

八角枫及其亚种叶绿体基因组序列结构及系统发育分析

杨小英, 刘畅, 曾宪法, 刘雄伟, 赵杰宏, 俸婷婷, 周英*

(贵州中医药大学药学院, 药食两用资源应用与开发研究中心, 贵州 贵阳 550025)

摘要: 八角枫为八角枫科 Alangiaceae 八角枫属 *Alangium* 常用药用植物, 是贵州特色苗药之一, 该属植物分化强烈、形态变异复杂, 各药材活性成分和药效存在显著差异。本文利用 Illumina 高通量测序技术对八角枫亚种 (稀花八角枫、伏毛八角枫) 和毛八角枫叶绿体基因组进行测序, 完成其组装注释与结构特征解析, 并从 NCBI 下载同属植物八角枫、高山八角枫进行基因组比较分析。最后, 选择 NCBI 中已发表的与八角枫科亲缘关系较近的物种进行系统发育树构建。结果如下: 稀花八角枫、伏毛八角枫和毛八角枫叶绿体基因组全长为 156 670、156 672、156 871 bp, 为典型的四段式结构, 均包含 1 个长单拷贝区 (LSC)、1 个短单拷贝区 (SSC) 及 2 个反向重复区 (IRa 和 IRb)。均注释到 133 个基因, 其中包括 88 个蛋白编码基因、37 个 tRNA 基因和 8 个 rRNA 基因。叶绿体基因组中均发现了正向重复、回文重复和反向重复 3 种类型的长重复序列和 SSR 位点。基因组比较分析表明 5 个物种边界过渡区域存在多样性, 叶绿体基因组未发现重排或倒置, 在 *ndhA*、*ycf1*、*rpl16*、*ycf2* 及 *petD* 等基因的编码区存在显著差异, 这些位点为八角枫的分子鉴别提供了新位点。系统分析中毛八角枫单独聚为一支, 稀花八角枫和伏毛八角枫聚为一支, 支持率为 93, 亲缘关系较近, 表明系统发育树可用于物种鉴别。本文可为八角枫属分类鉴定、保证八角枫药材临床用药安全及规范八角枫药材市场提供依据。

关键词: 八角枫; 亚种; 叶绿体基因组; 重复序列; 系统发育

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2022)10-3229-11

Sequence structure and phylogenetic analysis of the chloroplast genomes of *Alangium chinense* (Lour.) Harms and its different subspecies

YANG Xiao-ying, LIU Chang, ZENG Xian-fa, LIU Xiong-wei, ZHAO Jie-hong,
FENG Ting-ting, ZHOU Ying*

(Research Center for Application and Development of Medicinal and Food Resources, College of Pharmacy,
Guizhou University of Traditional Chinese Medicine, Guiyang 550025, China)

Abstract: *Alangium chinense* is a commonly used medicinal plant of Alangiaceae *Alangium*, one of the characteristic Miao medicines in Guizhou. The genus is strongly differentiated and has complex morphological variation, with significant differences in active ingredients and potency among herbs. In this paper, we sequenced the chloroplast genomes of *Alangium chinense* subsp. Pauciflorum, *Alangium chinense* subsp. Strigosum and *Alangium kurzii* Craib using Illumina high-throughput sequencing technology. The genomes of *Alangium chinense* subsp. Pauciflorum, *Alangium chinense* subsp. Strigosum and *Alangium kurzii* Craib chloroplasts were sequenced, their

收稿日期: 2022-04-25; 修回日期: 2022-05-24.

基金项目: 国家重点研发计划项目 (2018YFC1708100); 贵州省科技计划项目 (黔科合平台人才[2019]5407); 贵州省“十四五”中医药、民族医药重点学科 [QZYYZDXK(JS)-2021-03]; 贵州省中医药管理局中医药、民族医药科学技术研究课题 (QZYY-2021-098); 贵州中医药大学 2021 年度科研创新和探索专项 (2018YFC170810101, 2018YFC170810201).

*通讯作者 E-mail: yingzhou71@126.com

DOI: 10.16438/j.0513-4870.2022-0490

assembly annotation and structural characterization were completed, and the genomes of the congeners *Alangium chinense* and *Alangium alpinum* were downloaded from NCBI for comparative analysis. Finally, the phylogenetic tree was constructed by selecting published species with close affinity to *Alangium chinense* family from NCBI. The results were as follows: *Alangium chinense* subsp. Pauciflorum, *Alangium chinense* subsp. Strigosum and *Alangium kurzii* Craib chloroplast genomes were 156 670, 156 672 and 156 871 bp in length, with a typical four-segment structure, all containing one long single copy region (LSC), one short single copy region (SSC) and two inverted repeat regions (IRa and IRb). All of them were annotated to 133 genes, including 88 protein-coding genes, 37 tRNA genes and 8 rRNA genes. Three types of long repeat sequences and SSR loci, forward repeats, palindrome repeats and reverse repeats, were found in all chloroplast genomes. Comparative genomic analysis showed that there was diversity in the boundary transition regions of the five species, no rearrangements or inversions were found in the chloroplast genome, and there were significant differences in the coding regions of *ndhA*, *ycf1*, *rpl16*, *ycf2*, and *petD* genes, which provided new loci for molecular identification of Patagonia. In the phylogenetic analysis, *Alangium kurzii* var. *kurzii* clustered as a single species, and *Alangium chinense* subsp. Pauciflorum and *Alangium chinense* subsp. Strigosum clustered as a single species, with a support rate of 93 and close affinity, indicating that the phylogenetic tree can be used for the species identification. This paper can provide a basis for the taxonomic identification of *Alangium*, ensure the safety of clinical use of anise herbs and regulate the market of *Alangium chinense* herbs.

Key words: *Alangium chinense* (Lour.) Harms; subspecies; chloroplast genome; repeat sequence; phylogenetic

八角枫为八角枫 *Alangium chinense* (Lour.) Harms 及瓜木 *Alangium platani folium* (Sieb. Et Zucc.) Harms 的干燥细须根 (白龙须) 或干燥支根 (白金条)^[1], 始载于清·罗思举《《简易草药》》, 别名八角金盘 (《本草从新》)、五角枫 (《简易草药》)、木八角 (《本草纲目拾遗》) 等^[2]; 其性温味辛、苦, 归肝、肾、心经, 具有祛风除湿、舒筋通络及散瘀止痛等功效^[3,4], 临床常用于治疗关节炎、外伤出血和跌打损伤等病症; 对中枢神经系统、呼吸系统、平滑肌、心血管系统等均有良好的作用^[5]; 为贵州地区苗族特色药, 分布于贵州各地, 东南亚及非洲东部各国也有分布^[6]。《中国植物志》记载八角枫共 4 个亚种, 分别为八角枫 (原亚种) *Alangium chinense* (Lour.) Harms subsp. *chinense*、伏毛八角枫 (亚种) *Alangium chinense* subsp. *strigosum*、稀花八角枫 (亚种) *Alangium chinense* subsp. *pauciflorum* 和深裂八角枫 (亚种) *Alangium chinense* subsp. *triangulare*^[7]。八角枫药用范围较广, 在《贵州苗药·兴仁卷》中记载了 36 种含有八角枫的民用处方, 且作为“消痹灵合剂”“风湿定片”等制剂的主药^[8]; 其传统的功能主治疗效确切, 极具研究开发前景。由于八角枫具有毒藜碱成分^[9], 该类成分对人体具有毒性, 为确保用药的安全性, 对八角枫植物准确的鉴定显得尤为重要。

叶绿体 (chloroplast, cp) 起源于古细菌入侵植物细胞^[10], 是一种拥有自身遗传物质的多功能细胞器^[11], 为质体 (plastid) 中的一种类型, 基因组大小多为 150~200 kb, 且结构较为保守^[12], 在光合作用、氨基酸的合成及固碳中起到重要作用^[13]; 其全基因组包含大量遗

传信息, 被广泛应用于植物分子进化及系统发育的研究, 并在药用植物的遗传转化和分子选育等方面发挥着重要作用^[14]。叶绿体基因组富含 SSR 位点, 与核基因组相比^[15], 叶绿体基因组在基因组成、基因含量及排列顺序上高度保守^[16], 可作为“超级条形码”进行物种鉴定, 在中药材分类鉴定研究中得到广泛应用^[17,18]。

目前, 八角枫属物种的叶绿体基因组研究十分缺乏, NCBI 数据库中有八角枫 (MG524996) 和高山八角枫 (MG525003) 完整叶绿体基因组数据, 但是仅记录叶绿体基因组总长 (分别为 157 889 bp、156 673 bp) 等基础信息, 未进行深入挖掘研究, 因此, 本研究利用 Illumina HiSeq 高通量测序技术对八角枫的两个亚种 (稀花八角枫和伏毛八角枫) 与毛八角进行叶绿体基因组测序分析, 与既有的八角枫、高山八角枫叶绿体基因组比对, 通过该属的叶绿体基因组进行系统进化树的构建, 探讨八角枫属的系统进化与亲缘关系, 丰富该属物种的遗传数据信息, 以期八角枫属植物的鉴定、叶绿体基因组的变异及进一步研究其系统发育进化提供一定的参考, 保证八角枫药材用药安全、规范八角枫药材市场及保护八角枫种质资源提供依据。

材料与方法

材料与数据 稀花八角枫采自贵州省册亨县 (海拔 632 m, 105°46'19"E, 24°58'53"N), 伏毛八角枫采自贵阳市云岩区云关乡 (海拔 1 116 m, 106°46'24"E, 21°25'7"N), 毛八角枫采自清镇市九龙山 (海拔 1 230 m, 106°29'50"E, 26°40'26"N), 经贵州中医药大学魏升

华教授鉴定为稀花八角枫 *Alangium chinense* subsp. *Pauciflorum*、伏毛八角枫 *Alangium chinense* subsp. *Strigosum*、毛八角枫 *Alangium kurzii* Craib。取新鲜健康叶片, 用无菌水冲洗干净, 晾干后置于 $-80\text{ }^{\circ}\text{C}$ 冰箱保存。通过 NCBI (The National Center for Biotechnology Information, 美国国家生物信息中心) 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中检索了 28 种植物的叶绿体全基因组序列信息 (表 1)。

Table 1 The GenBank accession numbers for the chloroplast genomes of 28 species

Species	Family	Genus	Accession
<i>Alangium alpinum</i>	Alangiaceae	<i>Alangium</i>	MG525003
<i>Alangium chinense</i>	Alangiaceae	<i>Alangium</i>	MG524996
<i>Alangium Kurzii</i>	Alangiaceae	<i>Alangium</i>	MZ680502
<i>Nyssa sinensis</i>	Nyssaceae	<i>Nyssa</i>	KX904873
<i>Camptotheca acuminata</i>	Nyssaceae	<i>Camptotheca</i>	KX904871
<i>Syzygium aromaticum</i>	Myrtaceae	<i>Syzygium</i>	MN746306
<i>Rhodomyrtus tomentosa</i>	Myrtaceae	<i>Rhodomyrtus</i>	MK044696
<i>Psidium guajava</i> L.	Myrtaceae	<i>Psidium</i>	KY635879
<i>Trapa natans</i>	Trapaceae	<i>Trapa</i>	MZ617461
<i>Trapa bispinosa</i>	Trapaceae	<i>Trapa</i>	MW065552
<i>Trapa maximowiczii</i>	Trapaceae	<i>Trapa</i>	KY705084
<i>Stellera chamaejasme</i>	Thymelaeaceae	<i>Stellera</i>	MK681211
<i>Wikstroemia capitata</i>	Thymelaeaceae	<i>Wikstroemia</i>	MW073909
<i>Daphne genkwa</i>	Thymelaeaceae	<i>Daphne</i>	MT754180
<i>Daphne tangutica</i>	Thymelaeaceae	<i>Daphne</i>	MK455800
<i>Daphne retusa</i>	Thymelaeaceae	<i>Daphne</i>	MW245832
<i>Edgeworthia chrysantha</i>	Thymelaeaceae	<i>Edgeworthia</i>	MT135123
<i>Aquilaria yunnanensis</i>	Thymelaeaceae	<i>Aquilaria</i>	MG656407
<i>Aquilaria sinensis</i>	Thymelaeaceae	<i>Aquilaria</i>	MN720647
<i>Hippophae rhamnoides</i>	Elaeagnaceae	<i>Hippophae</i>	KY794808
<i>Hippophae neurocarpa</i>	Elaeagnaceae	<i>Hippophae</i>	MK787303
<i>Hippophae goniocarpa</i>	Elaeagnaceae	<i>Hippophae</i>	MW791524
<i>Hippophae tibetana</i>	Elaeagnaceae	<i>Hippophae</i>	MN643620
<i>Hippophae gyantsensis</i>	Elaeagnaceae	<i>Hippophae</i>	MK552375
<i>Hippophae salicifolia</i>	Elaeagnaceae	<i>Hippophae</i>	MW392804
<i>Elaeagnus multiflora</i>	Elaeagnaceae	<i>Elaeagnus</i>	LC522136
<i>Elaeagnus glabra</i>	Elaeagnaceae	<i>Elaeagnus</i>	LC522137
<i>Elaeagnus angustifolia</i>	Elaeagnaceae	<i>Elaeagnus</i>	MK131395

基因组 DNA 提取与测序 利用植物基因组 DNA 提取试剂盒提取样品总 DNA。0.8% 琼脂糖电泳检测 DNA 样品是否有降解及杂质, 并估算 DNA 浓度, Nanodrop 分光光度计检测样品浓度及纯度, 然后通过 Illumina HiSeq PE150 构建插入片段为 350 bp 的文库, 采用 Illumina Novaseq6000 高通量测序平台进行双末端测序, 测序策略为 PE150 (Pair-End 150)。

叶绿体全基因组序列组、拼接与注释 采用 SPAdes v.3.15.2 软件 (<http://cab.spbu.ru/software/spades/>) 进行拼接组装^[19], 并采用 PCR 与 Sanger 测序的方法进行验证。利用 CpGAVAS^[20]和 ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) 对叶绿体基因组进行注释。运用 Blastn 和 Blastp 的方法与已报道的近缘物种的叶绿体

基因组的编码蛋白和 rRNA 进行比对, 验证结果的准确性并进行修正。使用 ARWEN^[21] (Version 1.2, <http://mbio-serv2.mbioekol.lu.se/ARWEN/>) 进行 tRNA 的注释, 生成 tRNA 二级结构图。最后利用 OGDRAW 软件进行叶绿体基因组物理图谱的绘制^[22,23]。

重复序列和密码子偏好分析 利用 REPuter 软件 (<http://bibiserv.techfak.uni-bielefeld.de/reputer/>) 对 3 个八角枫属物种的长重复序列进行分析^[24], 海明距离 (Hamming distance)、最大计算重复次数 (maximum computed P repeats)、最小重复片段大小 (minimal repeat size) 参数设置为 3、5 000、30。利用 Perl 语言在线软件 MISA^[25] (<http://pgrc.ipk-gatersleben.de/misa/>) 对叶绿体基因组的 SSR (simple-sequence-repeat) 序列进行分析, 参数阈值设置为: 1、2、3、4、5、6, 核苷酸参数为 8、4、4、3、3、3, 且两个 SSRs 之间的距离不小于 100 bp。采用软件 CodonW (<http://codonw.sourceforge.net>) 分析密码子使用情况^[26]。

叶绿体全基因组比较及共线性分析 用 IRscope 工具 (<https://irscope.shinyapps.io/irapp/>) 进行可视化, 比对叶绿体基因组的 4 个区域的边界信息, 并观察收缩与扩张在 IRs 区域的现象, 包括稀花八角枫、伏毛八角枫、毛八角枫、八角枫和高山八角枫。采用 mVISTA 软件^[27], 并选用检测基因重排和倒位的全局比对模式 (Shuffle-LAGAN), 以稀花八角枫为参考基因组, 对八角枫属其余 4 种叶绿体全基因组序列同源性进行了比较研究。采用 Geneious 10.2.2 软件中 Mauve 多重基因组比对法对八角枫的叶绿体基因组进行共线性比较。

系统进化分析 本研究选择已发表的八角枫科植物为内类群和与八角枫科亲缘关系较近的为外类群, 利用 MAFFT 7 (<https://mafft.cbrc.jp/alignment/server/>) 在线软件对其叶绿体基因组进行全基因组比对, 并利用 MEGA11 软件, 使用邻接法 (NJ 法: neighbor-joining) 构建系统进化树。同时, 采用 IQ-TREE 2.0.5 (<http://www.iqtree.org/>) 软件^[28], 最大似然法 (ML 法: maximum likelihood method) 构建系统发育树, 构树参数设置为: -m MFP -B 1 000 -alrt 1 000, 优构树模型为: GTR+F+R2。

结果

1 八角枫的叶绿体基因测序结果分析

稀花八角枫、伏毛八角枫、毛八角枫叶绿体基因组图谱示于图 1。叶绿体基因组结构为双链环状, 具有典型的 4 部分结构, 包括一段大的单拷贝区 LSC (large single-copy) 和一段小的单拷贝区 SSC (small single-copy), 以及将这两段分开的序列相同、方向相反的两个反向重复区 (inverted repeat region, IRa 和 IRb)^[29]。

稀花八角枫、伏毛八角枫、毛八角枫的叶绿体基因组范围在 156 670~156 871 bp, 其中最大的是毛八角枫, 最小的是稀花八角枫, 最大与最小差为 201 bp。

2 叶绿体基因组特征比较分析

2.1 叶绿体基因组结构特征比较分析 稀花八角枫、伏毛八角枫、毛八角枫的 SSC 长为 18 573~18 593 bp, LSC 长为 86 179~86 386 bp, 2 个相同的反向重复序列 IRa 和 IRb 长为 25 950~25 956 bp; IR 区均有 38 个基因。稀花八角枫和伏毛八角枫叶绿体基因组的 GC 含量均为 37.69%, 无差异; 毛八角枫的叶绿体基因组的 GC 含量均为 37.70%; 稀花八角枫、伏毛八角枫、毛八角枫的叶绿体基因数量无差异, 均注释到 133 个基因, 且基因的注释结果相同, 蛋白编码基因较多, 为 88 个, rRNA 为 8 个, tRNA 数量为 37 个 (表 2)。

2.2 毛八角枫叶绿体基因组功能及分类 根据功能把毛八角枫叶绿体基因组分为 3 大类: 分别是与光合

Table 2 The basic information of chloroplast genomes of the three *Alangium* species

Genome feature	<i>Alangium chinense</i>	<i>Alangium chinense</i>	<i>Alangium kurzii</i>
	subsp. Pauciflorum	subsp. Strigosum	Craib
Total length/bp	156 670	156 672	156 871
GC content/%	37.69	37.69	37.7
LSC length/bp	86 179	86 179	86 386
SSC length/bp	18 591	18 593	18 573
IR length /bp	25 950	25 950	25 956
Number of genes	133	133	133
Number of protein-coding genes	88	88	88
Number of rRNA genes	8	8	8
Number of tRNA genes	37	37	37
Number of genes in IR	38	38	38

作用有关的基因、复制有关的基因、未知功能的蛋白质基因。在这些基因中, 有 7 个蛋白质编码基因 (*ndhB*、*rpl2*、*rpl23*、*rps12*、*rps7*、*ycf15*、*ycf2*)、7 个 tRNA 编码基

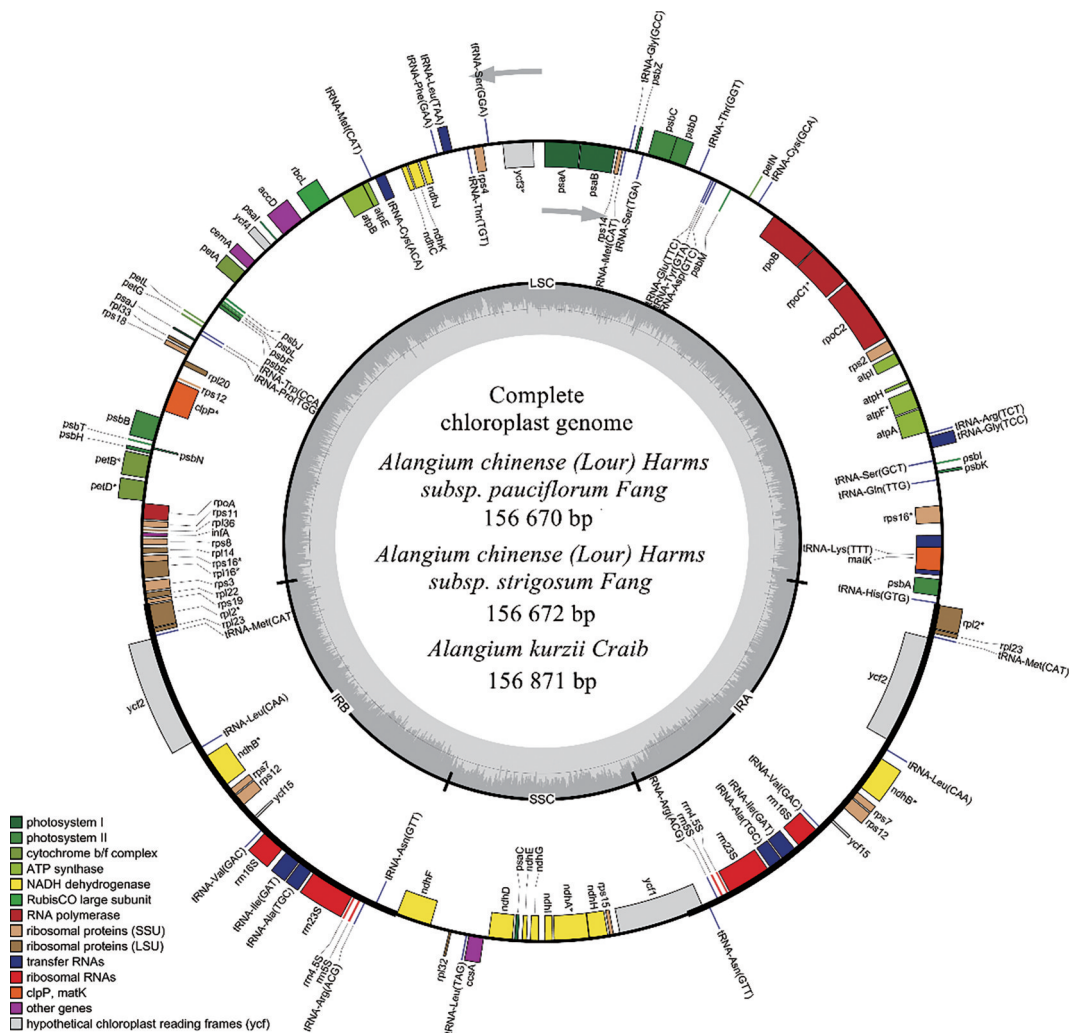


Figure 1 Chloroplast genome map of the three *Alangium* species. The length of the chloroplast genome was shown in the circle. Genes inside the circle are transcribed clockwise, and those outside are transcribed counter clockwise. Darker and lighter areas in the inner circle indicate GC content and AT content, respectively

因 [*trnA-MET* (CAT)、*trnA-Leu* (CAA)、*trnA-Val* (GAC)、*trnA-Ile* (GAT)、*trnA-Ala* (TGC)、*trnA-Arg* (ACG)、*trnA-Asn* (GTT)] 和 4 个 rRNA 编码基因 (*rrn16S*、*rrn23S*、*rrn4.5S*、*rrn5S*) 位于 IR 区 (表 3)。稀花八角枫、伏毛八角枫和毛八角枫基因组类似。

2.3 叶绿体基因组的重复序列分析 叶绿体基因组含有丰富的重复序列, 广泛的应用于群体遗传结构和物种鉴定等研究^[30,31]。利用 REPuter 在线工具对长重复序列进行分析, 发现稀花八角枫、伏毛八角枫、毛八角枫皆包含正向重复 (forward repeats, FR)、回文重复 (palindrome repeats, PR)、反向重复 (reverse repeats, RR), 但其数量稍有不同 (表 4), 稀花八角枫和伏毛八角枫叶绿体基因组共鉴定到 63 个长重复序列, 包括 30 个正向重复 (18 个 30~39 bp, 6 个 40~49 bp, 2 个 50~59 bp, 4 个 60~69 bp)、31 个回文重复 (19 个 30~39 bp, 6 个 40~49 bp, 2 个 50~59 bp, 4 个 60~69 bp)、2 个反向重复 (30~39 bp); 毛八角枫叶绿体基因组共鉴定 60 个长重复序列, 包括 30 个正向重复 (18 个 30~39 bp, 6 个 40~49 bp, 2 个 50~59 bp, 4 个 60~69 bp)、29 个回文重复 (18 个 30~39 bp, 5 个 40~49 bp, 2 个 50~59 bp, 4 个 60~69 bp)、1 个反向重复 (30~39 bp)。在稀花八角枫、伏毛八角枫、毛八角枫叶绿体基因组长重复序列数量分布中, 大多数重复序列长度都在 30~39 bp 之间, 且毛八角枫叶绿体基因组有 1 个反向重复, 均未鉴定到互补重复序列。

Table 3 Gene composition in chloroplast genomes of the *Alangium kurzii* Craib. *Contains one intron; **Contains two introns; rpl: Ribosomal protein large; rps: Ribosomal protein subunit; rpo: RNA polymerase; psa: Photosynthetic apparatus; ndh: NADH dehydrogenase; pet: Polypeptide; rbc: Ribulose biphosphate carboxylase; mat: Mature enzyme gene; clp: Caseinolytic protease; cem: Chloroplast envelope membrane protein; acc: Acetyl-CoA carboxylase; ccs: TypeC cytochrome synthesis gene; ycf: Hypothetical chloroplast openreading frame

Category	Gene group	Gene name
Self-replication	Large subunit of ribosome	<i>rpl2**</i> (x2), <i>rpl14</i> , <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (x2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	Small subunit of ribosome	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (x2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> (x2), <i>rps14</i> , <i>rps15</i> , <i>rps16**</i> (x2), <i>rps18</i> , <i>rps19</i>
	RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1*</i> , <i>rpoC2</i>
	rRNA	<i>rrn4.5S</i> (x2), <i>rrn5S</i> (x2), <i>rrn16S</i> (x2), <i>rrn23S</i> (x2)
	tRNA	<i>tRNA-Val</i> (x2), <i>tRNA-Tyr</i> , <i>tRNA-Trp</i> , <i>tRNA-Thr</i> (x2), <i>tRNA-Ser</i> (x3), <i>tRNA-Pro</i> , <i>tRNA-Phe</i> , <i>tRNA-Met</i> (x4), <i>tRNA-Lys</i> , <i>tRNA-Leu</i> (x4), <i>tRNA-Ile</i> (x2), <i>tRNA-His</i> , <i>tRNA-Gly</i> (x2), <i>tRNA-Glu</i> , <i>tRNA-Gln</i> , <i>tRNA-Cys</i> (x2), <i>tRNA-Asp</i> , <i>tRNA-Asn</i> (x2), <i>tRNA-Arg</i> (x3), <i>tRNA-Ala</i> (x2)
Photosynthesis	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
	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>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
	NADH dehydrogenase	<i>ndhA*</i> , <i>ndhB**</i> (x2), <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>
	Cytochrome b/f complex	<i>petA</i> , <i>petB*</i> , <i>petD*</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>
	Large subunit of rubisco	<i>rbcL</i>
Other genes	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
	Protease	<i>clpP*</i>
	Envelope membrane protein	<i>cemA</i>
	Subunit of Acetyl-carboxylase	<i>accD</i>
	C-type cytochrome synthesis	<i>ccsA</i>
	Open reading frames	<i>ycf1</i> , <i>ycf2</i> (x2), <i>ycf3*</i> , <i>ycf4</i> , <i>ycf15</i> (x2)

Table 4 Repeat sequences in the chloroplast genomes of the three *Alangium* species

Species	<i>Alangium chinense</i> subsp. Pauciflorum	<i>Alangium chinense</i> subsp. Strigosum	<i>Alangium kurzii</i> Craib
Palindrome repeats	31	31	29
Forward repeats	30	30	30
Reverse repeats	2	2	1
Complement repeats	0	0	0
Total	63	63	60

简单重复序列 (simple sequence repeat, SSRs), 通常以 1~6 个碱基串联重复序列为微卫星序列, 在整个叶绿体基因组中均有分布。SSR 广泛分布于叶绿体基因组中, 通常被用作物种鉴定的重要分子标记^[32], SSR 标记还因其共显性遗传、多等位基因性质, 已经成为品种鉴定和 DNA 指纹图谱构建的重要途径^[33]。

稀花八角枫、伏毛八角枫的叶绿体基因组均包括 5 种类型的 SSRs, 以单核苷酸类最多均为 122 个, 其余的 4 种核苷酸重复序列均相同, 二核苷酸、三核苷酸、四核苷酸、五核苷酸重复序列分别 50、9、4、1 个。毛八角枫叶绿体基因组包含 4 种类型的 SSRs, 未检测到五核苷酸重复序列, 以单核苷酸类的重复序列最多为 120 个 (表 5)。

2.4 叶绿体基因组内含子分析 稀花八角枫、伏毛八

Table 5 Types and amounts of SSRs in the chloroplast genomes of the three *Alangium* species

SSR type	Repeat type	<i>Alangium chinense</i> subsp. Pauciflorum	<i>Alangium chinense</i> subsp. Strigosum	<i>Alangium kurzii</i> Craib
		Mono	A/T	31
Di	AT/AT	7	7	7
Tri	AAG/CTT	2	2	2
	AAT/ATT	6	6	4
	AGC/CTG	1	1	1
Tetra	AAAT/ATTT	2	2	2
	AATT/AATT	1	1	0
	ACAT/ATGT	1	1	1
Penta	AAGAC/CTTGT	1	1	0
Total		52	55	53

角枫、毛八角枫的叶绿体基因组中内含子编码序列 (coding sequences, CDS) 长度和基因全长 (gene length) (表6)。稀花八角枫、伏毛八角枫、毛八角枫的叶绿体基因组中共检测到 10 个内含子的蛋白编码基因: *rps16*、*atpF*、*rpoC1*、*ycf3*、*clpP*、*petB*、*petD*、*rpl2*、*ndhB* 和 *ndhA*。稀花八角枫与伏毛八角枫的编码序列长度与基因长度没有改变, 毛八角枫的编码序列长度和基因长度与其他两个存在少量变化。

3 叶绿体基因组密码子偏好性分析

不同物种基因组中同义密码子使用频率 (relative synonymous codon usage, RSCU) 不同, 即存在密码子使用偏向性, 可以为物种进化研究提供重要的信息^[34]。根据 RSCU 值计算稀花八角枫、伏毛八角枫和毛八角枫叶绿体基因组蛋白编码基因的密码子使用模式, 密

码子 RSCU 值等于 1.00, 表示该密码子无偏好性, RSCU 值大于 1.00 则表示该密码子使用频率较高。

八角枫叶绿体基因组密码子偏好性分析结果示于表7, 结果表明, 稀花八角枫、伏毛八角枫和毛八角枫的叶绿体基因组均包含 64 种密码子, 其中 61 种密码子编码 21 种氨基酸, 其余 3 个为终止密码子。在毛八角枫所有氨基酸密码子中, 编码亮氨酸 (Leu) 的密码子数最多, 蛋氨酸 (Met) 的密码子数最少, RSCU > 1 的密码子共 30 个, 其中以 A/U 结尾的密码子 28 个, RSCU ≤ 1 的密码子共 34 个, 其中 30 个为 G/C 结尾的密码子; 在伏毛八角枫所有氨基酸密码子中, 编码亮氨酸 (Leu) 的密码子数最多, 蛋氨酸 (Met) 的密码子数最少, RSCU > 1 的密码子共 30 个, 其中以 A/U 结尾的密码子 27 个, RSCU ≤ 1 的密码子共 34 个, 其中 29 个为 G/C 结尾的密码子; 在稀花八角枫所有氨基酸密码子中, 编码亮氨酸 (Leu) 的密码子数最多, 蛋氨酸 (Met) 的密码子数最少, RSCU > 1 的密码子共 32 个, 其中以 A/U 结尾的密码子 27 个, RSCU ≤ 1 的密码子共 32 个, 其中 27 个为 G/C 结尾的密码子。这些结果表明 3 个八角枫叶绿体基因组对 A/U 结尾密码子的偏好性高于 G/C 结尾密码子。

4 IR-LSC/SSC 区的收缩和扩张分析

叶绿体基因组是由 SSC、LSC、IRa、IRb 4 个区间构成的环状结构, 其中 IR 区域相对保守, 其收缩与扩张的现象代表着植物的进化, 同时也会引起叶绿体基因的 GC 含量的变化, 不同的植物的叶绿体基因组大小也与其有关^[35]; 通过分析 IR-LSC/SSC 区域的边界信息, 对研究叶绿体基因组结构的差异、物种进化等具有

Table 6 Coding sequences (CDS) length and gene length of intron in the chloroplast genomes of the three *Alangium* species

Intron gene	<i>Alangium chinense</i> subsp. Pauciflorum	<i>Alangium chinense</i> subsp. Strigosum	<i>Alangium kurzii</i> Craib
<i>rps16</i>	CDS length/bp	258/411	258/411
	Gene length/bp	411/1 113	411/1 113
<i>atpF</i>	CDS length/bp	567	567
	Gene length/bp	1 266	1 266
<i>rpoC1</i>	CDS length/bp	2 061	2 061
	Gene length/bp	2 808	2 808
<i>ycf3</i>	CDS length/bp	507	507
	Gene length/bp	1 944	1 944
<i>clpP</i>	CDS length/bp	588	588
	Gene length/bp	2 051	2 051
<i>petB</i>	CDS length/bp	648	648
	Gene length/bp	1 412	1 412
<i>petD</i>	CDS length/bp	483	483
	Gene length/bp	1 229	1 229
<i>rpl2</i>	CDS length/bp	825/828	825/828
	Gene length/bp	1 490/1 490	1 490/1 490
<i>ndhB</i>	CDS length/bp	1 533/1 533	1 533/1 533
	Gene length/bp	2 212/2 212	2 212/2 212
<i>ndhA</i>	CDS length/bp	1 092	1 092
	Gene length/bp	2 207	2 207

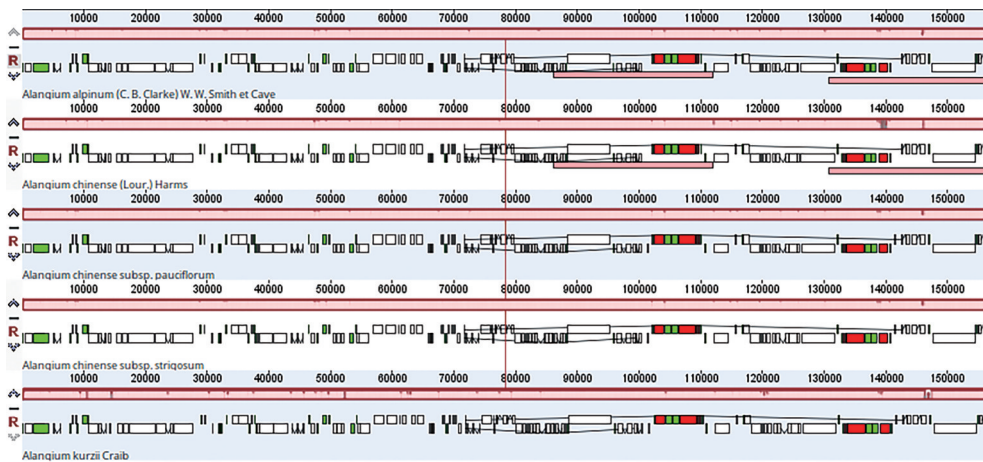


Figure 3 Collinearity analysis of chloroplast genomes from five *Alangium* species

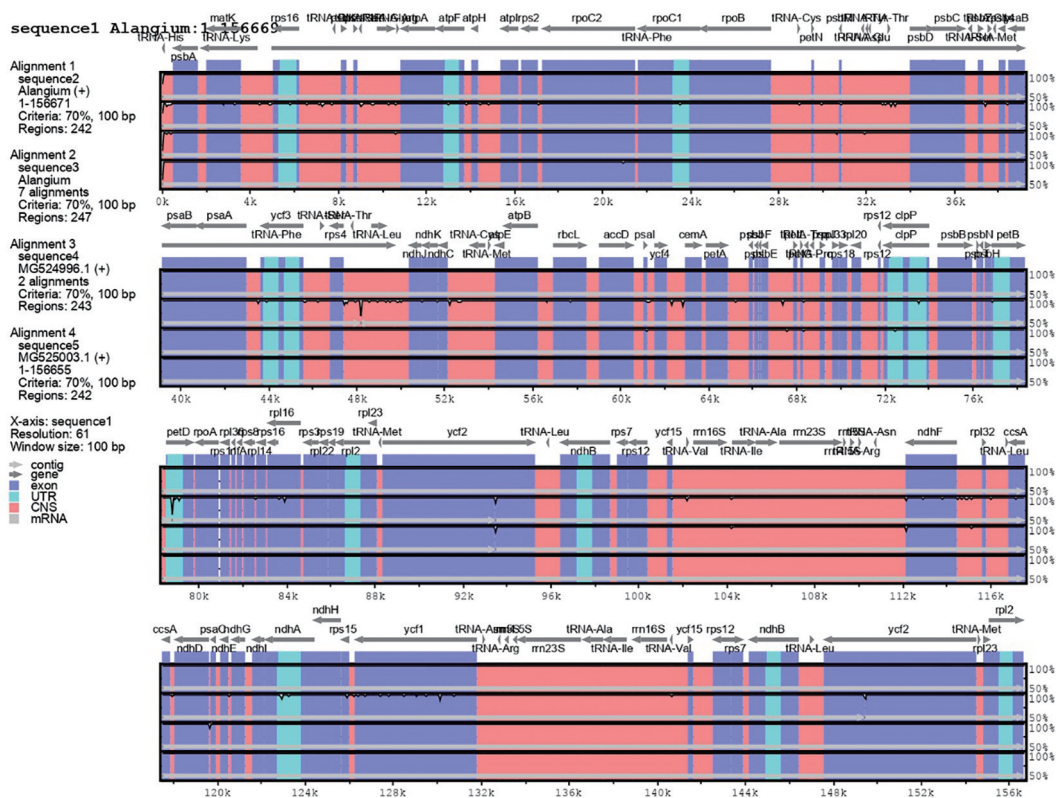


Figure 4 Global alignment in the chloroplast genomes of the five *Alangium* species. The top arrows show transcription direction; blue color indicates the exons of protein-coding genes (exon); pink color shows conserved non-coding sequence (CNS); The X-axis represents the positions in the cp genome while Y-axis represents percentage identity within 50%–100%

显的差异, 变异程度较高, 其中LSC区的变异程度最高; 而在基因编码区差异不明显, 在IRa区的变异程度最低, 具有较高的保守性。稀花八角枫、伏毛八角枫和毛八角枫叶绿体基因组中在 *ndhA*、*ycf1*、*rpl16*、*ycf2* 以及 *petD* 等基因的编码区存在显著差异。

7 系统发育分析

根据本研究中稀花八角枫、伏毛八角枫和毛八角枫, 选择已发表的八角枫科植物为内类群, 并选择了来

自蓝果树科、桃金娘科、瑞香科等共28个物种的叶绿体基因组为外类种群, 分别采用最大似然法 (ML法) 和邻接法 (NJ法) 构建了31种植物的系统发育树。ML法系统发育树显示 (图5), 各科的物种都聚成一枝; 进化树31个物种共有28个节点, 可确定为3个主要分支; 稀花八角枫、伏毛八角枫、高山八角枫、八角枫与毛八角枫聚为一小支, 且支持率为100%, 其中稀花八角枫与伏毛八角枫组成了小分支, 说明两者亲缘关系密

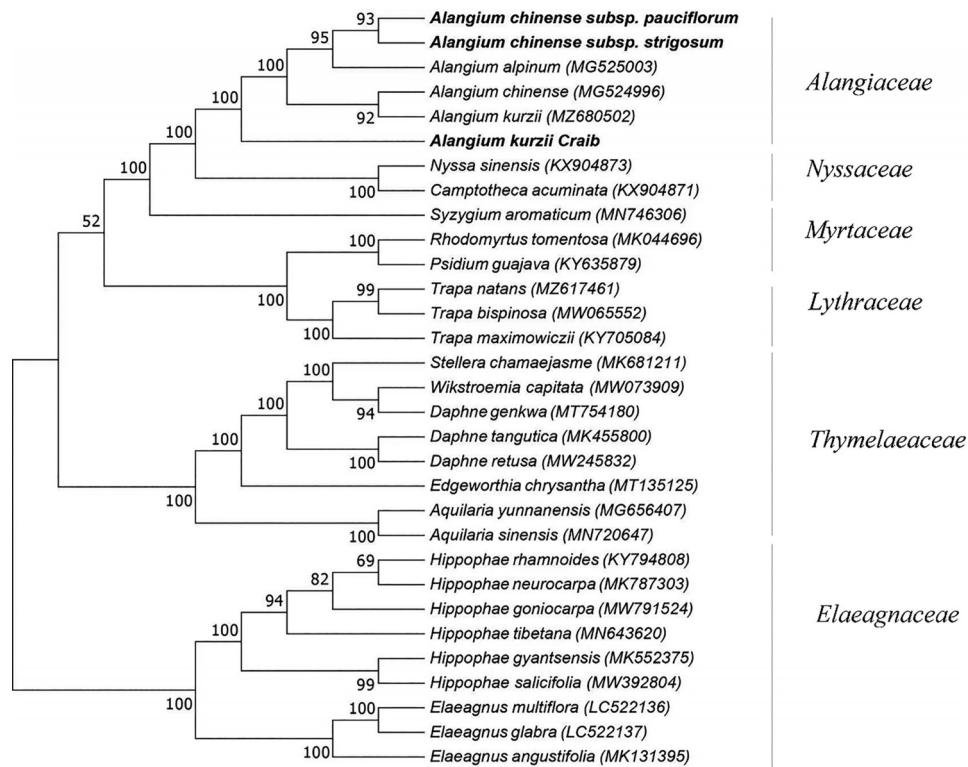


Figure 5 Maximum likelihood method (ML) phylogenetic tree was constructed based on the whole chloroplast genome sequence

切, 暗示它们在进化上具有共线性, 同时也表明叶绿体基因组构建的发育树可用于八角枫的鉴定。邻接法的结果与最大似然法构树结果基本一致, 各科的物种都聚成一支, 仅在科之间存在细微差异 (图6)。

讨论

稀花八角枫、伏毛八角枫和毛八角枫的叶绿体基因组分析表明, 叶绿体基因组均为双链环形结构, 均具有4个边界 (1个SSC、1个LSC、1个IRa和1个IRb),

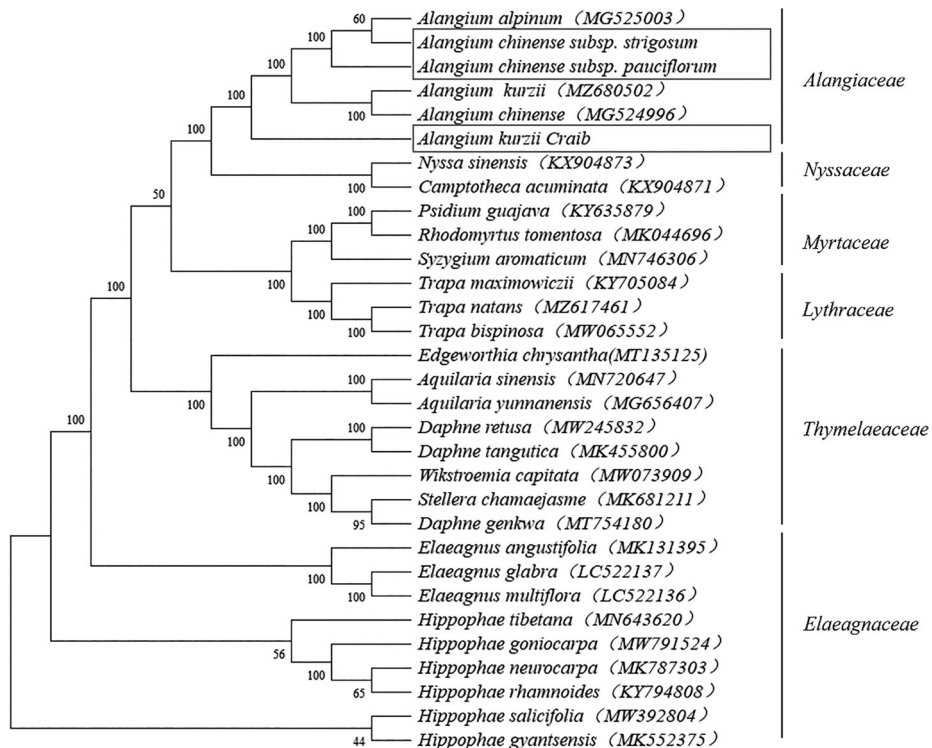


Figure 6 Neighbor-joining (NJ) phylogenetic tree was constructed based on the whole chloroplast genome sequence

叶绿体基因组长度在 156 kb 左右, 最大和最小差为 201 bp。SSC 长度在 18 kb 左右, LSC 长度在 86 kb 左右, 而两个相同的反向重复序列 IRa 和 IRb 的长度在 25 kb 左右, 均编码 133 个基因, 主要是蛋白编码基因数目居多; 利用 REPuter 在稀花八角枫、伏毛八角枫和毛八角枫中共检测出 60~63 个长重复序列, 正向重复所占的比例最高; 叶绿体基因组边界分析表明, 八角枫的 2 个亚种叶绿体基因组进化较保守, 而毛八角枫边界基因位置具有差异, 这可能也会影响叶绿体基因组总长度, 说明亲缘关系较近的物种此结构更加相似。

SSR 位点可被用于遗传多样性分析、物种鉴别、连锁作图及分子标记辅助选择育种等方面^[37,38]。鉴定出 186~179 个 SSR 位点, 单碱基重复居多数且重复 A/T 类型为主; 八角枫两亚种与毛八角枫相比, 除单碱基、二碱基和三碱基重复外, 在其余碱基重复的类型及数目上均有较大差异, 例如在稀花八角枫中检测到的四碱基重复类型为 AATT/AATT, 而在毛八角枫中未检测出, 且毛八角枫中并未有五碱基重复的 SSR 类型, 证实八角枫属叶绿体基因组间 SSR 多态性广泛, 基于叶绿体 SSR 位点开发的标记, 可用作种类的鉴别。密码子偏好性分析显示, 在所有氨基酸密码子中, 稀花八角枫、伏毛八角枫和毛八角枫叶绿体基因组对 A/U 结尾密码子的偏好性高于 G/C 结尾密码子。变异分析结果表明, 稀花八角枫、伏毛八角枫和毛八角枫叶绿体基因组中在 *ndhA*、*ycf1*、*rpl16*、*ycf2* 及 *petD* 等基因的编码区存在差异。系统进化分析结果显示, 稀花八角枫与伏毛八角枫亲缘关系密切。

虽然大多数高等被子植物叶绿体基因组 IR 区中的基因突变率常低于位于两个长短单拷贝区中的基因^[39], 但是叶绿体基因组中两个反向重复序列区与长、短单拷贝区的边界会存在扩张与收缩, 也会像基因和内含子的缺失一样影响到叶绿体基因组从而显示出高的突变性^[40]。本研究中稀花八角枫、伏毛八角枫和毛八角枫叶绿体基因组 IR 区较为恒定, 具有较高的保守性。

将稀花八角枫、伏毛八角枫和毛八角枫与八角枫属亲缘关系近的 28 个植物叶绿体全基因组序列进行比对后构建系统发育树, 分析结果表明稀花八角枫与伏毛八角枫亲缘关系最近, 暗示它们在进化上具有共线性。八角枫的叶绿体基因组信息将用于种群基因组研究和系统发育分析, 为更好地开发和利用该物种做出贡献。

本研究对稀花八角枫、伏毛八角枫和毛八角枫叶绿体基因组进行研究, 并分析了其叶绿体基因组结构特征, 挖掘其叶绿体基因组的重复序列位点并基于目前已测序和发表八角枫叶绿体基因组进行了边界分析、共线性分析以及变异分析等分析, 揭示了八角枫叶

叶绿体基因组的结构和一般变异特征, 并利用叶绿体基因组数据构建系统发育树, 揭示八角枫属物种之间的系统关系, 为药用植物八角枫的资源筛选、鉴定、保存及遗传多样性分析提供了分子依据。

作者贡献: 杨小英负责实验工作、数据分析、论文撰写与修改; 刘畅参与实验及数据分析; 曾宪法、刘雄伟参与稿件修改; 赵杰宏参与样品采集; 俸婷婷参与数据分析; 周英负责实验设计、论文指导和稿件修改。

利益冲突: 本文的研究内容无任何利益冲突。

References

- [1] Guizhou Provincial Drug Administration. Quality Standards for Chinese Herbal Medicine and Ethnic Medicine in Guizhou Province (贵州省中药材、民族药材质量标准) [S]. Guiyang: Guizhou Science and Technology Press, 2003: 11.
- [2] Xiong NQ, Zha J. Research progress of Hmong medicine *Alangium chinense* [J]. Sci Technol Asso Forum (科协论坛), 2013, (08): 108-109.
- [3] Wang BX. Modern Pharmacology of Chinese Medicine (现代中药药理学) [M]. Tianjin: Tianjin Science and Technology Press, 1997: 491.
- [4] Jiang YY, Wu WH, Peng RZ, et al. Preliminary study on the biopharmacology of the Chinese herb *Alangium chinense* [J]. Guangdong Pharm J (广东药学), 2004, 14: 8-10.
- [5] Ma CY, Yang HJ. Resarch progress on the traditional Bai medicine: piarpiaoryier (*Alangium chinense*) [J]. Chin J Ethnomed Ethnopharm (中国民族民间医药杂志), 2011, 20: 6-8.
- [6] Yue YD, Fu XB, Sun GD. Chemical Constituents of *Alangium chinense* [J]. Mod Chin Med (中国现代中药), 2022, 24: 805-811.
- [7] Editorial Committee of Flora of China, Chinese Academy of Sciences. Flora of China (中国植物志) [M]. Beijing: Science Press, 1983, 52: 166-168.
- [8] Meng YY, Du HZ, Wang XB, et al. Research progress on chemical constituents and pharmacological effects of *Alangium chinense* (Lour.) Harms [J]. Studies Trace Elements Health (微量元素与健康研究), 2021, 38: 40-43.
- [9] Zhang CY, Zhang LJ, Hu YL, et al. The pathological observation of acute intoxication by *Alangium chinense* in mice [J]. J Forensic Med (法医学杂志), 2009, 25: 329-331.
- [10] Dyal SD, Brown MT, Johnson PJ. Ancient invasions: from endosymbionts to organelles [J]. Science, 2004, 304: 253-257.
- [11] Chiba Y. Cytochemical studies on chloroplasts I: cytologic demonstration of nucleic acids in chloroplasts [J]. Cytologia, 1951, 16: 259-264.
- [12] Neuhaus HE, Emes MJ. Nonphotosynthetic metabolism in plastids [J]. Annu Rev Plant Physiol Plant Mol Biol, 2000, 51: 111-140.
- [13] Zhang M, Yin YP, Zhou LJ, et al. Comparative study on chloroplast genomes of the three Magnolia species [J]. Guihaia (广西植

- 物), 2022. <https://kns.cnki.net/kcms/detail/45.1134.q.20210622.1333.004.htm>.
- [14] Zhao Q, Yu JX, Qing YW, et al. Assembly and sequence analysis of chloroplast genome of *Championella sarcorrhiza* based on high-throughput sequencing [J]. *Chin Tradit Herb Drugs* (中草药), 2021, 52: 1744-1750.
- [15] Green BR. Chloroplast genomes of photosynthetic eukaryotes [J]. *Plant J*, 2011, 66: 34-44.
- [16] Kane NC, Cronk Q. Botany without borders: barcoding in focus [J]. *Mol Ecol*, 2010, 17: 5175-5176.
- [17] 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.
- [18] Zhang J, Lu JH, Wang QQ, et al. Characteristics of chloroplast genome of *Glycyrrhiza eurycarpa* P. C. Li from Xinjiang with comparison and phylogenetic analysis of the chloroplast genomes of the medicinal plants of *Glycyrrhiza* [J]. *Acta Pharm Sin* (药学报), 2022, 57: 1516-1525.
- [19] 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: 455-477.
- [20] Liu C, Shi L, Zhu Y, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences [J]. *BMC Genomics*, 2012, 13: 715.
- [21] Laslett D, Canbäck B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences [J]. *Bioinformatics*, 2008, 24: 172-175.
- [22] Stephan G, Pascal L, Ralph B. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes [J]. *Nucl Acids Res*, 2019, 47: W59-W64.
- [23] Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes [J]. *Curr Gen*, 2007, 52: 267-274.
- [24] Zhao R, Yin SY, Jiang CH, et al. Comparative analysis of chloroplast genomes of medicinal plants of *Aristolochiaceae* [J]. *China J Chin Mater Med* (中国中药杂志), 2022, 47: 2932-2937.
- [25] Thiel T, Michalek W, Varshney RK, et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum Vulgare* L.) [J]. *Theor Appl Gen*, 2003, 106: 411-422.
- [26] Li RJ, Wu LW, Xin TY, et al. Analysis of chloroplast genomes and development of specific DNA barcodes for identifying the original species of *Rhei Radix et Rhizoma* [J]. *Acta Pharm Sin* (药学报), 2022, 57: 1495-1505.
- [27] Frazer KA, Pachter L, Poliakov A, et al. VISTA: computational tools for comparative genomics [J]. *Nucleic Acids Res*, 2004, 32: W273-W279.
- [28] Li SC, Guo SH, Bei ZL. Chloroplast genome sequence and phylogenetic analysis of the *Berberis* [J]. *Chin Tradit Herb Drugs* (中草药), 2022, 53: 818-826.
- [29] Li J, Jiang B, Su YJ et al. Structure and phylogenetic analysis of the *Pseudotaxus chienii* (W. C. Cheng) W. C. Cheng chloroplast genome [J]. *Plant Sci J* (植物科学学报), 2021, 39: 5-13.
- [30] Wei D, Yan YH, Liu YP, et al. Comparative transcriptome and proteome profiling of chlorophyll metabolism pathway in four types of *Magnolia officinalis* [J]. *China J Chin Mater Med* (中国中药杂志), 2020, 45: 3826-3836.
- [31] Jiao Y, Jia HM, Li XW, et al. Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*) [J]. *BMC Genomics*, 2012, 13: 201.
- [32] Shen LQ. Complete Chloroplast Genomes of Three Medicinal Plants: Comparative and Evolutionary Analysis in *Lamiaceae* (唇形科三种药用植物叶绿体全基因组及科内的比较与进化分析) [D]. Hangzhou: Zhejiang University, 2018.
- [33] Wang R, Ou JM, Li X, et al. Construction of identification cards of medicinal plum varieties based on simple sequence repeats markers [J]. *J Anhui Univ Chin Med* (安徽中医药大学学报), 2020, 39: 62-67.
- [34] Zhang XL, Liu B, Fang FJ, et al. Chloroplast genome structure characteristic and phylogenetic analysis of *Arachis pintoi* [J]. *Mol Plant Breed* (分子植物育种), 2021. <https://kns.cnki.net/kcms/detail/46.1068.S.20210831.1738.016.htm>.
- [35] Huang H, Shi C, Liu Y, et al. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships [J]. *BMC Evol Biol*, 2014, 14: 151.
- [36] Liang FP, Wen XN, Gao HY, et al. Chloroplast genomes of *Asteraceae* species [J]. *Gen Appl Biol* (基因组学与应用生物学), 2018, 37: 5437.
- [37] Powell W, Morgante M, McDevitt R, et al. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines [J]. *Proc Natl Acad Sci U S A*, 1995, 92: 7759-7763.
- [38] Pugh T, Fouet O, Risterucci AM, et al. A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers [J]. *Theor Appl Genet*, 2004, 108: 1151-1161.
- [39] Maier RM, Neckermann K, Igloi GL, et al. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing [J]. *J Mol Biol*, 1995, 251: 614-628.
- [40] Dempewolf H, Kane NC, Ostevik KL, et al. Establishing genomic tools and resources for *Guizotia abyssinica* (L. f.) Cass. -the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome [J]. *Mol Ecol Resour*, 2010, 10: 1048-1058.