

## 红花不同花色变异类型叶绿体基因组特征与系统进化分析

丁怡宁, 毕光耀, 胡赛文, 周 艳, 李贺敏\*, 夏 至\*

河南农业大学农学院, 河南 郑州 450046

**摘要:**目的 以红花 *Carthamus tinctorius* 不同花色变异类型为材料, 比较分析其叶绿体基因组结构及其与近缘类群的系统发育关系, 为中药材红花种质资源鉴定和群体遗传学研究奠定基础。方法 利用华大 MGISEQ-2000PE150 测序平台, 双末端测序策略对红花不同花色类型全基因组 DNA 建库测序, 用 NOVOPlasty 组装叶绿体基因组, 采用最大似然法 (maximum Likelihood, ML) 构建系统进化树。结果 红花 3 种花色变异类型叶绿体基因组全长均为 153 114 bp, 完全一致, GC 值均为 37.8%, 具 1 个典型的四分区域结构, 包括 1 个大单拷贝区 (large single copy region, LSC)、1 对反向重复区 (inverted repeats, IR) 和 1 个短单拷贝区 (small single copy, SSC), 4 个区序列长度分别为 84 128、25 194、18 598 bp。红花的叶绿体基因组均有 130 个基因, 其中编码蛋白基因、rRNA 基因与 tRNA 基因的数量均为 84、8、38 个。系统进化分析表明, 红花 3 种花色变异类型聚在一支, 与菜蓟族的铺散矢车菊构成一单系分支, 具有 100% 支持率。结论 叶绿体基因组数据支持红花 3 种花色变异类型来源为同一种, 药用植物红花 (所在红花属) 隶属于菊科菜蓟族的矢车菊亚族。

**关键词:** 红花; 叶绿体基因组; 组装; 系统发育; 近缘类群

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2023)01-0262-10

DOI: 10.7501/j.issn.0253-2670.2023.01.028

## Characterization and phylogenetic analysis of complete chloroplast genome of different color medicinal plant *Carthamus tinctorius*

DING Yi-ning, BI Guang-yao, HU Sai-wen, ZHOU Yan, ZHOU Yan, LI He-min, XIA Zhi

College of Agronomy, Henan Agricultural University, Zhengzhou 450046, China

**Abstract: Objective** In order to provide foundation for *Carthamus tinctorius* molecular identification and population genetics studies, we sequenced and assembled the complete chloroplast genome of three flower color variants *C. tinctorius*, and constructed the phylogenetic relationship between it and its relatives. **Methods** Genome DNA library was constructed with the paired-end strategy, and MGISEQ-2000PE150 platform was used to sequence in Beijing Genomics Institute (China). The complete chloroplast genome was assembled using NOVOPlasty software, and sequence analysis was performed based on gene annotation results. Phylogenetic analyses were performed using Maximum-Likelihood (ML) methods. **Results** The complete chloroplast genome of three different colors *C. tinctorius* has same length 153 144 bp with a GC content of 37.8%. The chloroplast genome exhibited a typical quadripartite structure, including a large single copy region (LSC), a pair of inverted repeats (IR), and a small single copy (SSC), and the sequence lengths were 84 128 bp, 25 194 bp, and 18 598 bp. The chloroplast genome harbored 130 genes, including 84 protein-coding genes, 8 rRNA genes, and 38 tRNA genes. Phylogenetic analyses result indicates that the three flower color variants *C. tinctorius* form one clade, and this clade is sister to *Centaurea diffusa* with bootstrap 100%. **Conclusion** Our result supported the three flower color variants of *C. tinctorius* from the same species. The medicinal plant *C. tinctorius* (genus *Carthamus*) belongs to Centaureinae of Cynareae in Asteraceae.

**Key words:** *Carthamus tinctorius* L.; chloroplast genome; assembly; phylogenetic analysis

收稿日期: 2022-07-06

基金项目: 河南省科技攻关项目 (222102110137); 河南省高等学校重点科研项目计划 (22A360010); 国家自然科学基金项目 (U1404302)

作者简介: 丁怡宁 (1997—), 女, 河南平顶山人, 硕士研究生, 研究方向为中药资源的分子鉴定。

\*通信作者: 李贺敏, 女, 副教授, 主要从事中药资源的分类鉴定, E-mail: lihemin2002@henau.edu.cn

夏 至, 教授, 主要从事中药资源的分子鉴定及分子生药学研究。E-mail: xiazhiemail@126.com

药用植物红花 *Carthamus tinctorius* L., 隶属于菊科 (Asteraceae) 红花属 *Carthamus* L., 为一年生草本<sup>[1]</sup>。红花原产中亚地区, 引种后在我国广为栽培, 以河南和新疆栽培面积大, 所产红花质量好。《中国药典》2020 年版规定红花有效成分为羟基红花黄色素 A<sup>[2]</sup>。近年来, 化学成分研究表明红花主要含黄酮类物质、脂肪酸、色素、酚酸、挥发油等多种活性成分<sup>[3]</sup>, 具有抗炎、抗肿瘤、抗心血管疾病等药理作用, 临床应用十分广泛<sup>[4]</sup>。药用植物红花在栽培过程中, 常出现花色分化现象, 除常见的橙红色花, 还有部分浅黄色和白色花。这些花色变异的类型是种内变异还是来源于同属不同物种, 需要进一步的形态和分子验证。目前, 药用植物红花的叶绿体基因组虽然已经报道<sup>[5]</sup>, 但是对于红花不同花色变异类型的叶绿体基因组组装、基因组特征和系统进化分析等未见报道。

叶绿体是植物最重要的细胞器之一, 具有一整套用于光合作用、能量代谢、蛋白质合成及氮、硫同化相关的基因, 分布在大小为 120~180 kb 的环状基因组上, 具有结构保守, 母性遗传等特点<sup>[6-7]</sup>。叶绿体基因组一般为闭环双链 DNA 结构, 陆生植物的叶绿体基因组结构通常由 1 个大单拷贝区域 (large single copy, LSC)、1 个短单拷贝区域 (small single copy, SSC) 和 2 个反向重复区域 (inverted repeat, IR) 组成<sup>[8]</sup>。叶绿体基因组拥有相对独立的基因组和遗传序列, 并且不像核基因组一般有着复杂的重复序列, 其基因序列保守, 间隔区变异位点丰富, 适宜的进化速率能够为植物不同等级的亲缘关系<sup>[9]</sup>, 系统进化关系及遗传多样性研究提供较为可靠的信息<sup>[10]</sup>。

近年来, 叶绿体全基因组测序研究的快速发展为药用植物的分子鉴定研究提供了新的平台及思路<sup>[11]</sup>。随着测序技术的不断改进, 测序平台的不断升级, 一系列组装和注释软件如 plasmid SPAdes<sup>[12]</sup>、NOVOPlasty<sup>[13]</sup>、GetOrganelle<sup>[14]</sup>等的开发与更新, 多种重要的药用植物叶绿体基因组已完成测序和分析, 如人参 *Panax ginseng* C. A. Meyer<sup>[15]</sup>、厚朴 *Houpoea officinalis* (Rehder & E. H. Wilson) N. H. Xia & C. Y. Wu<sup>[16]</sup>、红豆杉 *Taxus wallichiana* var. *chinensis* (Pilger) Florin<sup>[17]</sup>、肉苁蓉 *Cistanche deserticola* Ma<sup>[18]</sup>、三七 *Panax notoginseng* (Burkill) F. H. Chen ex C. Chow & W. G. Huang<sup>[19]</sup>、野菊 *Chrysanthemum indicum* Linnaeus<sup>[20]</sup>、石斛 *Dendrobium nobile* Lindl.<sup>[21]</sup>、地黄

*Rehmannia glutinosa* (Gaert.) Libosch. ex Fisch. et Mey.<sup>[22]</sup>、射干 *Belamcanda chinensis* (L.) Redouté<sup>[23]</sup>等的叶绿体全基因组序列的分析已有相关报道。本研究以红花不同花色变异类型为材料, 利用高通量测序方法测定其基因组 DNA 序列, 并对该植物的叶绿体基因组进行组装和注释。分析药用植物红花 3 种花色类型叶绿体基因组序列特征, IR 边界特征, 间隔区信息位点的特征, 并对红花及其近缘物种共 18 种 (21 个样品) 植物的叶绿体基因组序列进行系统发育分析, 验证其在科级系统发育中的位置, 为药用植物红花的种质资源的鉴定、开发和利用提供一定的理论依据。

## 1 材料

红花 *C. tinctorius* L. 3 种颜色 (橙红色、浅黄色、白色) 的新鲜叶片采集于河南省新乡市原阳县小吴庄村河南农业大学科教园区 (35°6'32"N, 113°56'34"E), 凭证标本号分别为 XZ-2020-30、XZ-2020-29、XZ-2020-28。叶片装入取样袋后带回实验室, 用无菌水冲洗数次, 晾干后置于 -80 °C 冰箱备用。样品由河南农业大学农学院中药材系高致明教授和夏至副教授鉴定为菊科红花属的红花。凭证标本保存于河南农业大学标本馆, 红花及其近缘物种叶绿体基因组序列来源于 NCBI 数据库, 实验材料详细信息见表 1。

## 2 方法

### 2.1 DNA 的提取和高通量测序

采用北京天根生化植物 DNA 提取试剂盒 (Tiangen Biotech Co., 中国) 提取样品红花新鲜叶片的总 DNA, 利用 1% 的琼脂糖凝胶电泳检测 DNA 完整性。样品送至华大生物科技公司 (北京) 后, 进行 NanoDrop2000 微量分光光度计 (Thermo Scientific, 美国) 检测总 DNA 的纯度和浓度。MGISEQ-2000 PE150 测序平台测序, 测序完成后, 利用华大自主开发的过滤软件 SOAPnuke 过滤参数, 过滤步骤: (1) 过滤接头: 测序 read 匹配上 adapter 序列的 25% 或者以上则删除整条 read; (2) 过滤低质量数据: 如果测序 read 中质量值低于 20 的碱基占整条 read 的 30% 或者以上则删除整条 read; (3) 去 N: 如果测序 read 中 N 含量占整条 read 的 1% 或者以上, 则删除整条 read; (4) 获得 Clean reads。数据以 FASTQ 格式储存, 用于后续的拼接和注释。

### 2.2 叶绿体基因组的拼接和注释

叶绿体基因组拼接采用 NOVOPlasty<sup>[13]</sup>程序, 插

表 1 植物样品来源

Table 1 Plant samples used in this study

物种 (取样的标本号)	来源	GenBank 登录号
下田菊 <i>Adenostemma Lavenia</i>	GenBank	MW583043
霍香蓟 <i>Ageratum conyzoides</i>	GenBank	MK905238
豚草 <i>Ambrosia artemisiifolia</i>	GenBank	MG019037
三脉紫菀 <i>Aster ageratoides</i>	GenBank	MW813970
紫菀 <i>A. tataricus</i>	GenBank	MH669275
红花 <i>Carthamus tinctorius</i>	GenBank	KM207677
红花 (白) XZ-2021-28	河南原阳	MZ779038
红花 (浅黄) XZ-2021-29	河南原阳	MZ779039
红花 (橙红) XZ-2021-30	河南原阳	MZ779040
铺散矢车菊 <i>Centaurea diffusa</i>	GenBank	KJ690264
野菊 <i>Chrysanthemum indicum</i>	GenBank	MH165290
<i>C. x morifolium cultivar</i>	GenBank	MT919691
菊花 <i>C. x morifolium</i>	GenBank	JQ362483
向日葵 <i>Helianthus annuus</i>	GenBank	NC_007977
狭苞橐吾 <i>Ligularia intermedia</i>	GenBank	MF539930
全缘橐吾 <i>L. mongolica</i>	GenBank	MF539932
离舌橐吾 <i>L. veitchiana</i>	GenBank	MF539933
阔苞菊 <i>Pluchea indica</i>	GenBank	MG452144
花梗阔苞菊 <i>P. pteropoda</i>	GenBank	MW554520
假臭草 <i>Praxelis clematidea</i>	GenBank	KF922320
钻叶紫菀 <i>Symphyotrichum subulatum</i>	GenBank	MN541093
假还阳参 <i>Crepidiastrum lanceolatum</i>	GenBank	MK358413
蒲公英 <i>Taraxacum mongolicum</i>	GenBank	KU736961

入片段大小设为 150 bp。过滤后的 reads 用 Geneious 11.0.3<sup>[24]</sup> 拼接软件组装成重叠群, 并对组装中的简并碱基, 进行人工修正。利用 Geneq-Annotation of Organellar (<https://chlorobox.mpimgolm.mpg.de/geseq.html>), 结合 NCBI 上已报道的红花 (GenBank 登录号: KM207677) 注释结果对红花的 3 个样品叶绿体全基因组进行基因注释, 参数为默认值, 最后进行手动调整。tRNA 用 ANAGORNV1.2.38 (<https://chlorobox.mpimgolm.mpg.de/geseq.html>) 预测。注释完成后, 提交到 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/genbank/>), GenBank 登录号分别为 MZ779040 (橙红色, XZ2130), MZ779039 (浅黄色, XZ2129), MZ779038 (白色, XZ2128)。利用在线工具 OGDRAW-DRAW Organelle Genome Maps (<https://chlorobox.mpimgolm.mpg.de/OGDraw.html>) 绘制叶绿体结构图。

### 2.3 叶绿体基因组 IR 边界的收缩和扩张分析

IR 区域在叶绿体基因中具有高度保守性, IR 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>[25]</sup>。在植物进化过程中, IR/SC 边界不同程度的扩张和收缩是导致边界和基因组长度多样性的原因<sup>[26]</sup>。本研究使用 Geneious 11.0.3<sup>[22]</sup> 软件获得红花及菊科其他 11 种植物叶绿体基因组的 IRA/IRB、LSC 和 SSC 和边界基因的序列长度, 进行比较分析, 探讨菊科植物叶绿体基因组 IR 边界的收缩和扩张特征。并使用 Adobe illustrator 软件绘制菊科管状花亚科 12 个属植物叶绿体基因组 IR 边界对比图。

### 2.4 叶绿体基因组基因间隔区信息位点分析

相比叶绿体基因编码区, 叶绿体基因间隔区在近缘物种间往往具有更高的变异位点, 常被用来构建属间、属内、种间物种系统进化发育关系<sup>[27]</sup>。本研究基于红花及其近缘物种共 18 种 (21 个样品) 植物的叶绿体基因组序列特征, 利用 Phylosuite v1.2.1<sup>[28]</sup> 提取红花及其近缘物种共 18 种 (21 个样品) 植物的叶绿体基因组共有的间隔区。统计这些间隔区的信息位点百分率, 为下一步构建菊科属级和属内种间物种系统进化关系提供分子标记。

## 2.5 系统发育分析

基于菊科(Asteraceae)管状花亚科(Carduoideae)红花及其近缘物种共18种(21个样品)叶绿体基因组数据(表1),同时以菊科(Asteraceae)舌状花亚科(Cichorioideae)假还阳参和蒲公英为外类群,利用phylosuite v1.2.1<sup>[26]</sup>软件基于MAFFT<sup>[29]</sup>进行多重比对。系统发育分析采用最大似然法(maximum Likelihood, ML),利用CIPRES Science Gateway服务器(<http://www.phylo.org/>)中RaxML-HPC2 8.2.12软件<sup>[30]</sup>,采用GTR+GAMMA+I建树模型,构建系统发育树。利用Bootstrap(BS)(1000次重复)检验各分支的支持率。



图1 红花(3种花色)不同开花时期的花瓣

Fig. 1 Three color petals at different flowering stages of *C. tinctorius*

等,序列长度分别为84 128、25 194、18 598 bp。3种花色的红花在LSC区域、SSC区域和IRs区域中GC含量也完全一致,分别为36.0%、31.4%、43.2%。

## 3.2 红花3种花色类型叶绿体基因组的组成和特点分析

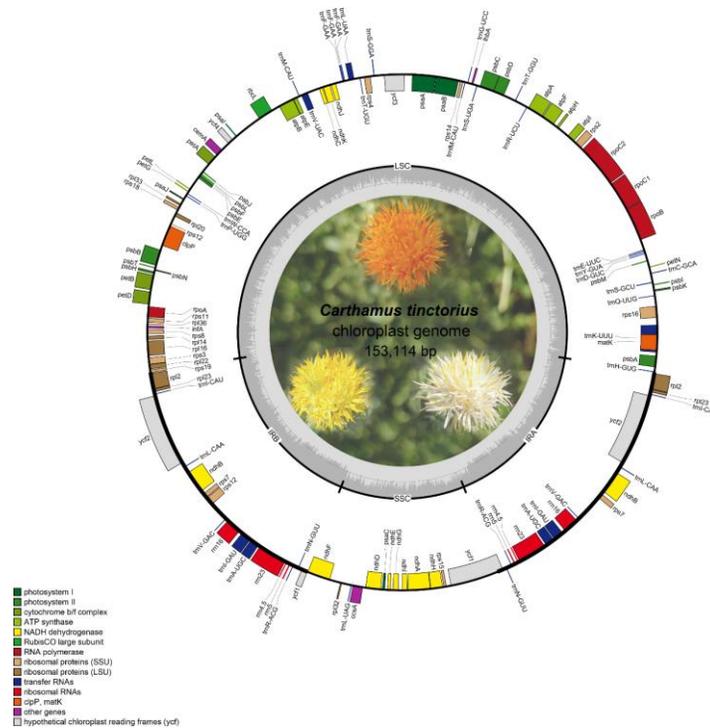
红花3种花色类型的叶绿体基因组完全一致,均包括130个基因,非重复基因115个,其中84个编码蛋白基因(PCG)、8个rRNA基因与38个tRNA基因。其中,蛋白质编码基因中与自我复制相关基因除rRNA基因和tRNA基因外,还包括13个核糖体小亚基基因、11个核糖体大亚基基因和4个RNA聚合酶亚基基因;光合作用相关的基因有44个,包括12个NADH脱氢酶基因、5个光合系统I基因、14个光合系统II基因、6个细胞色素复合体编码基因、6个ATP合酶基因、1个二磷酸核酮糖羧化酶大亚基基因;此外还有5个其他功能基

## 3 结果与分析

### 3.1 红花3种花色类型的叶绿体基因组结构特征

橙红色、浅黄色、白色这3种花色的红花,花期一致,植株形态无明显差异,随花期的变化,其颜色一直保持稳定,并且花色差异明显(图1)。测序结果去除接头和高质量的数据,组装和注释后得到橙红色、浅黄色、白色红花的叶绿体基因组。结果表明,红花3种花色类型的叶绿体全基因组均为共价闭合的双链环状分子(图2),全长均为153 114 bp,是一个典型的4分区域结构,包括1个大单拷贝区(LSC)、1对反向重复区(IR)和1个短单拷贝区(SSC),红花3种花色类型的4个区长度均相

因及7个未知功能基因(表2)。在tRNA中trnA-UGC、trnI-CAU、trnI-GAU、trnL-CAA、trnN-GUU、trnR-ACG和trnV-GAC各有2个拷贝, trnF-GAA有3个拷贝;4个核糖体RNA(rrn5 rRNA、rrn4.5 rRNA、rrn23 rRNA、rrn16 rRNA)均有2个拷贝,分别位于反向重复区IRA和IRB。核糖体蛋白大小亚基编码的基因中,rps7和rpl2这2个基因均有2个拷贝,其余为1个拷贝。NADH脱氢酶亚基中的ndhB基因及未知功能蛋白基因ycf1和ycf2的拷贝数均为2。红花叶绿体基因组中有21个基因有内含子。其中, trnA-UGC、trnI-CAU、trnI-GAU、trnK-UUU、trnL-UAA、trnV-UAC、rps16、rpl2、rpl16、rpl23、rpoC1、ndhA、ndhB、petB、petD和atpF各有1个内含子,而rps12、clpP、ycf3具2个内含子。matK基因位于tmK-UUU基因内,整个编码区为trnK-UUU内含子的一部分,存在序列共用现象。



LSC 和 SSC: 大单拷贝区域、小单拷贝区域; IRA 和 IRB: 2 个反向重复区域; 内圈深色部分: GC 含量  
 LSC and SSC: Large single copy and small single copy; IRA and IRB: two inverted repeats; the darker gray area in inner circle: GC content

图 2 红花 (3 种花色) 叶绿体基因组环状图

Fig. 2 Map of *C. tinctorius* (three color) chloroplast genome

表 2 红花叶绿体基因组编码的基因

Table 2 Genes present in chloroplast genome of *C. tinctorius*

基因分组	基因名称	数量
tRNA	<i>trnA-UGC<sup>ac</sup></i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA<sup>d</sup></i> , <i>trnFM-CAU</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU<sup>c</sup></i> , <i>trnI-GAU<sup>ac</sup></i> , <i>trnK-UUU<sup>b</sup></i> , <i>trnL-CAA<sup>c</sup></i> , <i>trnL-UAA<sup>a</sup></i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU<sup>c</sup></i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG<sup>c</sup></i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC<sup>c</sup></i> , <i>trnV-UAC<sup>a</sup></i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>	38
rRNA	<i>rrn5 rRNA<sup>c</sup></i> , <i>rrn4.5 rRNA<sup>c</sup></i> , <i>rrn23 rRNA<sup>c</sup></i> , <i>rrn16 rRNA<sup>c</sup></i>	8
核糖体蛋白小亚基	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7<sup>c</sup></i> , <i>rps8</i> , <i>rps11</i> , <i>rps12<sup>b</sup></i> , <i>rps14</i> , <i>rps15</i> , <i>rps16<sup>a</sup></i> , <i>rps18</i> , <i>rps19</i>	13
核糖体蛋白大亚基	<i>rpl2<sup>ac</sup></i> , <i>rpl14</i> , <i>rpl16<sup>a</sup></i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23<sup>a</sup></i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	11
RNA 聚合酶	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1<sup>a</sup></i> , <i>rpoC2</i>	4
NADH 脱氢酶亚基	<i>ndhA<sup>a</sup></i> , <i>ndhB<sup>ac</sup></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
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>	14
细胞色素 b/f 复合物亚基	<i>petA</i> , <i>petB<sup>a</sup></i> , <i>petD<sup>a</sup></i> , <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<sup>a</sup></i> , <i>atpH</i> , <i>atpI</i>	6
Rubisco 大亚基	<i>rbcL</i>	1
蛋白酶	<i>clpP<sup>b</sup></i>	1
成熟酶	<i>matK</i>	1
包裹膜蛋白	<i>cemA</i>	1
细胞色素 C 合成酶	<i>ccsA</i>	1
翻译起始因子	<i>infA</i>	1
未知功能蛋白	<i>ycf1<sup>c</sup></i> , <i>ycf2<sup>c</sup></i> , <i>ycf3<sup>b</sup></i> , <i>ycf4</i> , <i>lhbA</i>	7
总计		130

a 和 b 分别表示含有 1 个和 2 个内含子 c-含有 2 个拷贝基因 d-含有 3 个拷贝基因

a and b represent 1 and 2 introns respectively, c-contains 2 copy genes, d contains 3 copy genes

### 3.3 菊科部分物种叶绿体全基因组特征比较分析

菊科植物叶绿体基因组特征的比较分析见表3, 红花及其近缘物种共 18 种 (21 个样品) 植物叶绿体基因组的序列长度范围为 150 063~153 318 bp, 其中下田菊的叶绿体全基因组长度最短 (150 063 bp), 钻叶紫菀的叶绿体基因组最长 (153 318 bp)。不同花色的红花叶绿体基因组全长完全相同, 均为 153 114 bp, 介于菊科 18 个物种的叶绿体基因组长度范围之内。红花及其 17 种近缘植物叶绿体基因组 GC 含量的范围为 37.0%~37.8%, 钻叶紫菀的 GC 含量最低 (37.0%), 不同花色的红花 GC 值完全相同, 且含量最高 (37.8%)。

利用 Geneious 11.0.3 软件获取红花及 17 种近缘

菊科植物的 IR 区、LSC 区和 SSC 区序列。结果表明, IR 区的序列长度范围为 23 776~25 238 bp, 其中假臭草的 IR 区最短 (23 776 bp), 铺散矢车菊的 IR 区最长 (25 238 bp), 不同花色的红花 IR 区的长度完全一致均为 25 194 bp, 介于菊科 21 个样品 IR 区长度范围之内。LSC 区序列长度范围为 82 017~94 697 bp, 其中下田菊的 LSC 区长度最短 (82 017 bp), 紫菀的 LSC 区长度最长 (94 697 bp), 不同花色的红花 LSC 区的长度完全一致均为 84 128 bp, 介于菊科 18 个物种 LSC 区长度范围之内。SSC 区序列长度范围介于 17 949~18 598 bp, 其中豚草的 SSC 区最短 (17 949 bp), 不同花色的红花 SSC 区的长度完全一致且最长 (18 598 bp)。

表 3 菊科植物 18 种 (21 个样品) 的叶绿体基因组特征

Table 3 Chloroplast genome characteristics of 18 species (including 21 samples) plants of Asteraceae

物种	叶绿体基因组长度/bp	GC 值/%	LSC 长度/bp	SSC 长度/bp	IR 长度/bp
下田菊 <i>Adenostemma lavenia</i>	150 063	37.6	82 017	18 142	24 952
霍香蓟 <i>Ageratum conyzoides</i>	151 325	37.4	83 162	18 293	24 928
豚草 <i>Ambrosia artemisiifolia</i>	152 223	37.6	84 078	17 949	25 098
三脉紫菀 <i>Aster ageratoides</i>	153 031	37.3	84 896	18 269	24 953
紫菀 <i>A. tataricus</i>	152 992	37.3	94 697	18 250	25 022
红花	153 114	37.8	84 128	18 598	25 194
红花 (白)	153 114	37.8	84 128	18 598	25 194
红花 (浅黄)	153 114	37.8	84 128	18 598	25 194
红花 (橙红)	153 114	37.8	84 128	18 598	25 194
铺散矢车菊 <i>Centaurea diffusa</i>	152 559	37.7	83 596	18 487	25 238
野菊	151 095	37.5	83 038	18 323	24 867
<i>Chrysanthemum x morifolium cultivar</i>	151 109	37.5	82 931	18 350	24 914
菊花 <i>Chrysanthemum x morifolium</i>	151 033	37.5	82 780	18 353	24 950
向日葵 <i>Helianthus annuus</i>	151 104	37.6	83 530	18 308	24 633
狭苞橐吾 <i>Ligularia intermedia</i>	151 152	37.5	83 258	18 232	24 831
全缘橐吾 <i>L. mongolica</i>	151 118	37.5	83 244	18 214	24 830
离舌橐吾 <i>L. veitchiana</i>	151 253	37.5	83 330	18 247	24 838
阔苞菊 <i>Pluchea indica</i>	152 298	37.5	84 129	18 093	25 038
花梗阔苞菊 <i>P. pteropoda</i>	152 300	37.5	84 127	18 093	25 040
假臭草 <i>Praxelis clematidea</i>	151 410	37.2	85 309	18 549	23 776
钻叶紫菀 <i>Symphotrichum subulatum</i>	153 318	37.0	85 238	18 226	24 927

### 3.4 红花及菊科 11 种植物的叶绿体全基因组边界的特征

红花及本研究选取的管状花亚科 11 个属共 12 种植物叶绿体基因组的 IR-LSC 和 IR-SSC 边界比较显示 (图 3), 不同花色的红花, 其边界具有高度的保守性, 各边界侧翼基因完全相同, 扩张程度也完全一致。红花与其余 11 种菊科植物叶绿体全基因组相比, 边界处具有相同的基因, 但是基因的长度仍有差异, 部分边界基因也存在变化。LSC/IRb 边界 (JLB) 边界扩张范围显示, 红花与其余 11 种菊科

植物相比, JLB 边界均位于 rps19 基因内, 但扩张程度稍有差异, 红花 rps19 基因向 LSC 区域扩张 220 bp, 向 IRb 扩张 59 bp。SSC/IRb (JSB) 边界扩张范围显示, 红花与 11 种菊科植物 JSB 边界侧翼基因为 ycf1 和 ndhF 基因。铺散矢车菊和狭苞橐吾 JSB 边界位于 ycf1 基因左翼, 距离分别为 0、1 bp, 而红花与其余 9 种菊科植物的 JSB 边界均位于 ycf1 基因内, 但扩张程度稍有差异。SSC/IRa (JSA) 边界扩张范围显示, 而红花与紫菀、铺散矢车菊、狭苞橐吾和假臭草这 5 种菊科植物的 JSA 边界均位于

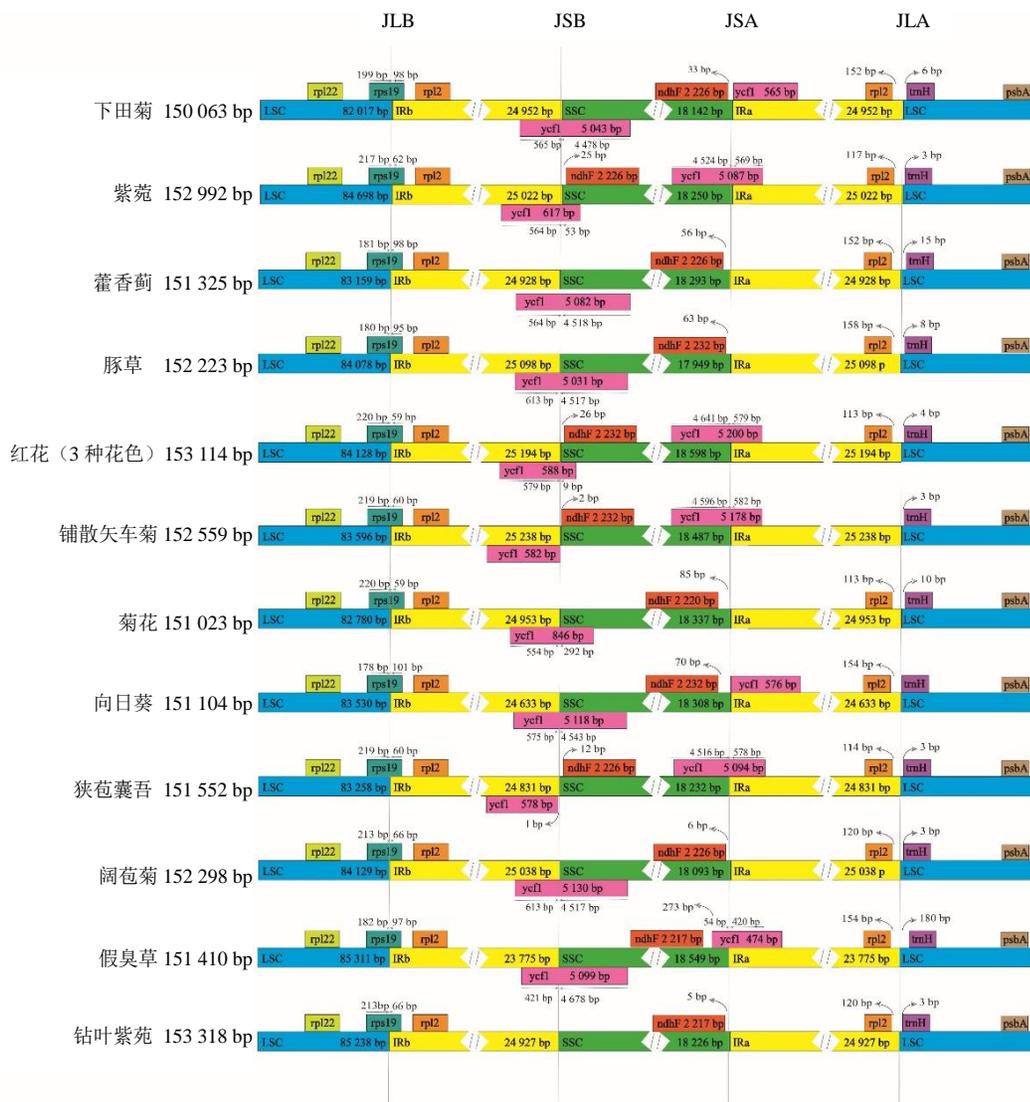


图 3 红花及菊科 11 种植物的叶绿体全基因组边界图

Fig. 3 Border regions of *C. tinctorius* and five Asteraceae plants Chloroplast genome

*ycf1* 基因内，但扩张程度稍有差异，而下田菊与向日葵 JSA 边界位于 *ycf1* 基因左翼、藿香蓟、豚草、菊花、阔苞菊和钴叶紫菀的 JSA 边界位于 *ndhF* 基因的右翼；LSC/IRa (JLA) 边界扩张范围显示，JLA 边界，边界处具有相同的基因，左翼为 *rpl2* 基因，右翼为 *trnH* 基因，但扩张程度稍有差异。此外，铺散矢车菊的叶绿体基因组在 IRb 区域，显示具有 *rpl2* 基因，IRa 区域存在 *rpl2* 基因序列，但是没有注释，因此 LSC/IRa 边界 (JLA) 中，*rpl2* 基因名称没有显示。

### 3.5 菊科叶绿体基因组基因间隔区信息位点分析

基于红花及其近缘物种共 18 种 (21 个样品) 的叶绿体基因组间隔区信息序列特征统计表明 (表 4)，在 21 个叶绿体基因组间隔区中，变异位点百分率

变化范围为 5.8%~28.8%，最高的 *trnH*-GUG-*psbA* 基因间隔区，其变异位点百分率为 28.8%。变异位点百分率超过 15% 的有 6 个，分别为 *ndhC*-*trnV*-UAC、*psbE*-*petL*、*psbM*-*trnD*-GUC、*rps18*-*rpl20*、*trnH*-GUG-*psbA*、*trnS*-GCU-*trnC*-GCA。变异位点百分率介于 10%~15% 的有 11 个，分别为 *atpB*-*rbcL*、*atpH*-*atpF*、*atpI*-*atpH*、*cemA*-*petA*、*psaI*-*ycf4*、*psaJ*-*rpl33*、*psbK*-*psbI*、*rpoC2*-*rps2*、*rps18*-*rpl20*、*trnC*-GCA-*petN*、*trnT*-GGU-*psbD*、*ycf4*-*cemA*。变异位点百分率低于 10% 的有 4 个，分别为 *psaA*-*ycf3*、*psaA*-*ycf3*、*rps2*-*atpI*、*trnP*-UGG-*psaJ*。这些变异位点百分率较高的叶绿体基因间隔区，能提供足够多的信息位点，为菊科属间和种间物种进化关系及分子鉴定提供较高的分辨率。

表 4 菊科植物 18 种 21 个样品叶绿体基因间隔区矩阵位点信息

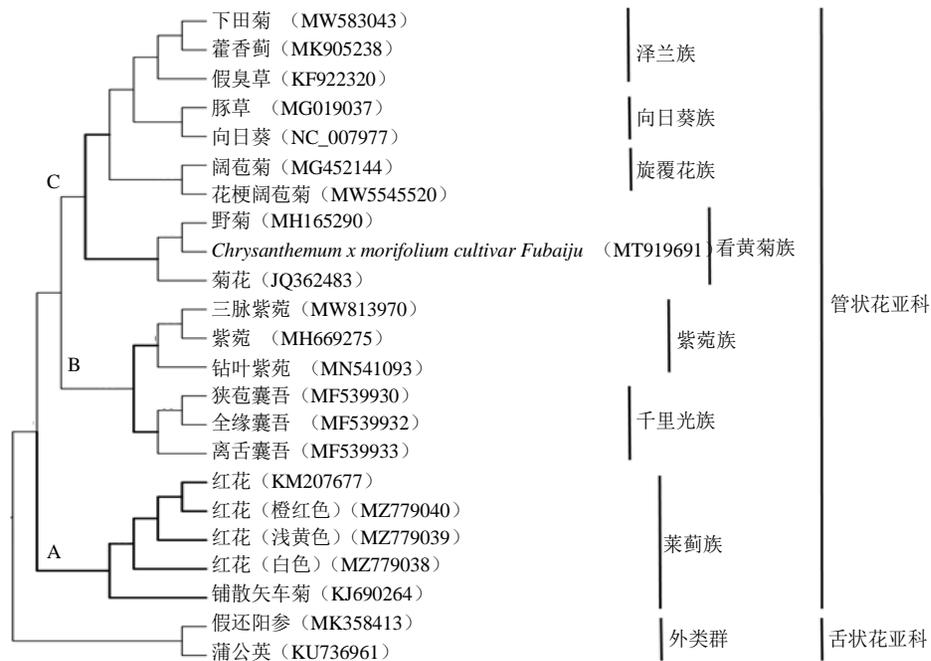
Table 4 Information of chloroplast intergenic region in 18 species (including 21 samples) of Asteraceae

编号	基因间隔区	信息位点/%	非信息位点/%	一致位点/%	编号	基因间隔区	信息位点/%	非信息位点/%	一致位点/%
1	atpB-rbcL	10.3	18.9	70.8	12	rpl20-rps12	7.7	5.2	87.1
2	atpH-atpF	14.4	6.9	78.7	13	rpoC2-rps2	13.1	5.9	81.0
3	atpI-atpH	14.7	6.0	79.3	14	rps2-atpI	5.8	4.6	89.6
4	cemA-petA	10.8	7.4	81.8	15	rps18-rpl20	16.2	8.0	75.8
5	ndhC-trnV-UAC	21.0	9.0	70.0	16	trnC-GCA-petN	10.3	8.6	81.1
6	psaA-ycf3	7.2	3.0	89.8	17	trnH-GUG-psbA	28.8	13.5	57.7
7	psaI-ycf4	12.7	6.6	80.7	18	trnP-UGG-psaJ	9.8	14.6	75.6
8	psaJ-rpl33	11.6	5.6	82.8	19	trnS-GCU-trnC-GCA	16.5	11.8	71.7
9	psbE-petL	15.4	6.7	77.9	20	trnT-GGU-psbD	14.8	11.0	74.2
10	1psbK-psbI	10.3	6.7	83.0	21	ycf4-cemA	13.7	7.5	78.8
11	psbM-trnD-GUC	18.0	18.9	63.1					

3.6 菊科叶绿体全基因组系统发育分析结果

基于红花及其近缘物种共 18 种 (21 个样品) 的叶绿体全基因组的系统发育树结果显示 (图 4), 菊科的管状花亚科类群构成一单系分支, 具有 100% 支持率。同一族取样的物种聚在一起, 单系性得到 100% 支持。在系统树上, 取样类群可以分为 3 个主要分支, A 分支是菜蓟族构成的单系分支, 具有

100% 支持率。红花 3 种不同花色类型与 NCBI 数据库的红花样品 (登录号: KM207677) 聚在一支, 具有 100% 支持率, 进一步与铺散矢车菊聚在一支, 支持率均为 100%。分支 B 包括 2 个族, 紫菀族和千里光族均构成单系分支具有 100% 的支持率。分支 C 包括 4 个族, 泽兰族、向日葵族、旋覆花族和看黄菊族构成单系分支具有 100% 的支持率。



分支上部数值表示 ML 分析的 BS 分析对该分支的支持强度 (>50%)

BS values of NJ (>50%) are shown above and below branches

图 4 基于叶绿体全基因组构建的系统发育树

Fig. 4 Phylogenetic tree based on complete chloroplast genome

#### 4 讨论

红花作为我国传统的大宗药材之一,具有重要的药用和经济价值。本研究完成了药用植物红花 3 种花色类型的叶绿体基因组的测序, 组装与注释。结果表明, 红花 3 种花色的叶绿体基因组序列完全一致, 具有典型的 4 分区域结构(图 2), 包括 1 个 LSC 区、1 对 IR 区和 1 个 SSC 区, 长度均为 153 114 bp。叶绿体基因组数据支持红花 3 种花色类型, 来源为同一种红花, 准确鉴定了红花 3 种花色变异类型的基原物种来源, 确保了红花临床用药的安全。红花花色变异可能与花青素生物合成途径相关转录因子分子调控机制有关<sup>[31]</sup>, 浅黄色的红花变异类型可能来源于橙红色和白色的变异类型的自然杂交。

对比红花与菊科其他物种的叶绿体基因组, 结果表明红花叶绿体基因组长度、IR 区和 LSC 区长度, 均位于菊科其他 17 个物种叶绿体基因组长度范围之内。但红花叶绿体基因组 GC 值及 SSC 区序列长度均高于菊科其他 17 个物种。红花与菊科其他 11 个属共 12 种植物的叶绿体基因组 IR 边界分析(图 3)显示, 其 LSC/IRb 边界(JLB), SSC/IRb 边界(JSB), SSC/IRa 边界(JSA)和 LSC/IRa 边界(JLA)的侧翼基因完全相同, 各边界基因序列的扩张长度略有差异。铺散矢车菊的叶绿体基因组在 IRb 区域具有 *rpl2* 基因, 同样 IRa 区域也具有 *rpl2* 基因, 但没有注释, 因此 LSC/IRa 边界(JLA)中, 铺散矢车菊 *rpl2* 基因名称没有显示(图 3)。相对于叶绿体基因组编码区基因具有较高的保守性, 叶绿体基因组的基因间隔区往往具有丰富的变异位点, 基于菊科管状花亚科 7 个族 18 种植物(21 个样品)的叶绿体基因组间隔区信息序列特征统计表明(表 4), 变异位点百分率超过 15% 的基因间隔区有 6 个, 分别为 *ndhC-trnV-UAC*、*psbE-petL*、*psbM-trnD-GUC*、*rps18-rpl20*、*trnH-GUG-psbA*、*trnS-GCU-trnC-GCA* 基因间隔区。其中 *trnH-GUG-psbA* 基因间隔区变异位点最高为 28.8%。这些变异位点百分率较高的叶绿体基因间隔区, 能提供足够多的信息位点, 为菊科属下大范围的物种鉴定, 杂交起源, 多倍体物种的形成和系统进化分析提供可靠的分子标记片段。

为进一步界定药用植物红花在菊科的系统位置, 探讨红花属与其近缘属的系统发育关系。基于菊科 18 种(21 个样品)叶绿体基因组全长构建系统发育树结果显示, 取样的管状花亚科所有类群聚

在一单系分支具有 100% 支持率, 红花不同花色类型的样品聚在一起具有 100% 支持率, 红花(所在红花属)与铺散矢车菊(所在矢车菊属)聚在一支, 构成姐妹群(分支 A)具有 100% 支持率。中国植物志记载<sup>[1]</sup>, 菊科菜蓟族的矢车菊亚族(Centaureinae)主要的特征在于它的瘦果有侧生着生面, 而药用植物红花(所在红花属)具有典型的瘦果有侧生着生面的形态特征。本研究叶绿体基因组数据分子结果支持传统分类研究结果<sup>[1]</sup>, 再次验证药用植物红花(所在红花属)隶属于菜蓟族的矢车菊亚族。

菊科是被子植物中种类最多的一个科, 包含大量的药用植物。同时菊科植物也是分布最广的一个类群, 是双子叶植物中进化地位最高的一个科, 物种形态变异多样<sup>[1]</sup>。本研究首次报道了药用植物红花不同花色变异类型的叶绿体全基因组, 综合分析菊科药用植物如红花、菊花、野菊花等的叶绿体基因组序列、结构和特征, 筛选出一批高变异的叶绿体基因组的间隔区。这为菊科药用植物的分子鉴定, 种质资源保护, 系统进化关系及遗传多样性研究奠定基础。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] 中国科学院中国植物志编辑委员会. 中国植物志(第三十八卷) [M]. 北京: 科学出版社, 1986: 56.
- [2] 中国药典 [S]. 一部. 2020: 1088.
- [3] 杨玉霞, 吴卫, 郑有良. 红花研究进展 [J]. 四川农业大学学报, 2004, 22(4): 365-369.
- [4] 陈梦, 赵丕文, 孙艳玲, 等. 红花及其主要成分的药理作用研究进展 [J]. 环球中医药, 2012, 5(7): 556-560.
- [5] Wu Z H, Liao R, Dong X, et al. Complete chloroplast genome sequence of *Carthamus tinctorius* L. from PacBio sequel platform [J]. *Mitochondrial DNA B Resour*, 2019, 4(2): 2635-2636.
- [6] Twyford A D, Ness R W. Strategies for complete plastid genome sequencing [J]. *Mol Ecol Resour*, 2017, 17(5): 858-868.
- [7] Yu X Y, Zuo L H, Lu D D, et al. Comparative analysis of chloroplast genomes of five *Robinia* species: Genome comparative and evolution analysis [J]. *Gene*, 2019, 689: 141-151.
- [8] Zhang F, Wang T, Shu X, et al. Complete chloroplast genomes and comparative analyses of *L. chinensis*, *L. anhuiensis*, and *L. aurea* (Amaryllidaceae) [J]. *Int J Mol Sci*, 2020, 21(16): E5729.

- [9] 邢少辰, Liu C J. 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.
- [10] 张靖雯, 姜在民, 蔡靖. 紫丁香与羽叶丁香叶绿体DNA提取方法研究 [J]. 西北林学院学报, 2018, 33(4): 95-99.
- [11] 倪梁红, 赵志礼, 米玛. 药用植物叶绿体基因组研究进展 [J]. 中药材, 2015, 38(9): 1990-1994.
- [12] Antipov D, Hartwick N, Shen M, *et al.* plasmidSPAdes: Assembling plasmids from whole genome sequencing data [J]. *Bioinformatics*, 2016, 32(22): 3380-3387.
- [13] Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data [J]. *Nucleic Acids Res*, 2017, 45(4): e18.
- [14] Jin J J, Yu W B, Yang J B, *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes [J]. *Genome Biol*, 2020, 21(1): 241.
- [15] Kim K J, Lee H L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants [J]. *DNA Res*, 2004, 11(4): 247-261.
- [16] 李西文, 胡志刚, 林小涵, 等. 基于 454FLX 高通量技术的厚朴叶绿体全基因组测序及应用研究 [J]. 药学报, 2012, 47(1): 124-130.
- [17] Zhang Y Z, Ma J, Yang B X, *et al.* The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (Taxaceae): Loss of an inverted repeat region and comparative analysis with related species [J]. *Gene*, 2014, 540(2): 201-209.
- [18] Li X, Zhang T C, Qiao Q, *et al.* Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae) [J]. *PLoS One*, 2013, 8(3): e58747.
- [19] 宋菊, 龙月红, 林丽梅, 等. 五加科植物叶绿体基因组结构与进化分析 [J]. 中草药, 2017, 48(24): 5070-5075.
- [20] Ma Y P, Zhao L, Zhang W J, *et al.* Origins of cultivars of *Chrysanthemum*—Evidence from the chloroplast genome and nuclear *LFY* gene [J]. *J Syst Evol*, 2020, 58(6): 925-944.
- [21] Biswal D, Konhar R, Debnath M, *et al.* Chloroplast genome sequence annotation of *Dendrobium nobile* (Asparagales: Orchidaceae), an endangered medicinal orchid from northeast India [J]. *PLoS Curr*, 2017, 9: ecurrents.tol.cf1709613759c2223eb582c0fa694cc7.
- [22] Xia Z, Li C C, Hu S W, *et al.* The complete chloroplast genome of Chinese medicine cultivar species of *Rehmannia glutinosa* (Orobanchaceae) [J]. *Mitochondrial DNA B Resour*, 2021, 6(1): 290-292.
- [23] Li C C, Hu S W, Ding Y N, *et al.* The complete chloroplast genome of Chinese medicinal herb *Belamcanda chinensis* (L.) Redouté (Iridaceae) [J]. *Mitochondrial DNA B Resour*, 2021, 6(2): 331-332.
- [24] Kearse M, Moir R, Wilson A, *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. *Bioinformatics*, 2012, 28(12): 1647-1649.
- [25] Wang W B, Yu H, Wang J H, *et al.* The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae) [J]. *Int J Mol Sci*, 2017, 18(11): 2288.
- [26] Dong W L, Wang R N, Zhang N Y, *et al.* Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution [J]. *Int J Mol Sci*, 2018, 19(3): 716.
- [27] 李世茂. 基于叶绿体 DNA 基因间隔序列的菊花与近缘种亲缘关系研究 [D]. 武汉: 华中农业大学, 2010.
- [28] Zhang D, Gao F L, Jakovlić I, *et al.* PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies [J]. *Mol Ecol Resour*, 2020, 20(1): 348-355.
- [29] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability [J]. *Mol Biol Evol*, 2013, 30(4): 772-780.
- [30] Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies [J]. *Bioinformatics*, 2014, 30(9): 1312-1313.
- [31] 李莹, 高振蕊, 张驰, 等. 花青素合成途径中分子调控机制的研究进展 [J]. 生态学杂志, 2015, 34(10): 2937-2942.

[责任编辑 时圣明]