

单叶铁线莲叶绿体基因组的序列特征与系统发育分析

韦海忠¹, 潘丽芹¹, 田盛野², 汤紫依², 何海叶², 张慧娟², 蒋明^{2*}

1. 台州科技职业学院, 浙江 黄岩 318020

2. 台州学院生命科学学院, 浙江 椒江 318000

摘要: 目的 以药用植物单叶铁线莲 *Clematis henryi* 叶片为材料, 在高通量测序、组装和序列分析的基础上, 明确叶绿体因组的序列特征和系统发育关系。方法 利用 CTAB 法提取叶片基因组 DNA, 通过 NovaSeq 6000 平台测序, 再用 NovoPlasty 组装叶绿体基因组, 借助 PhyML 生成系统发育树。结果 单叶铁线莲的叶绿体基因组全长 159 707 bp, GC 值为 38.0%, 大单拷贝区、小单拷贝区和反向重复序列的长度分别为 79 449、18 100、31 079 bp; 单叶铁线莲的叶绿体基因组总共有基因 137 个, 其中蛋白质编码基因、tRNA、rRNA 的数量分别为 91、36、8, 另有 2 个假基因, 分别为 $\Psi ycf1$ 和 $\Psi infA$ 。序列比对结果表明, 单叶铁线莲与牯牛铁线莲叶绿体基因组的相似性最高, 达 99.6%; 共检测到 41 个 SSR, 其中 40 个为单核苷酸重复, 1 个为三核苷酸重复。系统发育分析结果表明, 单叶铁线莲与牯牛铁线莲和大叶铁线莲聚于一组, 支持率达 100%。结论 单叶铁线莲叶绿体基因组的组装、序列分析和系统发育分析, 为该药用植物的遗传结构和遗传多样性研究奠定了基础。

关键词: 单叶铁线莲; 牯牛铁线莲; 大叶铁线莲; 叶绿体基因组; 序列特征; 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(2022)12-3766-08

DOI: 10.7501/j.issn.0253-2670.2022.12.023

Chloroplast genome sequence characterization and phylogenetic analysis of *Clematis henryi*

WEI Hai-zhong¹, PAN Li-qin¹, TIAN Sheng-ye², TANG Zi-yi², HE Hai-ye², ZHANG Hui-juan², JIANG Ming²

1. Taizhou Vocational College of Science and Technology, Huangyan 318020, China

2. College of Life Sciences, Taizhou University, Jiaojiang 318000, China

Abstract: Objective To confirm chloroplast genome sequence characteristics and phylogenetic relationships of *Clematis henryi* based on high-throughput sequencing, genome assembly, and sequence analysis. **Methods** The CTAB-based DNA extraction method was used to isolate leaf genomic DNA, and high-throughput sequencing was performed with the HiSeq X Ten Platform. The chloroplast genome was assembled using NovoPlasty and a phylogenetic tree was generated by PhyML. **Results** The full chloroplast genome of *C. henryi* was 159 707 bp in length and its GC content was 38.0%. The sizes of large single copy, small single copy, and inverted region were 79 449 bp, 18 100 bp, and 31 079 bp, respectively. Totally, there were 137 genes, and the numbers of protein-coding genes, tRNAs, and rRNAs were 91, 36, and eight, respectively. Additionally, $\Psi ycf1$ and $\Psi infA$ were recognized as two pseudogenes. Sequence comparison results showed that the highest similarity existed between *C. henryi* and *C. guniuensis*, which reached 99.6%. A total of 41 SSRs were detected, in which 40 were mono-nucleotide and one was tri-nucleotide. Phylogenetic analysis indicated that *C. henryi*, *C. guniuensis* and *C. heracleifolia* gathered in the same clade, with a support rate of 100%. **Conclusion** The assembly, sequence analysis, and phylogenetic analysis of *C. henryi* chloroplast genome provided insight into studies on both genetic structure and genetic diversity of this medicinal plant.

Key words: *Clematis henryi* Oliv.; *C. guniuensis* Oliv.; *C. heracleifolia* Oliv.; chloroplast genome; sequence characterization; phylogenetic analysis

铁线莲属 *Clematis* L. 为毛茛科 (Ranunculaceae) 本, 全世界约有 300 种, 主要产于热带、亚热带和多年生木质或草质藤本植物, 少数为直立灌木或草寒带地区, 我国有 100 多种, 全国各地均有分布^[1]。

收稿日期: 2021-12-02

基金项目: 台州市 211 人才工程经费资助 (2012 年度)

作者简介: 韦海忠, 男, 硕士, 副教授, 浙江东阳人, 研究方向为植物生理生化。E-mail: hzwei910@126.com

*通信作者: 蒋明, 男, 博士, 教授, 浙江嵊州人, 研究方向为植物基因组学、植物逆境生物学及其分子调控。E-mail: jiangming1973@139.com

铁线莲属植物的不少种类具有较好的药用价值,《中国药典》2020年版记载,小木通 *C. armandii* Franch.、绣球藤 *C. montana* Buch.-Ham. ex DC.、威灵仙 *C. chinensis* Osbeck、棉团铁线莲 *C. hexapetala* Pall.和东北铁线莲 *C. manshurica* Rupr.等植物的干燥藤茎、根或根茎可供药用^[2];《中国民族药辞典》则记录了40余种铁线莲属药用植物,如女萎 *C. apoofolia* DC.、甘青铁线莲 *C. tangutica* (Maxim.) Korsh.、西藏铁线莲 *C. tenuifolia* Royle 和单叶铁线莲 *C. henryi* Oliv.等^[3-4]。铁线莲属植物富含皂苷类、黄酮类、木脂素类、萜类和酚类化合物,具有抗肿瘤、抗炎、镇痛、抗氧化和抗菌等作用,用于治疗风湿痹痛、口舌生疮和筋脉拘挛等^[5-6]。

单叶铁线莲又名地雷根、雪里开、拐子药等,为木质藤本植物,根部膨大、单叶对生、花萼白色或淡黄色(图1)。单叶铁线莲在我国的分布范围较广,长江流域各省均有,生长在溪边、山谷、林下、路旁和灌丛等,是一种民间十分常用的中草药^[1]。

单叶铁线莲的根和叶可供药用,含熊果酸、胡萝卜苷、甘露醇和皂苷等,具有行气活血、清热解毒和消肿止痛等功效^[7-9]。目前,有关单叶铁线莲的研究集中在化学成分、组织培养、资源调查和药理作用等方面,而该植物叶绿体基因组相关的研究未见报道^[7-10]。叶绿体基因的数量较少,但它们在光合作用及氨基酸、激素、维生素和次生代谢物生物合成中起中重要作用,而叶绿体基因组在解析植物系统发育地位、物种鉴定、进化分析等方面有着重要意义^[11]。近年来,已有部分铁线莲属植物的叶绿体完成了组装,如绣球藤、灌木铁线莲 *C. fruticosa* Turcz.、牯牛铁线莲 *C. guniuensis* W. Y. Ni, R. B. Wang et S. B. Zhou 等,明确了它们的基因组结构、序列特征、基因组成及系统发育关系等^[11-13]。本研究拟在高通量测序的基础上组装单叶铁线莲的叶绿体基因组,以明确其序列特征、基因组成及系统发育关系,为该药用植物的遗传结构和遗传多样性研究奠定基础。



A-地上部分 B-花 C-果序
A-aerial parts B-flowers C-infructescence

图1 单叶铁线莲

Fig. 1 *C. henryi* Oliv.

1 材料与仪器

1.1 材料

单叶铁线莲的叶片采自浙江省台州市黄岩区茅畲乡山卡村,台州学院蒋明教授鉴定为单叶铁线莲 *C. henryi* Oliv., 伴生植物有雀梅藤 *Sageretia thea* (Osbeck) Johnst.、鸡腿堇菜 *Viola acuminata* Ledeb.、短尾柯 *Lithocarpus brevicaudatus* (Skan) Hayata、黄檀 *Dalbergia hupeana* Hance 和秀丽野海棠 *Bredia amoena* Cogn.等。采集时间为2020年12月,将健康叶片放入取样袋,带回实验室后先用大量的自来水冲洗,再用无菌水清洗3~4次,晾干后置于-80℃冰箱备用。

1.2 仪器

移液枪(德国艾本德公司);高速冷冻离心机(美

国贝克曼库尔特有限公司);电泳仪(北京六一生物科技有限公司);凝胶成像系统(美国Bio-rad公司);Thinkpad移动工作站(联想集团有限公司);超低温冰箱(日本Sanyo电机公司)。

2 方法

2.1 DNA的提取和测序

叶片先用液氮快速冷冻,用研棒磨成粉末后采用十六烷基三甲基溴化铵(hexadecyltrimethylammonium bromide, CTAB)法提取基因组DNA,经电泳检测和浓度测定后置于-20℃冰箱备用。基因组DNA用超声波处理,经末端修复、加A尾、加测序接头、纯化、PCR扩增等步骤完成文库构建,再在Illumina高通量测序平台NovaSeq 6000上测序,获得原始数据。

2.2 叶绿体基因组的组装

经去除接头和低质量的数据, 获得 clean reads。以转子莲 *C. patens* Morr. et Decne. 的 *matK* (AY515250) 基因为种子序列, 利用 NOVOPlasty 软件的 Perl 程序拼接单叶铁线莲的叶绿体基因组, 参数采用默认值^[14]。

2.3 基因组的注释和比较

利用 Geneious prime 的 Find repeats 插件检测叶绿体基因组的反向重复 (inverted repeat, IR) 序列, 确定大单拷贝区 (large single copy, LSC) 和小单拷贝区 (small single copy, SSC) 的范围。单叶铁线莲叶绿体基因组采用 Dual Organellar GenoMe Annotator (DOGMA, <http://dogma.cccb.utexas.edu/>) 注释, 并作手工调整^[15]; tRNA 的预测采用 tRNAscan-SE (http://lowelab.ucsc.edu/tRNA_scan-SE/) 和 ARAGORN 工具^[16-17]; 最后 Organellar Genome DRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) 生成圈图, 勾选 Tidy up annotation, 其他参数采用默认值^[18]。

2.4 简单重复序列和核苷酸多态性分析

利用 MISA (MicroSAteLLite Identification Tool) 提供的 Perl 程序分析简单重复序列 (Simple sequence repeat, SSR) 位点, 程序在 WINDOWS 命令窗口运行。单核苷酸最低重复次数设定为 10, 二核苷酸最低重复次数为 5, 三核苷酸、四核苷酸、五核苷酸和六核苷酸的最低重复次数为 4^[19]。核苷酸多态性分析采用 DnaSP 5.0 软件, 窗口大小、步移尺寸分别设为 600 bp 和 200 bp。

2.5 系统发育分析

用于多重比对和系统发育分析的叶绿体基因组全长序列下载自 NCBI 数据库, 它们是槭叶铁线莲 *C. acerifolia* Maxim. (NC_039844)、曲柄铁线莲 *C. repens* Finet et Gagnep. (NC_039578)、长瓣铁线莲 *C. macropetala* Ledeb. (NC_041477)、美花铁线莲 *C. potaninii* Maxim. (NC_058760)、绣球藤 (NC_057507)、灌木铁线莲 (MT083932)、甘青铁线莲 (MK253446)、芹叶铁线莲 *C. aethusifolia* Turcz. (MK253462)、牯牛铁线莲 (MN527334)、大叶铁线莲 *C. heracleifolia* DC. (KY120885)、铁线莲 *C. florida* Thunb. (NC_058885)、转子莲 (MW900175)、柱果铁线莲 *C. uncinata* Champ. (NC_039846)、威灵仙 (MZ351432) 和圆锥铁线莲 *C. terniflora* DC. (KJ956785) 等。

MAFFT 7.388 程序用于铁线莲属植物及细叶黄

乌头 *Aconitum barbatum* Pers. 叶绿体基因组的两两比对和多重比对, 然后采用 jModelTest 获得最佳替代模型, 最后基于最大似然 (Maximum likelihood, ML) 法构建系统发育树^[20-22]。建树时以毛茛科的细叶黄乌头为外类群, 自举检测 (Bootstrap) 次数为 1000。

3 结果与分析

3.1 叶绿体基因组的结构

利用 NOVOPlasty 的 Perl 程序完成 clean reads 的拼接, 综合运用多个软件或在线工具完成了叶绿体基因组的注释。结果表明, 单叶铁线莲的叶绿体基因组为一个典型的四分体结构, 由 LSC、SSC 和 2 个 IR 组成 (图 2)。叶绿体基因组的全长为 159 707 bp, GC 值为 38.0%; LSC、SSC 和 IR 序列的长度分别为 79 449、18 100、31 079 bp, GC 值分别为 36.3%、31.4%、42.0%。

3.2 基因组成

基因注释结果表明, 单叶铁线莲叶绿体基因组共有 137 个基因, tRNA 基因的数量最多, 共 36 个, 核糖体蛋白大小亚基基因的数量其次, 为 31 个, 其中大亚基基因 14 个, 小亚基基因 17 个 (表 1)。tRNA 中, *trnA-UGC*、*trnI-CAU*、*trnI-GAU*、*trnL-CAA*、*trnN-GUU*、*trnR-ACG* 和 *trnV-GAC* 各有 2 个拷贝; 核糖体蛋白大小亚基基因中, *rps3*、*rps7*、*rps8*、*rps12*、*rps19*、*rpl2*、*rpl14*、*rpl16*、*rpl22* 和 *rpl23* 也各有 2 个拷贝; 此外, 所有核糖体 RNA 基因、*ndhB* 和 *ycf2* 有 2 个拷贝; *Ycf1* 和 *infA* 虽然各有 2 个拷贝, 但其中 1 份为假基因。*trnA-UGC*、*trnG-UCC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA*、*trnV-UAC*、*rps16*、*rpl2*、*rpl16*、*rpoC1*、*ndhA*、*ndhB*、*petB*、*petD* 和 *atpF* 各有 1 个内含子, 而 *rps12* 和 *ycf3* 有 2 个内含子。单叶铁线莲叶绿体基因组序列已上传 NCBI, 登录号为 OM650806。

3.3 SSR 检测

经 MISA 检测, 单叶铁线莲叶绿体基因组共有 41 个 SSR, 其中单核苷酸重复 40 个, 三核苷酸重复 1 个, 没有发现到其他类型的 SSR (表 2)。大部分 SSR 分布在基因间隔区 (intergenic spacer, IGS), 共 29 个, 占总数的 70.7%; 10 个出现在 *matK*、*trnG-UCC*、*rpoB*、*psbC*、*rpoA*、*ccsA* 和 *ycf1* 的编码区, 其中 *ycf1* 基因编码区内有 4 个 SSR; 另有 2 个分别位于 *atpF* 和 *ndhA* 的内含子中。SSR 出现在 LSR、SSR 和 IR 的次数分别为 25、14 和 2, 反向重复区中 SSR 仅占总数的 4.9%。

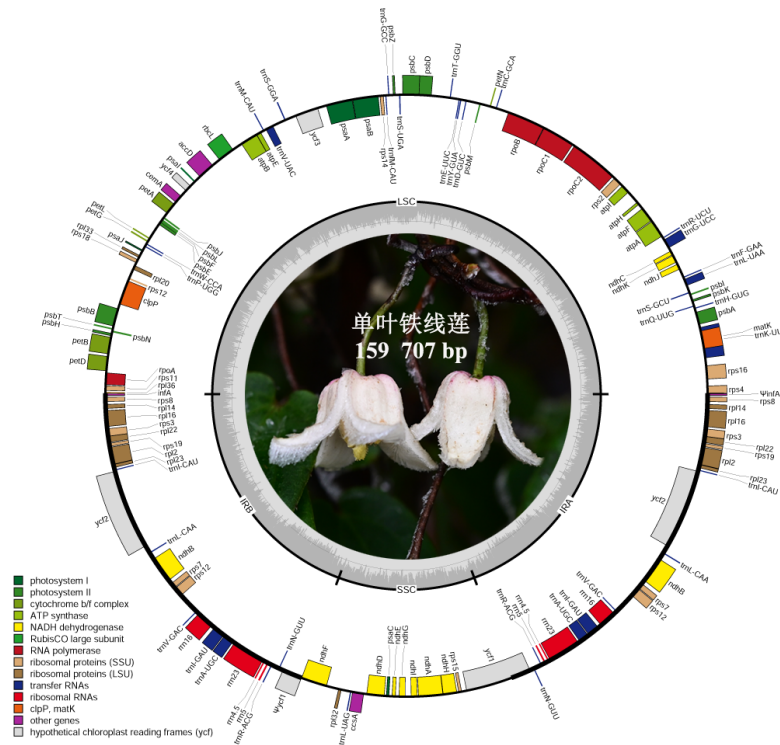


图 2 单叶铁线莲叶绿体基因组

Fig. 2 Chloroplast genome of *C. henryi*

表 1 单叶铁线莲叶绿体基因组上的基因

Table 1 Genes located on chloroplast genome of *C. henryi*

基因类别	基因名称	数量
转运 RNA	<i>trnA-UGC</i> (×2)*, <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnM-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> *, <i>trnH-GUG</i> , <i>trnI-CAU</i> (×2), <i>trnI-GAU</i> (×2)*, <i>trnK-UUU</i> *, <i>trnL-CAA</i> (×2), <i>trnL-UAA</i> *, <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> (×2), <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> (×2), <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnV-GAC</i> (×2), <i>trnV-UAC</i> *, <i>trnW-CCA</i> , <i>trnY-GUA</i>	36
核糖体 RNA	<i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)	8
核糖体蛋白小亚基	<i>rps2</i> , <i>rps3</i> (×2), <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> (×2), <i>rps11</i> , <i>rps12</i> (×2)**, <i>rps14</i> , <i>rps15</i> , <i>rps16</i> *, <i>rps18</i> , <i>rps19</i> (×2)	17
核糖体蛋白大亚基	<i>rpl2</i> (×2)*, <i>rpl14</i> (×2), <i>rpl16</i> (×2)*, <i>rpl20</i> , <i>rpl22</i> (×2), <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	14
RNA 聚合酶	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> *, <i>rpoC2</i>	4
NADH 脱氢酶亚基	<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
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 b/f 复合物亚基	<i>petA</i> , <i>petB</i> *, <i>petD</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</i> *, <i>atpH</i> , <i>atpI</i>	6
Rubisco 大亚基	<i>rbcL</i>	1
成熟酶	<i>matK</i>	1
蛋白酶	<i>clpP</i> **	1
被膜蛋白	<i>cemA</i>	1
乙酰 CoA 羧化酶	<i>accD</i>	1
细胞色素 C 合成酶	<i>ccsA</i>	1
翻译起始因子	<i>infA</i> , Ψ <i>infA</i>	1
未知功能蛋白	<i>ycf1</i> , Ψ <i>ycf1</i> , <i>ycf2</i> (×2), <i>ycf3</i> ***, <i>ycf4</i>	6
合计		137

×2: 拷贝数为 2; Ψ : 假基因; *: 1 个内含子; **: 2 个内含子

×2: two copies; Ψ : pseudogene; *: one intron; **: two introns.

表 2 单叶铁线莲叶绿体基因组 SSR 的分布

Table 2 Distribution of SSRs within *C. henryi* chloroplast genome

编号	SSR 类型	大小	起始位置	结束位置	位置
1	(T) 10	10	2 949	2 958	IGS
2	(T) 10	10	4 034	4 043	matK
3	(A) 11	11	7 109	7 119	IGS
4	(A) 11	11	7 353	7 363	IGS
5	(A) 11	11	8 082	8 092	IGS
6	(A) 11	11	8 782	8 792	IGS
7	(A) 10	10	9 018	9 027	IGS
8	(T) 10	10	9 926	9 935	IGS
9	(A) 13	13	13 044	13 056	trnG-UCC
10	(T) 10	10	16 123	16 132	atpF 内含子
11	(T) 11	11	20 495	20 505	IGS
12	(T) 10	10	30 262	30 271	rpoB
13	(T) 10	10	36 495	36 504	IGS
14	(A) 12	12	36 681	36 692	IGS
15	(G) 10	10	39 204	39 213	psbC
16	(T) 11	11	41 406	41 416	IGS
17	(T) 10	10	50 099	50 108	IGS
18	(TTG) 6	18	50 487	50 504	IGS
19	(T) 10	10	52 220	52 229	IGS
20	(T) 14	14	59 632	59 645	IGS
21	(A) 11	11	63 167	63 177	IGS
22	(T) 12	12	69 620	69 631	IGS
23	(A) 11	11	77 671	77 681	IGS
24	(T) 10	10	77 938	77 947	rpoA
25	(T) 10	10	79 464	79 473	IGS
26	(A) 10	10	79 710	79 719	IGS
27	(T) 12	12	112 940	112 951	IGS
28	(A) 10	10	115 502	115 511	ccsA
29	(A) 10	10	115 963	115 972	IGS
30	(T) 10	10	118 116	118 125	IGS
31	(T) 10	10	118 333	118 342	IGS
32	(T) 15	15	118 804	118 818	IGS
33	(A) 11	11	119 534	119 544	IGS
34	(A) 14	14	121 516	121 529	ndhA 内含子
35	(A) 10	10	123 984	123 993	IGS
36	(T) 11	11	125 503	125 513	ycf1
37	(T) 10	10	125 909	125 918	ycf1
38	(T) 10	10	127 111	127 120	ycf1
39	(T) 13	13	127 269	127 281	ycf1
40	(T) 10	10	159 438	159 447	IGS
41	(A) 10	10	159 684	159 693	IGS

3.4 核苷酸多态性分析

利用 DnaSP 软件分析了 16 种铁线莲属植物叶绿体基因组的核苷酸多态性, 结果表明, 它们的核苷酸多态性范围为 0~0.023 57, 平均值为 0.003 252。rps11、ndhF、ndhD 和 rpl32-trnI-UAG

等基因或基因间隔区具有较高的变异, 它们的核苷酸多态性均大于 0.015, 其中, 基因间隔区 rpl32-trnI-UAG 的多态性最高, ndhD 次之, 为 0.016 9, 而 rps11、ndhF 的多态性分别为 0.016 5 和 0.015 8 (图 3)。

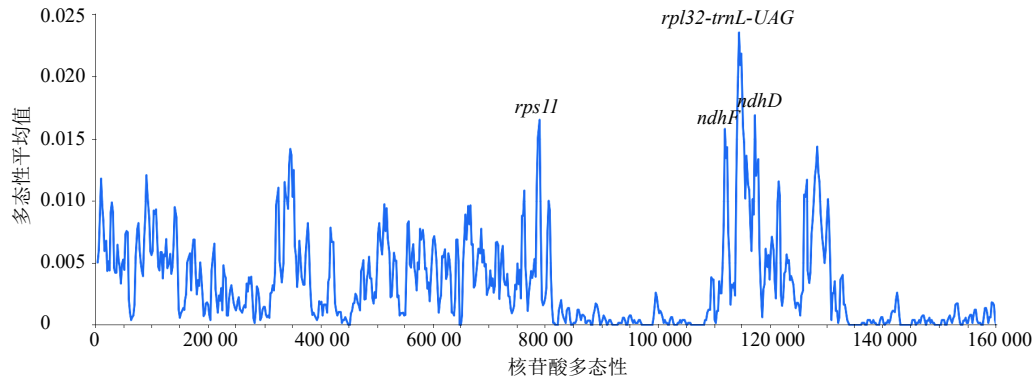


图3 核苷酸多态性的滑动窗口分析

Fig. 3 Sliding window analysis of nucleotide diversity

3.5 系统发育分析

序列比对结果表明，单叶铁线莲与牯牛铁线莲和大叶铁线莲的相似性最高，分别达 99.6%和 99.4%，与外类群细叶黄乌头的相似性最低，仅 79.8%。利用 jModelTest 工具获得分子进化模型，结果表明，GTR+G+I 为最佳替代模型，赤池信息标准 (akaike information criterion, AIC) 与贝叶斯信息标准 (bayesian information criterion, BIC) 分别为

632 080.765 08 和 632 493.356 54。系统发育分析结果表明，17 种铁线莲属植物在 ML 树上取为 4 组，绣球藤与美花铁线莲聚为一组，槭叶铁线莲、曲柄铁线莲和长瓣铁线莲聚为一组，铁线莲、转子莲、柱果铁线莲、圆锥铁线莲和威灵仙聚于另一组，单叶铁线莲先与大叶铁线莲和牯牛铁线莲聚于一个分支，支持率达 100%，它们再与灌木铁线莲、甘青铁线莲和芹叶铁线莲聚为一组，支持率为 99% (图 4)。

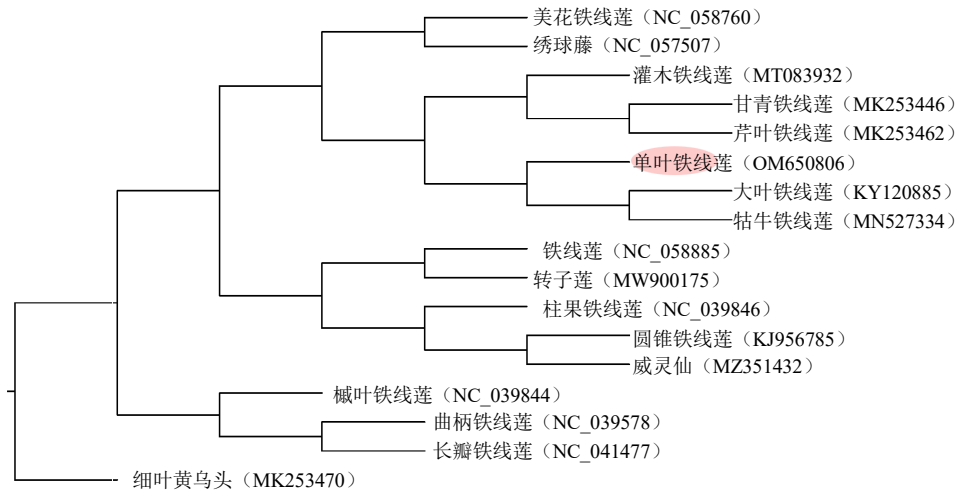


图4 基于叶绿体基因组序列构建的系统发育树

Fig. 4 A phylogenetic tree constructed based on chloroplast genome sequences

4 讨论

叶绿体是绿色植物和藻类的重要细胞器，它将光能转换成化学能，并把二氧化碳同化为碳水化合物；叶绿体起源于蓝藻，并保留很多古老的特征，叶绿体自带基因组，基因参与光合作用、硫同化、脂类代谢及淀粉、色素、激素和氨基酸等的生物合成^[23-24]。叶绿体基因组的大小通常为 120 000~170 000 bp，植物

叶绿体基因组的结构、基因的组成及排列的顺序具有一定的保守性^[25]。铁线莲属植物叶绿体基因组大小在 159 500 bp 左右，如绣球藤、灌木铁线莲和牯牛铁线莲的叶绿体基因组大小分别为 159 523、159 683、159 682 bp^[13]。本研究中，单叶铁线莲的叶绿体基因组为 159 707 bp，序列长度与已组装完成的同属物种差异不大^[11-13]。叶绿体基因组呈环状，

通常可分成4个部分,即LSC、SSC和2个反向重复区IR(IRA和IRB),但也有例外,如松科(Pinaceae)植物中IRB缺失,柏类植物中IRA丢失^[26-27]。本研究中,单叶铁线莲的叶绿体基因组为典型的四分体结构,即由LSC、SSC、IRA和IRB组成。

假基因化在叶绿体基因组中十分常见,碱基突变、插入或缺失等引起编码区产生终止密码子是假基因化较为常见的原因。在鸭跖草亚科(Commelinoideae)中,*accD*、*rpoA*和*ycf15*发生假基因化,它们由碱基插入或缺失引起,而杜若*Politia japonica* Thunb.和钩毛子草*Rhopalephora scaberrima* (Blume) Faden中*ndhB*的假基因化则由点突变引起^[28]。产生假基因的另一情况是编码区截短,通常发生在边界部分,如毛茛科乌头*A. carmichaelii* Debeaux和黄花乌头*A. coreanum* (Lévl.) Rapaics的*ycf1*位于IRA与SSC的边界,导致其假基因化^[29]。本研究中,*ycf1*有2份拷贝,它们分别位于IRB/SSC和SSC/IRA的边界,前者发生截短现象,成为假基因,另1份拷贝则正常,除*ycf1*外,2份*infA*中的1个拷贝也为假基因;类似的现象发生在牯牛铁线莲叶绿体基因组中,它的*ycf1*和*infA*均有2份拷贝,其中1份为假基因,不同的是,牯牛铁线莲的*rpl32*和*ndhK*也为假基因^[13]。*infA*是叶绿体中蛋白翻译起始的重要成分,该编码蛋白基因假基因化现象十分普遍,假基因化的主要原因为有害突变的积累,包括碱基插入、缺失和替换等^[30]。Millen等^[31]对16属17种茄科植物叶绿体基因组进行研究,发现所有物种的*infA*都是假基因,部分缺少起始密码子,3个物种发生移码框突变,而番茄*Lycopersicon esculentum* Miller叶绿体基因组的*infA*缺失前面124个碱基。在本研究中,单叶铁线莲*infA*的假基因化源于3'端序列的缺失。

SSR是指由1~6个核苷酸组成的基本单元经多次重复形成的一段DNA,它们分布在叶绿体基因组的部位。与核基因组相比,叶绿体基因组较小,SSR数量相对较少,但也可用于物种鉴定和分子辅助育种。Liu等^[32]以圆锥铁线莲叶绿体基因组为参照,开发SSR标记对11个铁线莲属植物的43个个体进行测试,发现SSR可将这些物种完全区分开。褐毛铁线莲变种*C. fusca* var. *coreana* Turcz.叶绿体基因组中有67个SSR,其中单核苷酸重复50个,二核苷酸-四核苷酸重复共17个^[33]。本研究中,单叶铁线莲叶绿体基因组共有41个SSR,数量远

少于褐毛铁线莲,另外,SSR类型也较少,仅2种,即单核苷酸重复和三核苷酸重复,其中单核苷酸重复占97.6%。SSR的数量和类型与物种相关,甚至在同一物种不同品种间也存在差异,可用于物种鉴定和亲缘关系分析等^[34]。单叶铁线莲在分类学上属于尾叶铁线莲组Sect. *Viorna*尾叶铁线莲亚组Subsect. *Connatae*,基于叶绿体基因组构建的系统发育树上,单叶铁线莲与同为尾叶铁线莲亚组的大叶铁线莲和牯牛铁线莲聚为一组,支持率达100%,分子聚类结果与形态分类一致。

本研究以药用植物单叶铁线莲叶片为材料,经过DNA提取、文库构建、高通量测序、组装和序列分析,明确了叶绿体因组的序列特征和系统发育关系,为后续开展该药用植物的遗传结构和遗传多样性研究打下了基础。

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-第十四卷 [M]. 北京: 科学出版社, 1980: 126.
- [2] 中国药典 [S]. 一部. 2020: 236.
- [3] 贾敏如, 张艺. 中国民族药辞典 [M]. 北京: 中国医药科技出版社, 2016: 36.
- [4] 张亚梅, 慕泽涇, 张普照, 等. 中国铁线莲属民族药用植物研究整理 [J]. 中国实验方剂学杂志, 2018, 24(2): 226-234.
- [5] 李杨, 张伟, 赫雪锋, 等. 铁线莲属植物的化学成分及药理作用研究进展 [J]. 中南药学, 2018, 16(3): 355-362.
- [6] 冯玲枫, 翁美娅, 田富饶, 等. 11种铁线莲属药用植物地上部分总黄酮含量比较研究 [J]. 中国现代中药, 2011, 13(7): 14-18.
- [7] 刘韶, 向大雄, 颜苗, 等. 单叶铁线莲化学成分研究 [J]. 中成药, 2007, 29(9): 1379-1380.
- [8] 孙晓博, 张凯, 李兰林, 等. 大孔吸附树脂分离纯化单叶铁线莲总皂苷的研究 [J]. 中草药, 2011, 42(10): 1977-1981.
- [9] 黄余磊, 吕枷薪, 蒋明, 等. 单叶铁线莲 *Clematis henryi* 愈伤组织诱导与植株再生 [J]. 浙江农业学报, 2011, 23(4): 731-735.
- [10] Mao C Y, Zhang X T, Shi J, et al. The complete chloroplast genome sequence of *Clematis montana* Buch. -Ham. (Ranunculaceae) and its phylogenetic analysis [J]. *Mitochondrial DNA B Resour*, 2020, 5(3): 2246-2247.
- [11] Dobrogojski J, Adamiec M, Robert L. The chloroplast genome: A review [J]. *Acta Physiol Plant*, 2020, 42(6): 98.

- [12] Yang Y C, Wang N, Zhang W, *et al.* The complete chloroplast genome of *Clematis fruticosa* Turcz. (Ranunculaceae) [J]. *Mitochondrial DNA B*, 2020, 5(2): 1908-1909.
- [13] Jiang M, Wang J F, Zhang H J. The complete plastome sequence of *Clematis guniuensis* (Ranunculaceae), a new plant species endemic to China [J]. *Mitochondrial DNA B Resour*, 2020, 5(1): 408-409.
- [14] 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.
- [15] Wyman S K, Jansen R K, Boore J L. Automatic annotation of organellar genomes with DOGMA [J]. *Bioinformatics*, 2004, 20(17): 3252-3255.
- [16] Lowe T M, Eddy S R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence [J]. *Nucleic Acids Res*, 1997, 25(5): 955-964.
- [17] Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences [J]. *Nucleic Acids Res*, 2004, 32(1): 11-16.
- [18] Lohse M, Drechsel O, Kahlau S, *et al.* OrganellarGenomeDRAW: a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. *Nucleic Acids Res*, 2013, 41: W575-W581.
- [19] Thiel T, Michalek W, Varshney R K, *et al.* Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) [J]. *Theor Appl Genet*, 2003, 106(3): 411-422.
- [20] Posada D. jModelTest: phylogenetic model averaging [J]. *Mol Biol Evol*, 2008, 25(7): 1253-1256.
- [21] Katoh K, Rozewicki J, Yamada K D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization [J]. *Brief Bioinform*, 2019, 20(4): 1160-1166.
- [22] Guindon S, Dufayard J F, Lefort V, *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0 [J]. *Syst Biol*, 2010, 59(3): 307-321.
- [23] Armbruster U, Pesaresi P, Pribil M, *et al.* Update on chloroplast research: New tools, new topics, and new trends [J]. *Mol Plant*, 2011, 4(1): 1-16.
- [24] Pogson B J, Ganguly D, Albrecht-Borth V. Insights into chloroplast biogenesis and development [J]. *Biochim Biophys Acta*, 2015, 1847(9): 1017-1024.
- [25] Smith D R. Does cell size impact chloroplast genome size? [J]. *Front Plant Sci*, 2017, 8: 2116.
- [26] Wu C S, Wang Y N, Hsu C Y, *et al.* Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny [J]. *Genome Biol Evol*, 2011, 3: 1284-1295.
- [27] Wu C S, Wang Y N, Liu S M, *et al.* Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: Insights into cpDNA evolution and phylogeny of extant seed plants [J]. *Mol Biol Evol*, 2007, 24(6): 1366-1379.
- [28] 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.
- [29] Park I, Yang S, Choi G, *et al.* The complete chloroplast genome sequences of *Aconitum pseudolaeve* and *Aconitum longecassidatum*, and development of molecular markers for distinguishing species in the *Aconitum* subgenus *Lycotoctonum* [J]. *Molecules*, 2017, 22(11): E2012.
- [30] Chandrasekaran C, Betrán E. Origins of new genes and pseudogenes [J]. *Nat Educ*, 2008, 1(1): 181.
- [31] Millen R S, Olmstead R G, Adams K L, *et al.* Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus [J]. *Plant Cell*, 2001, 13(3): 645-658.
- [32] Liu Z G, Korpelainen H. Improved characterization of *Clematis* based on new chloroplast microsatellite markers and nuclear ITS sequences [J]. *Hortic Environ Biotechnol*, 2018, 59(6): 889-897.
- [33] 陈文超, 苏琬涵, 刘志高, 等. 铁线莲属植物叶绿体微卫星引物开发及其遗传分析 [J]. *浙江农业学报*, 2017, 29(12): 2023-2031.
- [34] Tuler A C, Carrijo T T, Nória L R, *et al.* SSR markers: A tool for species identification in *Psidium* (Myrtaceae) [J]. *Mol Biol Rep*, 2015, 42(11): 1501-1513.