

旋覆花、湖北旋覆花和线叶旋覆花的叶绿体基因组比较分析和系统发育研究

吴茜, 姜梅, 陈海梅, 王立强, 黄林芳, 刘昶*

(中国医学科学院、北京协和医学院药用植物研究所, 教育部中药材生物活性物质与资源利用重点实验室, 教育部中药材资源工程研究中心, 北京 100193)

摘要: 旋覆花、湖北旋覆花和线叶旋覆花均为菊科旋覆花属药用植物, 其中湖北旋覆花为我国特有。为分析比较旋覆花属 3 个物种的基因组序列差异, 进而为该属植物分子鉴别和种质资源开发提供科学依据, 本文采用二代测序技术对旋覆花属 3 个物种的总 DNA 进行测序, 使用 NOVOPlasty 成功组装了完整的叶绿体基因组, 然后用 CPGAVAS2 对 3 个基因组的基因组组成及重复序列进行了分析比较。最后用 RAxML 对旋覆花属 3 个物种及其近缘物种进行了系统发育分析。结果显示, 旋覆花、湖北旋覆花和线叶旋覆花的叶绿体基因组全长分别为 150 754、150 909、150 812 bp, 均由一个大的单拷贝区 (LSC) (82 427、82 552、82 474 bp), 一个小的单拷贝区 (SSC) (18 431、18 453、18 434 bp) 和一对反向重复序列 (IRs) (24 948、24 952、24 952 bp) 组成。基因组注释结果显示, 3 个物种均共编码 111 个不同基因, 包括 79 个蛋白编码基因、28 个 tRNA 基因和 4 个 rRNA 基因。旋覆花、湖北旋覆花和线叶旋覆花叶绿体基因组 G/C 含量分别为 37.72%、37.71%、37.71%; 分别共检测到 32、33、34 个简单重复序列; 18、22、18 个串联重复序列和 33、37、38 个散在重复序列。系统发育分析结果表明, 旋覆花属 3 个物种与阔苞菊聚在一起; 表明旋覆花族与千里光族亲缘关系较近, 可能起源于千里光族, 而非茺蔚族。本研究为研究旋覆花属物种鉴定、亲缘关系、进化历史与遗传多样性研究奠定了坚实的基础。

关键词: 旋覆花; 湖北旋覆花; 线叶旋覆花; 叶绿体基因组; 重复序列; 系统进化

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2020)05-1042-08

Comparative analysis of three complete chloroplast genomes of *Inula* genus with phylogenetic analysis of 49 plants from *Carduoideae*

WU Xi, JIANG Mei, CHEN Hai-mei, WANG Li-qiang, HUANG Lin-fang, LIU Chang*

(Key Laboratory of Bioactive Substances and Resource Utilization of Chinese Herbal Medicine from Ministry of Education, Engineering Research Center of Chinese Medicine Resources from Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100193, China)

Abstract: *Inula japonica*, *Inula hupehensis* and *Inula linariifolia* are all medicinal plants of *Inula* L. in the Compositae family, and *Inula hupehensis* is endemic to China. In order to compare their genomic sequence differences and provide scientific basis for their germplasm conservation and development, we obtained and analyzed the complete chloroplast genomes of these three species. Total DNAs were extracted from fresh leaves and subjected to next-generation DNA sequencing. NOVOPlasty was used to assemble the chloroplast genomes from the sequence reads. CPGAVAS2 was used to annotate the genes and repeats in each genome. Lastly, phylogenomics

收稿日期: 2019-09-24; 修回日期: 2020-02-02.

基金项目: 中国医学科学院医学与健康创新工程创新基金资助项目 (CIFMS) [2016-I2M-3-016, 2017-I2M-1-013]; 国家自然科学基金资助项目 (81872966); 中国国家科技基础资源调查项目 (2018FY100705).

*通讯作者 Tel: 86-10-57833111, Fax: 86-10-62899715, E-mail: cliu6688@yahoo.com

DOI: 10.16438/j.0513-4870.2019-0774

analysis was conducted using RAxML. The results showed that the total length of the chloroplast genome of *Inula japonica*, *Inula hupehensis* and *Inula linaariifolia* is 150 754, 150 909, and 150 812 bp respectively, each consisting of a large single copy region, a small single copy region and a pair of inverted repeat regions. In addition, the G/C content of all three chloroplast genomes was approximately 37.7% and each encoded 111 unique genes, including 79 protein-coding, 28 tRNA and 4 rRNA genes. Meanwhile, 32, 33, 34 simple repeat sequences, 18, 22, 18 tandem repeat sequences and 33, 37, 38 scattered repeat sequences were identified in three species. Phylogenomic analysis showed that all three species of *Inula* L. and *Pluchea indica* were clustered together, with the relationship between Inuleae and Senecioneae closer, suggesting that Inuleae may have originated from the Senecioneae, not the Cardueae. The data in this study not only enriches the chloroplast genome database of *Inula* L., but also lays the foundation for the future studies of species identification, phylogenetic relationships, evolution history and genetic diversity of *Inula* species.

Key words: *Inula japonica*; *Inula hupehensis*; *Inula linaariifolia*; chloroplast genome; repeat sequence; phylogenetic evolution

旋覆花属 (*Inula* L.) 属于菊科 (Compositae), 包含约 100 种植物, 分布于欧洲、非洲和亚洲等地区, 部分为广布种, 部分为特有种。我国有 20 余种和多个变种, 进一步分为 5 组 11 系^[1,2]。《中国药典》2015 年版收载品种为菊科植物旋覆花 (*Inula japonica* Thunb) 和欧亚旋覆花 (*Inula britannica* L.) 的干燥头状花序, 具有降气、消痰、行水、止呕的功效, 用于风寒咳嗽、痰饮蓄结、胸膈痞闷、喘咳痰多、呕吐噎气、心下痞硬^[3]。旋覆花富含甾醇、倍半萜及黄酮类化合物, 其中, 倍半萜类化学成分被认为是菊科旋覆花属植物的特征性成分, 且不同种旋覆花属植物的倍半萜含量有显著差异^[4]。药理研究表明, 旋覆花属植物的提取物及单体成分在抗肿瘤、抗炎、抗血脂和护肝及治疗糖尿病等方面具有良好的生物活性^[5]。

该属植物地方习用品和伪混品较多, 且药用功效有一定差异^[6,7]。如部分地区使用线叶旋覆花 (*Inula linaariifolia* Turcz.), 云南、贵州使用同属植物水朝阳旋覆花 (*Inula helianthus-aquatuca* C. Y. Wu ex Ling), 湖北、四川使用湖北旋覆花 (*Inula hupehensis*), 甘肃使用沙旋覆花 (*Znula salsoloides* (Turca.) Ostenf.)。伪混品有柳叶旋覆花 (*Inula salicina* L.) 和山白菊 (*Aster ageroides* Turcz.)。药用功效方面, 水煎剂旋覆花有镇咳、祛痰和抗炎作用, 水朝阳旋覆花有镇咳和抗炎作用, 湖北旋覆花有抗炎和祛痰作用。目前, 市售品混淆现象严重, 有报道显示, 线叶旋覆花煎剂刺激性较大, 会导致呕吐现象^[8]; 此外, 近年来也有旋覆花引起皮肤过敏的报道^[9]。因此, 早在上个世纪 90 年代, 王建华等^[7]就对中药旋覆花的 7 种原植物进行过形态组织学研究, 使从药材来源上保证使用安全。但不能很好的区分干药材。刘义梅等^[10]采用提取植物 DNA, 通过比较 ITS 序列对旋覆花属 8 种药用植物 (旋覆花、欧亚旋

覆花、总状土木香、土木香、羊耳菊、泽兰羊耳菊、赤茎羊耳菊、显脉旋覆花) 进行了鉴定, 发现能很好的区分。施之琪等^[11]采用 HPLC 通过鉴定特征成分峰能区分旋覆花、水朝阳旋覆花、湖北旋覆花和山黄菊 2 种药材制成的配方颗粒, 为临床安全用药提供了参考。但从基因组水平, 对旋覆花属植物或产品进行区分的文章尚未见报道。

叶绿体基因组序列是植物物种鉴定和系统发育学研究的有效工具。与核基因组相比, 叶绿体基因组具有结构简单、分子量小和拷贝多的特点。同时, 叶绿体基因组具有高度保守性。在不同物种之间, 乃至同一物种不同个体间又存在着一定的局部区域的变异。叶绿体基因组的整体高度保守性和局部区域的变异性为其应用于植物系统发育学研究提供了基础与条件^[12]。近年来随着基因组测序技术的发展, 越来越多植物的叶绿体全基因组序列得到解析。截止 2019 年 9 月, GenBank 的 RefSeq 数据库已经收录了包括陆地植物、海洋植物在内的 3 490 种植物的叶绿体基因组信息^[13]。然而, 旋覆花属植物叶绿体基因组序列尚未见发布。本文通过比较分析旋覆花属 3 个物种 (旋覆花、湖北旋覆花和线叶旋覆花) 的叶绿体基因组结构信息, 为该属物种精准鉴定和系统发育研究提供参考, 进一步为旋覆花属植物种质资源保护与开发提供科学依据。关于旋覆花族的起源, 《中国植物志》上有 2 种不同的观点, 一种观点认为其起源于千里光族, 另一种观点认为其起源于菜蓟族^[2]。本研究旨在通过构建基因组系统发育树, 为阐明该族起源提供有力的证据。

材料和方法

植物材料、DNA 提取与测序 旋覆花采自河南南阳桐柏山脉, 线叶旋覆花采自河北鸽子窝公园, 由中国

医学科学院药用植物研究所黄林芳研究员鉴定;湖北旋覆花叶片采自华中药用植物园(湖北恩施),由湖北省农科院中药材研究所尤金文教授鉴定。新鲜叶片摘取后放入液氮迅速冷冻,随后放于-80 °C冰箱待用。总DNA提取采用植物基因组DNA提取试剂盒(天根生物,北京),利用琼脂糖凝胶, NanoDrop2000 微量分光光度计(Thermo Scientific, 美国)和 Qubit3.0 检测总DNA的浓度和质量,取500 ng DNA构建文库,采用Illumina Hiseq 2500平台测序,旋覆花、湖北旋覆花、线叶旋覆花测序数据大小分别为5.2G、5.1 G、5.7G。其原始数据已经提交到GenBank,相应的Bioproject ID为PRJNA604096。

叶绿体基因组组装和注释 选取拟南芥的*rbcl*基因作为种子序列,采用NOVOPlasty(v.2.7.2)从原始测序数据中组装叶绿体基因组^[14]。为验证组装的正确性,采用Bowtie2(V.2.0.1)^[15]将所有原始reads映射(mapping)到旋覆花属物种的叶绿体基因组组装序列上,通过reads映射的覆盖度及深度检测叶绿体基因组组装的正确性,尤其是各contig连接处的正确性。叶绿体基因组的注释采用CPGAVAS2^[16],注释信息的编辑、校正采用Apollo^[17]。基于CPGAVAS2软件绘制叶绿体基因组环状示意图^[18]。利用sequin软件提交旋覆花、湖北旋覆花和线叶旋覆花叶绿体基因组序列及注释至GenBank,登录号分别为MN974527、MN418395、MN954653。

重复序列分析 重复序列分析和密码子使用情况均使用CPGAVAS2分析获得。其中,使用MISA发现卫星序列^[19],参数设置阈值为:单、双、三、四、五、六核苷酸的单位数分别为10、6、5、5、5和5。使用TRF软件发现串联重复序列^[20],设定重复单元的大小 ≥ 7 。采用VMATCH软件发现散布重复序列^[21]。GC含量和密码子使用偏好均采用EMBOSS(v6.3.1)^[15]中的Cusp程序计算获得。

系统发育分析 从PlasDB数据库(<http://www.herbalgenomics.org/plasdb>)下载来自管状花亚科11个族49个属和2个外类群物种的所有叶绿体编码蛋白质序列,其中59个为共有蛋白质:*atpA*、*atpE*、*atpF*、*atpH*、*rps2*、*rps3*、*rps4*、*rps7*、*rps8*、*rps11*、*rps12*、*rps15*、*rps18*、*rps19*、*rpl2*、*rpl14*、*rpl16*、*rpl20*、*rpl22*、*rpl23*、*rpl32*、*rpl33*、*rpl36*、*rpoA*、*rpoB*、*rpoC2*、*petA*、*petB*、*petD*、*petG*、*petL*、*psaA*、*psaB*、*psaC*、*psaI*、*psaJ*、*psbA*、*psbD*、*psbE*、*psbF*、*psbH*、*psbI*、*psbM*、*psbN*、*psbT*、*ndhB*、*ndhC*、*ndhE*、*ndhF*、*ndhG*、*ndhH*、*ndhI*、*ndhJ*、*ndhK*、*rbcL*、*matK*、*cemA*、*ccsA*和*ycf4*。使用Clustal W进行多重序列比对。以拟南芥(*Arabidopsis thaliana*)和烟草(*Nicotiana_*

tabacum)作为外类群,利用RAxML(Randomized Axelerated Maximum Likelihood)软件实现的最大似然法构建系统发育树。具体参数设置为:“RAxMLHPC-PTHREADS-SSE3-f-N 1000-m PROTGAMMACPREV-x 551314260-p 551314260-o A_thaliana, N_tabacum T 20”。进化树枝的可信度用Bootstrap 1000次的平均值表示。

结果

1 旋覆花属3个物种叶绿体基因组结构特征比较

旋覆花、湖北旋覆花和线叶旋覆花的叶绿体基因组均为典型环状DNA分子,总长度分别为150 754、150 909和150 812 bp。三者均具有保守的四分状结构,包括一个大单拷贝区(large single copy region, LSC),长度分别为82 427、82 552、82 474 bp;一个小单拷贝区(small single copy region, SSC),长度分别为18 431、18 453、18 434 bp;和一对反向重复区(inverted repeat, IR),长度分别为24 948、24 952、24 952 bp(图1)。三者叶绿体基因组的整体G/C含量分别为37.72%、37.71%、37.71%,均低于IR区的G/C含量(43.01%、43.01%、43.00%),但均高于SSC区(31.29%、31.25%、31.25%)和LSC区(36.05%、35.94%、35.96%)的G/C含量(表1)。有研究者认为这是因为IR区存在4个高GC含量的rRNA基因^[22],也有研究者认为这表明LSC、SSC和IR区来源不同或具有不同选择压力^[23]。

Table 1 Summary of the plastome features for the three *Inula* L. species. LSC: Large single-copy region; SSC: Small single-copy region; IR: Inverted repeat

Features		<i>I. japonica</i>	<i>I. hupehensi</i>	<i>I. linaariifolia</i>
Length	Total	150 754	150 909	150 812
	LSC	82 427	82 552	82 474
	SSC	18 431	18 453	18 434
	IR	24 948	24 952	24 952
GC content/%	Total	37.72	37.71	37.71
	LSC	36.05	35.94	35.96
	SSC	31.29	31.25	31.25
	IR	43.01	43.01	43.00
No. of gene	Total	131	131	131
	Protein coding	87	87	87
	tRNA	36	36	36
	rRNA	8	8	8

在旋覆花属3个物种的叶绿体基因组中均预测到131个编码基因,包括111个不同的基因,其中蛋白质编码基因80个、转运RNA(transfer RNA)基因28个和核糖体RNA(ribosome RNA)基因4个(表2)。三者均检测到9个基因(*rps16*、*rpoC1*、*atpF*、*rpl2*、*ndhB*、

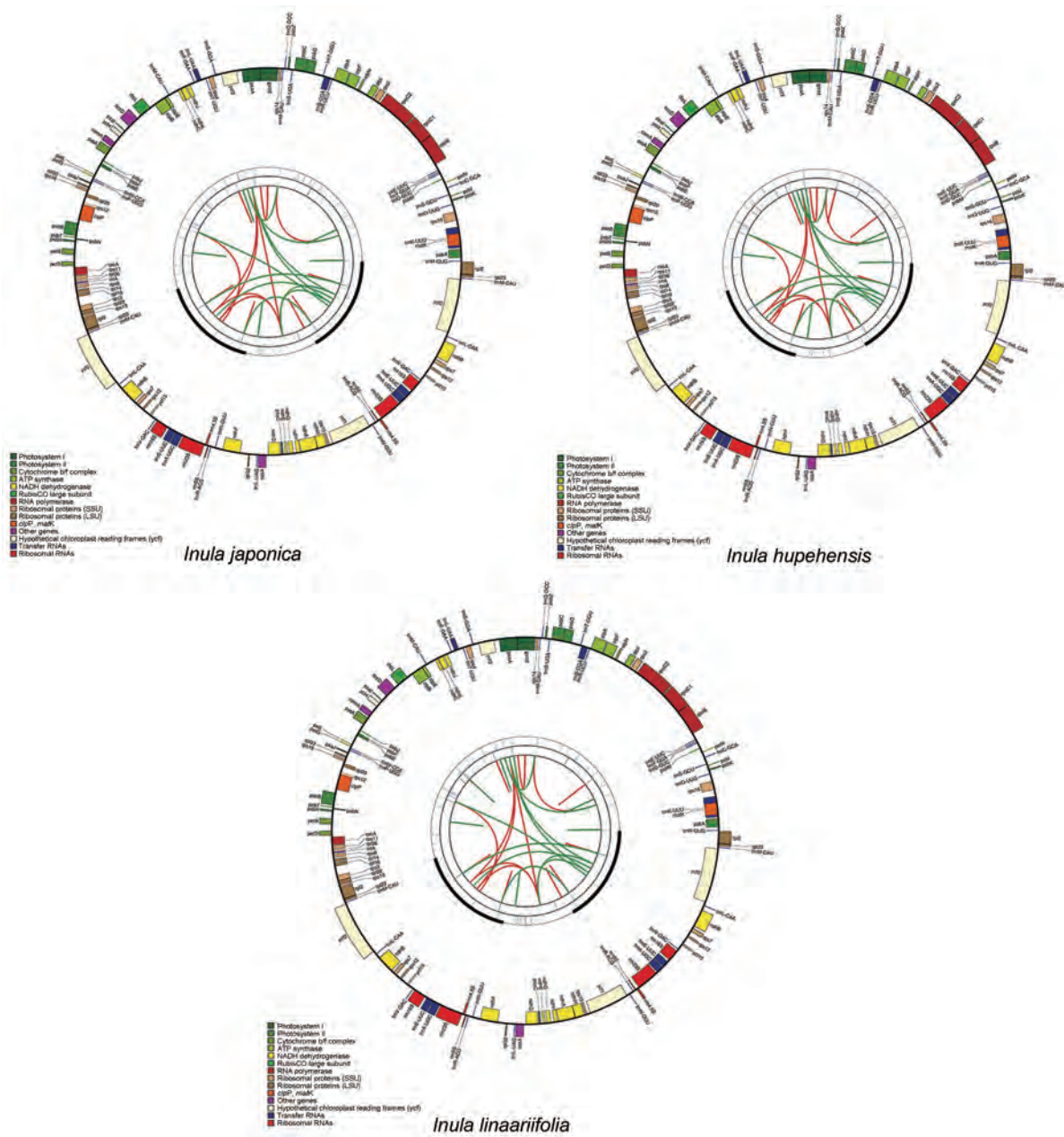


Figure 1 Graphic representation of features identified in three *Inula* L. species plastomes by using CPGAVAS2. The map contains four rings. From the center going outward, the first circle shows the forward and reverse repeats connected with red and green arcs, respectively. The next circle shows the tandem repeats marked with short bars. The third circle shows the microsatellite sequences identified using MISA. The fourth circle is drawn using draw gene map and shows the gene structure on the plastome. The genes were colored based on their functional categories, which are shown at the left corner.

ndhA、*petB*、*rpoA*、*ycf2* 含 1 个内含子 (intron), 2 个基因 (*ycf3*、*clpP*) 含 2 个内含子。旋覆花、湖北旋覆花和线叶旋覆花叶绿体全基因组中蛋白质编码区 (coding sequence, CDS) 的长度分别为 78 702、78 705、78 702 bp, 各自占其基因组总长度的 52.21%、52.15%、52.19%; rRNA 基因的长度均为 9 394 bp, 分别占各自基因组总长度的 6.23%、6.22%、6.23%; tRNA 基因的长度均为 2 726 bp, 均占各自基因组总长度的 1.81%。此外, 旋

覆花、湖北旋覆花和线叶旋覆花的叶绿体基因组非编码区均主要包括内含子和基因间隔区, 分别占各自整个基因组长度的 39.75%、39.81%、39.78%。

在旋覆花属 3 个物种中, 共发现 86 个蛋白编码基因, 编码 26 235 个密码子。其中包括编码 20 个氨基酸的 61 个唯一密码子和 3 个终止密码子。在这些密码子中, 编码频次最高的氨基酸为异亮氨酸, 编码 2 793 次, 占比 10.65%, 编码频次最低的氨基酸为半胱氨酸, 只

Table 2 Gene contents of the three *Inula* L. plastomes

Gene function	Group of genes	Gene names	Amount
rRNA	rRNA genes	<i>rrn23S</i> (×2), <i>rrn16S</i> (×2), <i>rrn5S</i> (×2), <i>rrn4.5S</i> (×2)	8
tRNA	tRNA genes	<i>trnH</i> -GUG, <i>trnK</i> -UUU, <i>trnQ</i> -UUG, <i>trnS</i> -GCU, <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnY</i> -GUA, <i>trnE</i> -UUC (×3), <i>trnR</i> -UCU, <i>trnS</i> -CGA, <i>trnT</i> -GGU, <i>trnS</i> -UGA, <i>trnG</i> -GCC, <i>trnM</i> -CAU (×2), <i>trnS</i> -GGA, <i>trnT</i> -UGU, <i>trnL</i> -UAA, <i>trnF</i> -GAA, <i>trnM</i> -CAU (×2), <i>trnW</i> -CCA, <i>trnP</i> -UGG, <i>trnL</i> -CAA (×2), <i>trnV</i> -GAC (×2), <i>trnA</i> -UGC (×2), <i>trnR</i> -ACG, <i>trnN</i> -GUU (×2), <i>trnL</i> -UAG, <i>trnR</i> -ACG	36
Self replication	Large subunit of ribosome	<i>rpl14</i> , <i>rpl16</i> , <i>rpl2</i> (×2), <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	11
	DNA dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>	4
	Small subunit of ribosome	<i>rps11</i> , <i>rps12</i> (×2), <i>rps14</i> , <i>rps15</i> , <i>rps16</i> , <i>rps18</i> , <i>rps19</i> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i>	14
Photosynthesis	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>	6
	Subunits of photosystem II	<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>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i> , <i>ycf3</i>	15
	Subunits of NADH-dehydrogenase	<i>ndhA</i> , <i>ndhB</i> (×2), <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
	Subunits of cytochrome b/f complex	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>	6
	Subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	5
	Large subunit of rubisco	<i>rbcL</i>	1
	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>	1
Other genes	c-type cytochrom synthesis gene	<i>ccsA</i>	1
	Envelope membrane protein	<i>cemA</i>	1
	Protease	<i>clpP</i>	1
	Translation initiation factor	<i>infA</i>	1
	Maturase	<i>matK</i>	1
	Conserved open reading frames	<i>ycf1</i> , <i>ycf15</i> (×2), <i>ycf2</i> (×2), <i>ycf3</i> , <i>ycf4</i>	7
Unknown			
Total			131

编码了294次,占比1.12%。3个物种中相同密码子的相对同义密码子用法 (relative synonymous codon usage, RSCU) 值相似。此外,3个物种质体蛋白质编码区 (CDS) 内密码子第三位的A/T含量 (69.80%、69.80%、69.81%),分别高于第一位 (54.35%、54.34%、54.35%) 和第二位的A/T含量 (61.88%、61.88%、61.88%)。与其他大多数陆地植物中观察到的趋势一致,密码子第三位的A/T含量较第一、二位的高^[24]。

2 重复序列分析

重复序列在基因组进化中起着重要的作用,如大片DNA的插入、删除、重排等,并且可以影响基因组的长度和基因的顺序^[25]。本文主要分析叶绿体基因组中的3种重复序列,即简单重复序列 (simple sequence repeat, SSR),又称微卫星重复序列 (microsatellite sequence)、串联重复序列 (tandem repeats sequence) 和散在重复序列 (scattered repetitive sequence)。在旋覆花属3个物种的叶绿体基因组中,共检测到42个SSR (表3)。这些SSR主要分布在IGS和内含子序列中,共检测到38次。大多数单核苷酸重复序列由A/T重复序列组成,AT/AT二核苷酸重复序列是最常见的类型。这与Wang等^[26]的报道一致,即来自质体的SSR通常由短polyA或polyT重复序列组成,在许多植物中很少含有串联G或C重复序列。在这3个物种的4个基因

的CDS中,共鉴定出4个SSRs,包括*rpoB*、*rpoC2*、*rpoA*和*ycf1*。结果表明,这3个质体在SSRs的数目和GC含量上具有高度的保守性。

在旋覆花、湖北旋覆花和线叶旋覆花叶绿体基因组中分别发现有18、22、18个串联重复序列,这些序列重复单元之间的相似性均大于90%。大多数重复只有2个重复单元。重复单元的长度从10 bp到32 bp不等。

散在重复序列的重复单元是以散在方式分布于基因组内,主要有回文重复和正向回复2种类型。以e-value小于 10^{-4} 为阈值,旋覆花叶绿体基因组散在重复序列包括回文重复序列15条、正向重复序列20条;湖北旋覆花为17、20条;线叶旋覆花为18、20条。最长的散布重复单元为48 bp,分布在2个*ycf2*基因外显子中。这些重复序列将为分子标记开发和系统发育分化研究提供有意义的线索。

3 基因组系统发育分析

对来自管状花亚科49个属的51个物种和2个外类群物种 (拟南芥,烟草) 的59个共有蛋白质序列用最大似然法构建进化树。总体上各属物种均与本族物种聚在一起,与传统分类结果一致 (图2)。如图所示,管状花亚科可分为4支,鼠麴草族 (Gnaphalieae)、春黄菊族 (Anthemideae) 聚为一支;紫菀族 (Astereae) 单独为一支;千里光族 (Senecioneae)、旋覆花族 (Inuleae)、向

Table 3 Details for the simple sequence repeats in the three *Inula L.* species

SSR nr.	Location	Region	Motif	<i>I. japonica</i>		<i>I. hupehensi</i>		<i>I. linariifolia</i>	
				Repeat number	Position	Repeat number	Position	Repeat number	Position
1	<i>rps16</i>	Intron	A	NA	NA	NA	NA	10	6 113
2	<i>rps16-trnQ</i>	IGS	A	13	7 070	14	7 090	10	7 071
3	<i>trnC-petN</i>	IGS	A	17	9 757	15	9 778	16	9 758
4	<i>trnC-petN</i>	IGS	T	NA	NA	10	97 96	NA	NA
5	<i>trnE-rpoB</i>	IGS	T	12	12 569	11	12 597	12	12 578
6	<i>rpoB</i>	Exon	A	10	13 255	10	13 282	10	13 264
7	<i>rpoC1</i>	Intron	A	14	16 486	11	16 529	13	16 511
8	<i>rpoC2</i>	Exon	A	11	22 783	NA	NA	11	22 807
9	<i>rpoC2-rps2</i>	IGS	T	NA	NA	10	23 156	10	23 140
10	<i>atpI-atpH</i>	IGS	A	12	24 927	NA	NA	NA	NA
11	<i>atpI-atpH</i>	IGS	T	10	25 872	10	25 911	10	25 896
12	<i>atpF-atpA</i>	IGS	T	10	27 816	10	27 855	10	27 840
13	<i>trnR-trnS</i>	IGS	A	10	29 622	NA	NA	NA	NA
14	<i>trnT-psbD</i>	IGS	C	NA	NA	10	31 115	NA	NA
15	<i>psbC-trnS</i>	IGS	T	13	34 455	11	34 525	18	34 480
16	<i>psaA-ycf3</i>	IGS	T	12	41 332	25	41 401	11	41 362
17	<i>psaA-ycf3</i>	IGS	A	NA	NA	10	41 416	NA	NA
18	<i>ycf3-trnS</i>	IGS	T	10	43 701	NA	NA	11	43 730
19	<i>trnT-trnL</i>	IGS	T	10	45 776	NA	NA	NA	NA
20	<i>trnT-trnL</i>	IGS	A	11	45 809	15	45 887	13	45 839
21	<i>ndhC-trnM</i>	IGS	T	13	49 503	13	49 584	13	49 534
22	<i>atpB-rbcL</i>	IGS	T	10	53 016	10	53 097	10	53 047
23	<i>accD-psaI</i>	IGS	A	NA	NA	13	57 181	NA	NA
24	<i>psaI-ycf4</i>	IGS	T	10	57 706	11	57 792	11	57 769
25	<i>ycf4-cemA</i>	IGS	T	NA	NA	10	59 414	NA	NA
26	<i>psbE-petL</i>	IGS	A	15	63 824	16	63 918	17	63 888
27	<i>rpl33-rps18</i>	IGS	TA	12	66 098	12	66 193	12	66 164
28	<i>clpP</i>	Intron	T	16	68 731	19	68 826	13	68 797
29	<i>rpoA</i>	Exon	T	10	76 303	10	76 426	10	76 357
30	<i>rps8-rpl14</i>	IGS	T	14	78 734	16	78 857	19	78 788
31	<i>ycf15-trnV</i>	IGS	A	16	98 083	20	98 208	20	98 130
32	<i>rrn5S-trnR</i>	IGS	A	12	105 665	12	105 794	12	105 716
33	<i>ndhF-rpl32</i>	IGS	T	11	110 369	NA	NA	12	110 420
34	<i>rpl32-trnL</i>	IGS	A	11	110 988	15	111 115	11	111 040
35	<i>rpl32-trnL</i>	IGS	A	10	111 643	10	111 774	10	111 695
36	<i>ccsA-ndhD</i>	IGS	T	NA	NA	NA	NA	10	113 112
37	<i>ccsA-ndhD</i>	IGS	T	NA	NA	NA	NA	10	113 155
38	<i>ndhD-psaC</i>	IGS	A	10	114 702	11	114 833	10	114 756
39	<i>ndhG-ndhI</i>	IGS	A	10	116 384	10	116 522	10	116 438
40	<i>ycf1</i>	Exon	TCT	NA	NA	15	125 744	NA	NA
41	<i>trnR-rrn5S</i>	IGS	T	12	127 507	12	127 658	12	127 561
42	<i>trnV-ycf15</i>	IGS	T	16	135 085	20	135 236	20	135 139

日葵族 (Heliantheae)、泽兰族 (Eupatorieae) 等聚为一支; 菜蓟族 (Cardueae) 单独为一支。旋覆花族内, 旋覆花属 3 个物种与阔苞菊属阔苞菊 (*Pluchea indica*) 聚为一类, bootstrap 值高 (100%); 旋覆花属 3 个物种中, 旋覆花与湖北旋覆花亲缘关系较近, 二者与线叶旋覆花相对较远。本结果显示旋覆花族与千里光族亲缘关系较近, 与菜蓟族关系较远。这与《中国植物志》上认为的旋覆花族起源于千里光族的观点一致^[2]。

讨论

本文首次报道了旋覆花属 3 个植物旋覆花、湖北旋覆花和线叶旋覆花的叶绿体基因组序列及结构组成。该属 3 个物种均为典型环状 DNA 分子, 且具有保

守的四分状结构, 与已报道的同族阔苞菊叶绿体基因组大小相似^[27]。共预测了 130 个蛋白质编码基因, 包括 111 个不同的基因, 其中蛋白质编码基因 80 个。

本研究从旋覆花、湖北旋覆花和线叶旋覆花叶绿体基因组中分别共检测到 32、33、34 个 SSR; 共发现 18、22、18 个串联重复序列; 散在重复序列 33、37、38 条。这些重复序列将为接下来旋覆花属物种分子标记开发, 物种或产品鉴定提供理论依据。本课题组下一步研究将针对旋覆花属 3 个物种的差异序列进行分析筛选, 寻找合适的区域进行引物设计, 并采集实体样本进行 PCR 验证, 从分子水平鉴定旋覆花属物种。

关于旋覆花族的起源, 本研究通过对菊科管状花亚科 11 族 48 属植物进行系统进化分析发现, 旋覆花族

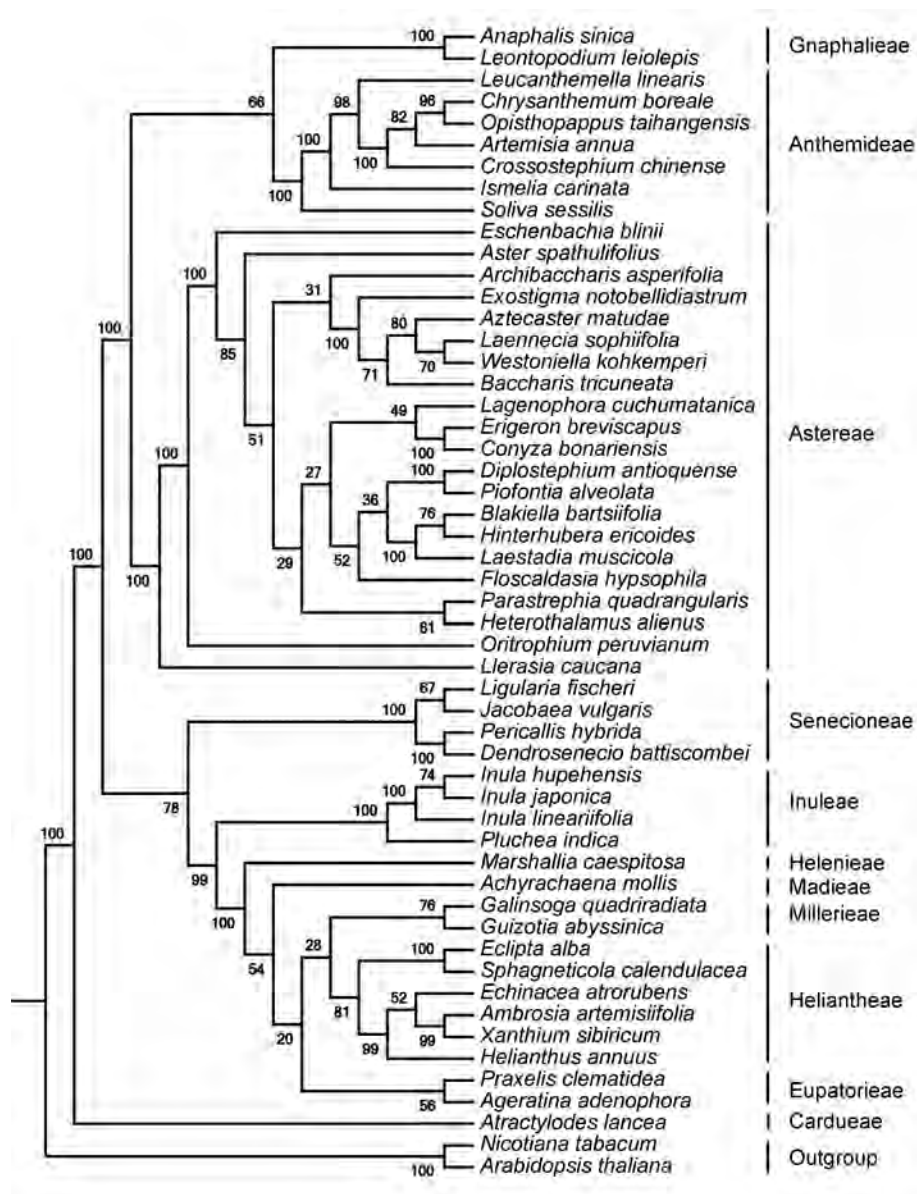


Figure 2 Molecular phylogenetic analyses. Plastome sequences of 59 common proteins present from 51 species were used to construct the phylogenetic tree with the maximum likelihood method implemented in the RaxML. Two taxa, namely, *N. tabacum* and *A. thaliana*, were used as outgroups. Tribes to which each species belongs are shown on the right side of the tree. Bootstrap supports were calculated from 1000 replicates

与千里光族亲缘关系较近,与菜蓟族亲缘关系较远,首次为旋覆花族起源提供了直接分子证据。

References

[1] Ling Y, Tseng YQ. Miscellaneous notes on Chinese Compositae [J]. J Univ Chin Acad Sci (中国科学院大学学报), 1978, 16: 82-86.
 [2] Editorial Board of Flora of China, Chinese Academy of Sciences. Flora of China Vol. 75 [M]. Beijing: Science Press, 1979, 75: 248-281.
 [3] Chinese Pharmacopoeia Committee. Pharmacopoeia of the

People's Republic of China (中华人民共和国药典) [S]. Part 1. Beijing: China Medical Science Press, 2015: 325.
 [4] Yang X, Liu H, He YJ, et al. Simultaneous determination of the contents of five compounds in *Inula* plants [J]. J Shenyang Pharm Univ (沈阳药科大学学报), 2012, 29: 116-120.
 [5] Fan LL, Chen JN, Zhang T, et al. Chemical constituents in Xuanfuhua (*Inula L.*) and their pharmacological activities [J]. Guid J Tradit Chin Med Pharm (中医药导报), 2017, 23: 40-43.
 [6] Wang JH, Qi Z, Jia GS, et al. Study on the pharmacological effect of the Traditional Chinese Medicine Xuanfuhua (*Inula L.*) and its local customs [J]. Beijing J Tradit Chin Med (北京中医), 1997, 1: 42-44.

- [7] Wang JH, Lou ZC. Herbalogical, morphological and histological studies of the Chinese drug XUAN-FU-HUA, Flos Inulae [J]. Acta Pharm Sin (药学报), 1983, 18: 950-964.
- [8] Li ZA. Adverse reactions caused by *Inula linariaefolia* [J]. J New Chin Med (新中医), 1985, 8: 55.
- [9] Gao PF, Wang YF, Zhang XJ. A case of skin hypersensitivity caused by XUN-FU-HUA decocting [J]. Clin J Tradit Chin Med (中医药临床杂志), 2004, 13: 473.
- [10] Liu YM, Yu S, Luo K, et al. ITS2 sequences analysis of eight medicinal plants in *Inula L.* [J]. Chin Tradit Herb Drug (中草药), 2014, 45: 410-414.
- [11] Shi ZQ, Fu JW, Cai CH, et al. Study on the characteristic HPLC chromatograms of *Inula japonica* formula granules [J]. Tradit Chin Drug Res Clin Pharm (中药新药与临床药理), 2017, 28: 765-770.
- [12] Olejniczak SA, Lojewska E, Kowalczyk T, et al. Chloroplasts: state of research and practical applications of plastome sequencing [J]. Planta, 2016, 244: 517-527.
- [13] NIH, U.S. National Library of Medicine. NCBI, Genome. [2019-09-24]. [https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/\[DB/OL\]](https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/[DB/OL]).
- [14] Dierckxsens N, Mardulyn P, Smits G. NOVO Plasty: *de novo* assembly of organelle genomes from whole genome data [J]. Nucleic Acids Res, 2017, 45: e18.
- [15] Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome [J]. Genome Biol, 2009, 10: R25.
- [16] Shi L, Chen H, Jiang M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. Nucleic Acids Res, 2019, 47: W65-W73.
- [17] Lewis SE, Searle SM, Harris N, et al. Apollo: a sequence annotation editor [J]. Genome Biol, 2002, 3: 82.
- [18] Lohse M, Drechsel O, Bock R. Organellar Genome DRAW (OG-DRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes [J]. Curr Gen, 2007, 52: 267-274.
- [19] Beier S, Thiel T, Munch T, et al. MISA-web: a web server for microsatellite prediction [J]. Bioinformatics, 2017, 33: 2583-2585.
- [20] Benson G. Tandem repeats finder: a program to analyze DNA sequences [J]. Nucleic Acids Res, 1999, 27: 573-580.
- [21] Kurtz S, Choudhuri JV, Ohlebusch E, et al. REPuter: the manifold applications of repeat analysis on a genomic scale [J]. Nucleic Acids Res, 2001, 29: 4633-4642.
- [22] Qian J. Study on the Chloroplast and Mitochondrial Genome of *Salvia miltiorrhiza* (丹参的叶绿体和线粒体基因组研究) [D]. Beijing: Chinese Academy of Medical Sciences, Peking Union Medical College, Tsinghua University, 2014.
- [23] Yang QQ, Jiang M, Wang LQ, et al. Complete chloroplast genome of *Allium chinense*: comparative genomic and phylogenetic analysis [J]. Acta Pharm Sin (药学报), 2019, 54: 173-181.
- [24] Nie XJ, Deng PC, Feng K, et al. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family [J]. Plant Mol Biol Reporter, 2014, 32: 828-840.
- [25] Tangphatsornruang S, Sangsrakru D, Chanprasert J, et al. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships [J]. DNA Res, 2009, 17: 11-22.
- [26] Wang S, Shi C, Gao LZ. Plastid genome sequence of a wild woody oil species, *Prinsepia utilis*, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes [J]. PLoS One, 2013, 8: e73946
- [27] Zhang Y, Zhang J, Yang Y, et al. Complete chloroplast genome of *Pluchea indica* (L.) Less. (Asteraceae) and its phylogenetic analysis [J]. Mitochondrial DNA B Resour, 2018, 2: 918-919.