUZAWA H, KOHCHI T, et al. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA[J]. Nature,1986(322),572-574
- [11] DANIELL H, LIN C S, YU M, et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering[J]. Genome Biology,2016,17(1):134
- [12] CHEN S, ZHOU Y, CHEN Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor[J]. Bioinformatics, 2018,34(17):i884-i890
- [13] LANGMEAD B, SALZBERG S L. Fast gapped-read alignment with Bowtie 2[J]. Nature Methods,2012,9(4): 357-359
- [14] BANKEVICH A, NURK S, ANTIPOV D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing[J]. Journal of Computational Biology,2012,19(5):455-477
- [15] TILLICH M, LEHWARK P, PELLIZZER T, et al. GeSeq - versatile and accurate annotation of organelle genomes[J]. Nucleic Acids Research,2017,45(W1):W6-W11
- [16] GREINER S, LEHWARK P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes[J]. Nucleic Acids Research,2019,47(W1):W59-W64
- [17] KUMAR S, STECHER G, LI M, et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms[J]. Molecular Biology and Evolution,2018,35(6):1547-1549
- [18] BEIER S, THIEL T, MÜNCH T, et al. MISA-web: a web server for microsatellite prediction[J]. Bioinformatics,2017,33(16):2583-2585
- [19] DARLING A C, MAU B, BLATTNER F R, et al. Mauve: multiple alignment of conserved genomic sequence with rearrangements[J]. Genome Research,2004,14(7):1394-1403
- [20] NAKAMURA T, YAMADA K D, TOMII K, et al. Parallelization of MAFFT for large-scale multiple sequence alignments[J]. Bioinformatics,2018,34(14):2490-2492
- [21] ROZAS J, FERRER-MATA A, SÁNCHEZ-DELBARRIO J C, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets[J]. Molecular Biology and Evolution,2017,34(12):3299-3302
- [22] DARRIBA D, POSADA D, KOZLOV A M, et al. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models[J]. Molecular Biology and Evolution,2020,37(1):291-294
- [23] KOZLOV A M, DARRIBA D, FLOURI T, et al. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference[J]. Bioinformatics,2019,35(21):4453-4455
- [24] HUELSENBECK J P, RONQUIST F. MRBAYES: Bayesian inference of phylogenetic trees[J]. Bioinformatics,2001,17(8):754-755
- [25] ZHANG Z, XIE P, GUO Y, et al. Easy353: A Tool to Get Angiosperms 353 Genes for Phylogenomic Research[J]. Molecular Biology and Evolution,2022,39(12):msac261
- [26] LI G, ZHANG L, XUE P. Codon usage pattern and genetic diversity in chloroplast genomes of *Panicum* species[J]. Gene,2021,802:145866

- [27] MWANZIA M V. 泽泻目植物叶绿体基因组结构重排及大小变异研究[D]. 北京:中国科学院大学,2020:1
- [28] 郝新艳,赵淑文,刘嘉伟,等. 杂花苜蓿叶绿体基因组特征及系统发育分析[J]. 草地学报,2023,31(6):1665-1672
- [29] SLIMP M, WILLIAMS L D, HALE H, et al. On the potential of Angiosperms353 for population genomic studies[J]. *Applications in Plant Sciences*,2021,9(7):e11419
- [30] DROUIN G, DAOUD H, XIA J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants[J]. *Molecular Phylogenetics and Evolution*, 2008,49(3):827-831

