

分子植物育种
Molecular Plant Breeding
ISSN 1672-416X, CN 46-1068/S

《分子植物育种》网络首发论文

题目：木犀科叶绿体基因组比较及 DNA 条形码分析
作者：刘蕊，王贝，郭朋，刘燕培，王一涵，尚富德，林楠
网络首发日期：2023-05-05
引用格式：刘蕊，王贝，郭朋，刘燕培，王一涵，尚富德，林楠. 木犀科叶绿体基因组比较及 DNA 条形码分析[J/OL]. 分子植物育种.
<https://kns.cnki.net/kcms/detail/46.1068.s.20230504.2241.020.html>



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

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

研究报告
Research Report

木犀科叶绿体基因组比较及 DNA 条形码分析

刘蕊 王贝 郭朋 刘燕培 王一涵 尚富德* 林楠*

河南农业大学生命科学学院, 郑州, 450002

*共同通讯作者, fudeshang@henu.edu.cn; linnan93@163.com

摘要 木犀科 (Oleaceae) 隶属于双子叶植物纲唇形目 (Lamiales), 该科植物因具有良好的经济价值和较高的园林观赏价值而被广泛栽培, 但传统的分类学对其准确鉴定存在困难, 严重阻碍了木犀科植物资源的保护和利用。本研究从 NCBI 下载了 27 个木犀科叶绿体基因组序列, 涵盖了木犀科 5 个族 4 个亚族 19 个属 8 个组 27 个种, 对其进行叶绿体基因组比较及 DNA 条形码分析。结果表明 (1) 木犀科植物的叶绿体基因组均为典型的四分体结构, 大小在 154 kb~159 kb, 包含 80 个蛋白编码基因, 30 个 tRNA 及 4 个 rRNA; (2) 四分体结构比较发现, LSC 与 IRb 及 SSC 与 IRa 区域的边界基因分别为 *rpl2* 和 *ycf1*, 但在族间呈现不同程度的收缩和扩张。共线性分析发现木犀科 5 个族中仅在素馨族 (Jasmineae) 中发生~20 kb 的倒位和 *accD* 基因的丢失, 明显区别于其他族; (3) 相较于国际通用条形码, 我们筛选到的高变异片段和叶绿体基因组数据对木犀科物种的鉴定能力大幅度提升; (4) 基于不同矩阵构建木犀科系统发育树, 结果支持将木犀科划分为 5 个族 4 个亚族, 与前人结果基本一致。因此, 本研究证实了叶绿体基因组在木犀科植物的鉴定中具有重要作用, 为未来木犀科植物的系统发育及遗传多样性的相关研究奠定理论基础。

关键词 木犀科; 叶绿体基因组; 四分体结构比较; DNA 条形码; 系统发育

Comparison and DNA Barcoding Analysis of Chloroplast Genomes in Oleaceae

Liu Rui Wang Bei Guo Peng Liu Yanpei Wang Yihan Shang Fude* Lin Nan*

College of Life Sciences, Henan Agricultural University, Zhengzhou, 450002

*Co-corresponding authors, fudeshang@henu.edu.cn; linnan93@163.com

Abstract Oleaceae, one of typical family within Lamiales, is widely cultivated in the world due to the high ornamental and economic value. The definition and discrimination have been challenging based on traditional taxonomy, which highly impedes the protection and utilization of Oleaceae plants. In the current study, 27 chloroplast genomes were downloaded from NCBI, covering 5 tribes, 4 subtribes, 19 genera, 8 sections and 27 species of Oleaceae. We are aiming to explore the comparison and DNA barcoding of chloroplast genomes in Oleaceae, and our results showed that (1) the chloroplast genomes of Oleaceae exhibited a typical quadripartite structure, and the length of chloroplast genomes of Oleaceae ranged from 154 kb to 159 kb, including 80 protein-coding genes, 30 tRNA, and 4 rRNA; (2) comparison of quadripartite structure found the boundary genes of LSC/IRb and SSC/IRa were *rpl2* and *ycf1*, respectively, but degrees of gene expansion and contraction have been various among different tribes of Oleaceae. Meanwhile, synteny analysis revealed that Jasmineae was distinct from other tribes of Oleaceae by ~20 kb inversion and the loss of *accD* gene; (3) compared to international universal barcodes, variable hotspot regions and chloroplast genomes greatly improved the identification rate of Oleaceae species; (4) our phylogenetic trees based on different matrixes were highly consistent and supported Oleaceae divided into 5 tribes and 4 subtribes, the result was basically consistent with previous studies. Therefore, this study confirmed that chloroplast genomes play an important role in the identification of plants in Oleaceae, and our study will lay a theoretical foundation in promoting phylogeny and the research of genetic diversity in Oleaceae.

Keywords Oleaceae; Chloroplast genome; Comparison of quadripartite structure; DNA barcoding; Phylogeny

木犀科 (Oleaceae) 隶属于双子叶植物纲唇形目 (Lamiales), 木犀科植物主要分布在热带和温带地区,

全球有 27 个属, 约 700 个种 (Green, 2004)。中国分布 180 余种, 特有种高达 85% 以上。中国也是连翘属 (*Forsythia*)、丁香属 (*Syringa*)、女贞属 (*Ligustrum*) 和木犀属 (*Osmanthus*) 的现代分布中心。木犀科植物分布广泛, 具有重要的经济和生态价值, 囊括多种药用、香料、油料及观赏植物, 因此常作为经济作物和园林观赏植物, 在世界多地有引种栽培。由于环境气候变化和人类活动等原因导致的生境破碎化加剧了木犀科优良种质资源的流失, 甚至多种植物濒危灭绝; 另一方面, 基于传统的形态分类方法对木犀科植物的鉴别和种间界定具有较大的挑战性, 进一步限制了其有效的保护和资源的开发利用。因此, 对木犀科植物种质资源的准确鉴定, 尤其是对具有较高抗逆性和适应能力的树种开展分子系统学和 DNA 条形码的研究是至关重要的。

国内外对木犀科的分类学研究, 早期主要依据果实的类型将木犀科进一步划分为梣族 (*Fraxineae*, 翅果)、丁香族 (*Syringaceae*, 蒴果)、木犀榄族 (*Oleeae*, 核果)、素馨族 (*Jasmineae*, 浆果双生), 奠定了木犀科的基本分类系统 (中国植物志编辑委员会, 1992)。随后基于染色体基数对木犀科的分类系统进一步完善, 主要是将连翘属 (*Forsythia*, $x=14$) 和雪柳属 (*Fontanesia*, $x=13$) 分别从丁香族 (*Syringaceae*) 和梣族 (*Fraxineae*) 中独立出来, 重新建立了连翘族 (*Forsythieae*, $x=14$) 和雪柳族 (*Fontanesieae*, $x=13$) (Johnson, 1957)。木犀科植物在实际应用中主要以花、果实为主要鉴定依据, 因细胞学研究对实验材料的要求极高, 以及在实际的鉴定中因地理重叠和同源性状等诸多问题对木犀科植物的准确鉴定带来一定的困难。随着高通量测序技术的发展、测序成本的降低以及生物信息数据处理技术的提高, 植物 DNA 条形码 2.0 概念的引入 (李德铎和曾春霞, 2015), 通过浅层测序技术 (Genome skimming) 获得叶绿体基因组 (Chloroplast genome, cp genome) 和核基因组的核糖体 DNA (Nuclear ribosomal DNA, nrDNA) 序列, 具有更多可变特征以提高物种识别能力, 整合结构变异和系统发育信息将为植物的分类与精准鉴定提供重要途径 (Fu et al., 2022)。

木犀科分子系统学研究方面, Wallander 和 Albert (2000) 基于 *rps16* 和 *trnL-F* 基因片段对木犀科 25 个属 76 个种构建系统发育树, 结果提出摒弃亚科水平, 建立 5 个族, 但由于信息位点有限, 木犀榄族 (*Oleeae*) 中的 (*Oleinae*) 和梣属 (*Fraxinus*) 的系统位置仍无法被解析。Dupin 等 (2020) 整合线粒体、叶绿体和核基因的系统发育证据进一步验证了木犀科植物分为以下 5 个族 (*Fontanesieae*, *Forsythieae*, *Oleeae*, *Myxopyreae*, *Jasmineae*) 和 4 个亚族 (*Oleinae*, *Fraxininae*, *Ligustrinae*, *Schreberinae*), 但 *Oleeae* 中一些属的系统发育关系仍不明晰。Dong 等 (2022) 分别基于叶绿体基因组、核 SNP 数据集构建系统发育树也均支持木犀科植物分为 5 个族和 4 个亚族。目前, 由于取样及分子标记的问题, 木犀科的系统发育关系仍未能得到完全解决。因此, 我们将整合已发表的木犀科叶绿体基因组进行比较分析并重建其系统发育关系: (1) 比较木犀科叶绿体基因组四分体结构变异情况及进化规律; (2) 分析蛋白编码基因和基因间隔区的核苷酸多态性 (π), 筛选高变异片段作为木犀科植物潜在的 DNA 条形码; (3) 基于不同数据集 (叶绿体基因组, 79 个蛋白编码基因, 106 个基因间隔区, 3 个国际通用条形码, 3 个高变异片段) 分别重建木犀科系统发育关系, 评估不同数据集对木犀科物种的鉴定能力, 最终为理解木犀科植物的进化关系及未来种质资源遗传多样性的研究提供重要的参考依据。

1 结果与分析

1.1 木犀科叶绿体基因组结构比较分析

通过对 27 个木犀科植物叶绿体基因组大小、边界基因的差异及基因的数目等进行统计和比较, 结果发现 27 个木犀科植物叶绿体基因组均是典型的四分体结构, 包括大单拷贝区 (large single copy, LSC)、小单拷贝区 (small single copy, SSC) 及 2 个反向重复区 (inverted repeats, IRs)。27 个木犀科植物叶绿体基因组大小在 154 849 bp (Trib. *Oleeae*: *Nestegis apetala*)-159 404 bp (Trib. *Jasmineae*: *Jasminum fruticans*), GC 含量在 37.5% ~ 38%, LSC 的长度介于 81 851 bp (Trib. *Myxopyreae*: *Myxopyrum hainanense*) - 91 710 bp (Trib. *Jasmineae*: *Jasminum fruticans*), SSC 的长度介于 16 888 bp (Trib. *Oleeae*: *Priogymnanthus hasslerianus*) ~ 19 047 bp (Trib. *Oleeae*: *Syringa yunnanensis*), IR 区的长度在 24 898 bp (Trib. *Jasmineae*: *Jasminum fruticans*) ~ 26 646 bp (Trib. *Oleeae*: *Priogymnanthus hasslerianus*)。含有的基因数目为 113-114 个。其中有 15 个基因 (*rps16*, *atpF*, *rpoC1*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *ndhA*, *trnG-GCC*, *trnL-UAA*, *trnI-GAU*, *trnA-UGC*, *trnV-UAC*, *trnK-UUU*) 含有 1 个内含子, 有 3 个基因 (*rps12*, *ycf3*, *clpP*) 具有 2 个内含子。通过比较基因的缺失

情况，发现仅在素馨族 (Jasmineae) 的 *Jasminum fruticans* 中缺少 *accD* 基因 (表 1)。

本研究选择代表木犀科族水平上的 5 个种 (*Forsythia mandschurica*, *Fontanesia philliraeoides* subsp. *fortunei*, *Syringa yunnanensis*, *Jasminum fruticans*, *Myxopyrum hainanense*)，对其四分体结构进行比较分析。结果发现，LSC 与 IRb 及 SSC 与 IRa 区域的边界基因分别为 *rpl2* 和 *ycf1* (图 1)。在 IRb 和 SSC 区，*Fontanesia philliraeoides* subsp. *fortunei* (Fontanesieae) 的 *ycf1* 向 SSC 区偏移了 59 bp，*Syringa yunnanensis* (Oleeae) 的 *ycf1* 向 SSC 偏移了 1 001 bp，在 SSC 和 IRa 区，*Syringa yunnanensis* (Oleeae) 的 *ycf1* 向 IRa 偏移了 871 bp。而 *Jasminum fruticans* (Jasmineae) 的 *ycf1* 向 IRa 偏移了 1 358 bp (图 1)。

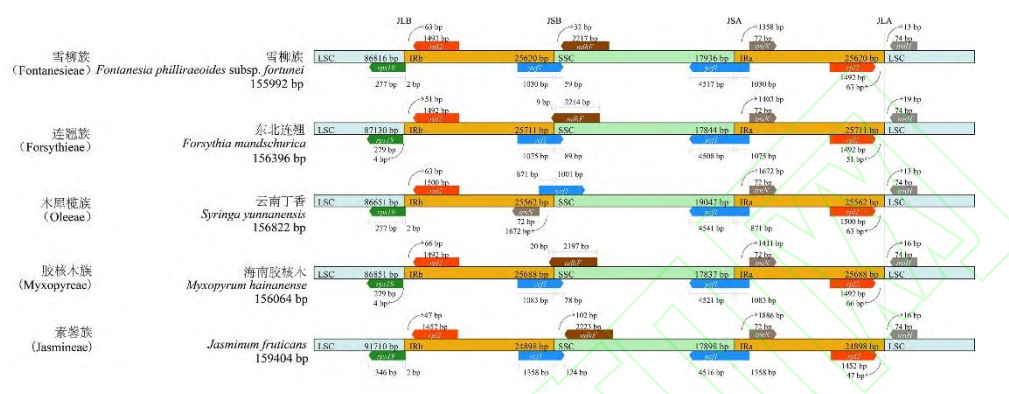


图 1 木犀科 5 个族叶绿体基因组四分体结构比较

注：未写中文名称表示该物种在国外分布

Figure 1 Comparison of quadripartite structure among five tribes chloroplast genomes of Oleaceae

Note: If the species name is not marked in Chinese, it means that the species is distributed abroad

表 1 27 个木犀科叶绿体基因组结构特征统计

Table 1 Statistics of structural features among 27 chloroplast genomes of Oleaceae

物种 Species	族 Tribe	登录号 Accession number	基因组大小 Genome size(bp)	GC 含量 GC content %	蛋白编码 基因 Protein coding genes	tRNA 数量 Number of tRNA	rRNA 数量 Number of rRNA	总基 因数 Total genes	LSC 长度 Length of LSC (bp)	SSC 长度 Length of SSC (bp)	IR 长度 Length of IR (bp)	边界基因 Boundary gene
翅果连翘 <i>Abeliophyllum distichum</i>	连翘族 Forsythieae	NC031445	155 982	37.8	80	30	4	114	86 742	17 834	25 703	<i>rpl2</i>
流苏树 <i>Chionanthus retusus</i>	木犀榄族 Oleeae	MH817897	155 637	37.8	80	30	4	114	86 483	17 858	25 648	<i>rpl2</i>
<i>Jasminum fruticans</i>	素馨族 Jasmineae	NC056244	159 404	37.5	79	30	4	113	91 710	17 898	24 898	<i>rpl2</i>
雪柳 <i>Fontanesia philliraeoides</i> subsp. <i>fortunei</i>	雪柳族 Fontanesieae	MG255754	155 992	37.7	80	30	4	114	86 816	17 936	25 620	<i>rpl2</i>
<i>Forestiera pubescens</i> var. <i>pubescens</i>	木犀榄族 Oleeae	MK299390	155 183	37.9	80	30	4	114	85 956	17 781	25 723	<i>rpl2</i>
东北连翘 <i>Forsythia mandschurica</i>	连翘族 Forsythieae	MH559275	156 396	37.8	80	30	4	114	87 130	17 844	25 711	<i>rpl2</i>
美国白栎 <i>Fraxinus americana</i>	木犀榄族 Oleeae	MH817907	155 602	37.9	80	30	4	114	86 433	17 765	25 702	<i>rpl2</i>
白蜡树 <i>Fraxinus chinensis</i>	木犀榄族 Oleeae	MW599993	155 610	37.8	80	30	4	114	86 486	17 760	25 682	<i>rpl2</i>
湖北栎 <i>Fraxinus hupehensis</i>	木犀榄族 Oleeae	NC052770	155 689	37.8	80	30	4	114	86 498	17 803	25 694	<i>rpl2</i>
水曲柳 <i>Fraxinus mandshurica</i>	木犀榄族 Oleeae	MK299393	155 560	37.8	80	30	4	114	86 332	17 820	25 705	<i>rpl2</i>
<i>Haenianthus salicifolius</i>	木犀榄族 Oleeae	MH817913	155 928	37.7	80	30	4	114	86 660	17,832	25 718	<i>rpl2</i>
海南胶核木 <i>Myxopyrum hainanense</i>	胶核木族 Myxopyreae	NC047485	156 064	37.7	80	30	4	114	81 851	17 837	25 688	<i>rpl2</i>

<i>Nestegis apetala</i>	木犀榄族 Oleeae	MG255758	154 849	37.8	80	30	4	114	85 898	17 505	25 723	<i>rpl2</i>
<i>Noronhia lowryi</i>	木犀榄族 Oleeae	MH817920	155 602	37.8	80	30	4	114	86 384	17 798	25 710	<i>rpl2</i>
<i>Notelaea microcarpa</i>	木犀榄族 Oleeae	MH817923	155 019	37.8	80	30	4	114	86 119	17 474	25 713	<i>rpl2</i>
木犀榄 <i>Olea europaea</i> subsp. <i>europaea</i>	木犀榄族 Oleeae	MG255762	155 872	37.8	80	30	4	114	86 601	17 787	25 742	<i>rpl2</i>
<i>Olea lancea</i>	木犀榄族 Oleeae	MH817878	155 863	37.8	80	30	4	114	86 633	17 798	25 716	<i>rpl2</i>
腺叶木犀榄 <i>Olea paniculata</i>	木犀榄族 Oleeae	MT560031	155 625	37.8	80	30	4	114	86 444	17 785	25 698	<i>rpl2</i>
云南木犀榄 <i>Olea tsoongii</i>	木犀榄族 Oleeae	MH817930	155 237	37.8	80	30	4	114	86 135	17 818	25 642	<i>rpl2</i>
<i>Osmanthus americanus</i>	木犀榄族 Oleeae	MH119143	155 553	37.8	80	30	4	114	86 277	17 840	25 718	<i>rpl2</i>
野桂花 <i>Osmanthus yunnanensis</i>	木犀榄族 Oleeae	NC042463	155 482	37.8	80	30	4	114	86 584	17 490	25 704	<i>rpl2</i>
狭叶总序桂 <i>Phillyrea angustifolia</i>	木犀榄族 Oleeae	MH817935	155 335	37.8	80	30	4	114	86 516	17 501	25 609	<i>rpl2</i>
<i>Picconia excelsa</i>	木犀榄族 Oleeae	MH817938	155 355	37.8	80	30	4	114	86 449	17 496	25 705	<i>rpl2</i>
<i>Priogymnanthus hasslerianus</i>	木犀榄族 Oleeae	MH817879	156 760	37.6	80	30	4	114	86 580	16 888	26 646	<i>rpl2</i>
<i>Schrebera arborea</i>	木犀榄族 Oleeae	MG255767	155 657	37.8	80	30	4	114	86 622	17 553	25 741	<i>rpl2</i>
紫丁香 <i>Syringa oblata</i>	木犀榄族 Oleeae	NC057990	155 569	37.9	80	30	4	114	86 200	17 903	25 733	<i>rpl2</i>
云南丁香 <i>Syringa yunnanensis</i>	木犀榄族 Oleeae	NC042468	156 822	38	80	30	4	114	86 651	19 047	25 562	<i>rpl2</i>

注：未写中文名称表示该物种在国外分布

Note: If the species name is not marked in Chinese, it means that the species is distributed abroad

以东北连翘 (*Forsythia mandschurica*) 的叶绿体基因组作为参考序列, 选择以上 5 个木犀科植物叶绿体基因组序列进行序列变异分析 (图 2), 结果发现, 木犀科叶绿体基因组的基因间隔区的序列差异要高于编码区, 其中 IR 区相较 LSC 区域要更保守 (图 2)。同时对木犀科植物叶绿体基因组进行共线性分析 (图 3), 结果发现仅在 *Jasminum fruticans* (Jasmineae) 的叶绿体基因组的 LSC 区域发生了~20 kb 的倒位现象, 发生倒位的基因在 *ycf4-ycf3* 之间, 而其它木犀科 4 个族物种的叶绿体基因组结构较为保守, 未发现有基因的重排与倒位现象 (图 3)。

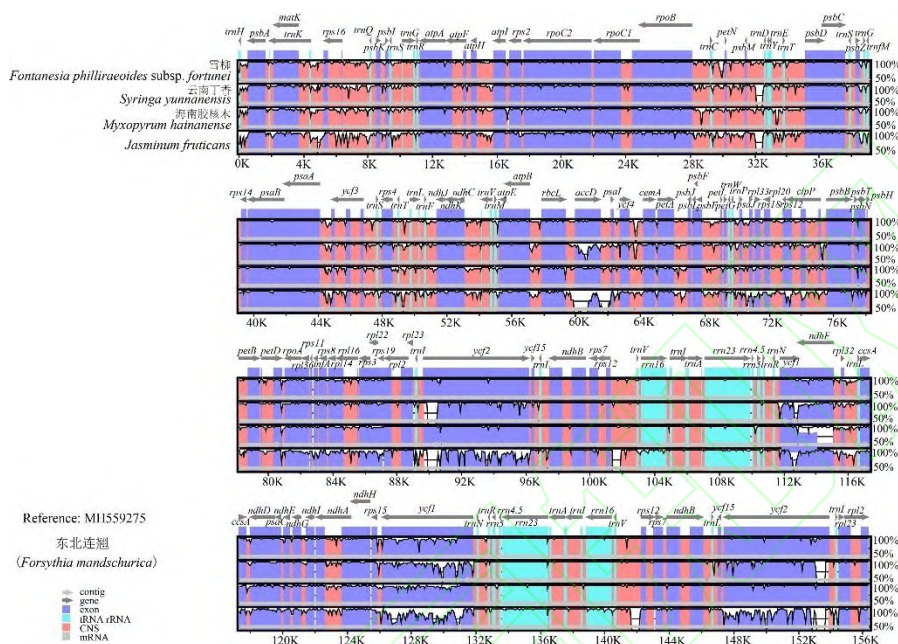


图 2 木犀科 5 个族叶绿体基因组变异比较 (以 *Forsythia mandschurica* 作为参考)

注: 上方灰色箭头表示基因和转录方向; 纵坐标表示范围从 50%-100% 的序列一致性百分比; CNS: 保守的非编码区; 未写中文名称表示该物种在国外分布

Figure 2 Structural alignment among five tribes chloroplast genomes of Oleaceae with *Forsythia mandschurica* as reference
Note: Grey arrow above the alignment indicate genes and transcriptional directions; the vertical scale indicates the percentage of identity, ranging from 50% to 100%; CNS: conserved non-coding sequences; if the species name is not marked in Chinese, it means that the species is distributed abroad

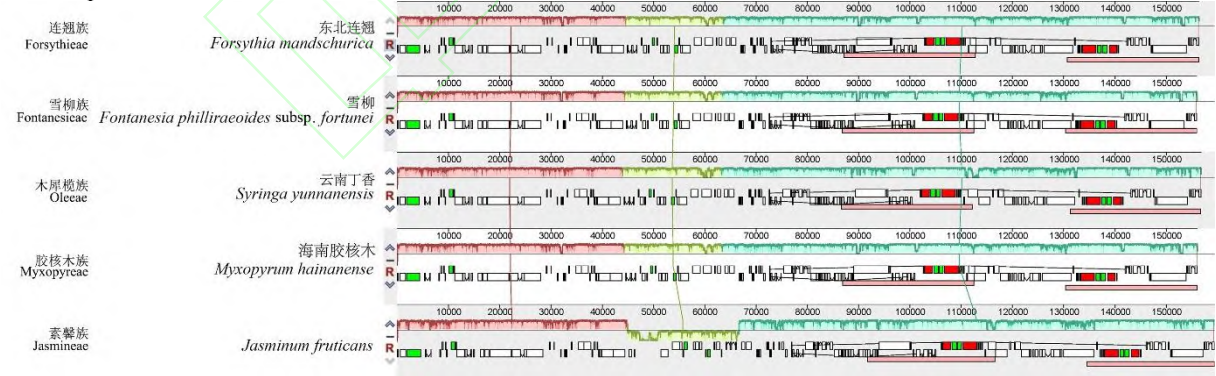


图 3 木犀科 5 个族叶绿体基因组共线性分析

注: 相同颜色代表同源区域, 白色矩形代表 CDS, 绿色代表 tRNA, 红色代表 rRNA; 未写中文名称表示该物种在国外分布

Figure 3 Synteny analysis among five tribes chloroplast genomes of Oleaceae
Note: Blocks of the same color represent homologous regions, the white rectangle is CDS, the green rectangle is tRNA, the red rectangle is rRNA; if the species name is not marked in Chinese, it means that the species is distributed abroad

1.2 通用条形码的评估及高变异片段筛选

通过对 27 个木犀科植物的蛋白编码基因及基因间隔区进行核苷酸多态性 (pi) 分析, 发现蛋白编码基因的 pi 值在 0.000 79~0.045 23 (图 4A), 在基因间隔区域内, 其 pi 值在 0~0.079 1 (图 4B)。综合比较木犀科叶绿体基因组中的变异位点占比、信息位点占比及比对的序列长度, 其中在蛋白编码区, 筛选出变异位点占比位于前三的基因分别是 *clpP* (0.261)、*rpl22* (0.164)、*rps18* (0.151) (表 2), 在基因间隔区, 筛选出变异位点占比位于前三的基因分别是 *trnH-psbA* (0.311)、*ndhF-rpl32* (0.247)、*rpl32-trnL*

(0.224) (表 2)。本研究进一步比较了国际通用条形码 (*matK*, *rbcL*, *trnH-psbA*) 和筛选出来的 3 个高变异片段 (*trnH-psbA*, *ndhF-rpl32*, *clpP*) 的种间遗传距离, 结果发现 *matK* 基因和 *rbcL* 基因的种间遗传距离均在 0~0.015 之间, 而 *trnH-psbA* 基因在 0~0.035 之间 (图 5A)。此外, 筛选出来的 2 个高变异片段 *ndhF-rpl32* 和 *clpP* 的种间遗传距离分别在 0.005~0.035 之间和 0~0.05 之间 (图 5B)。

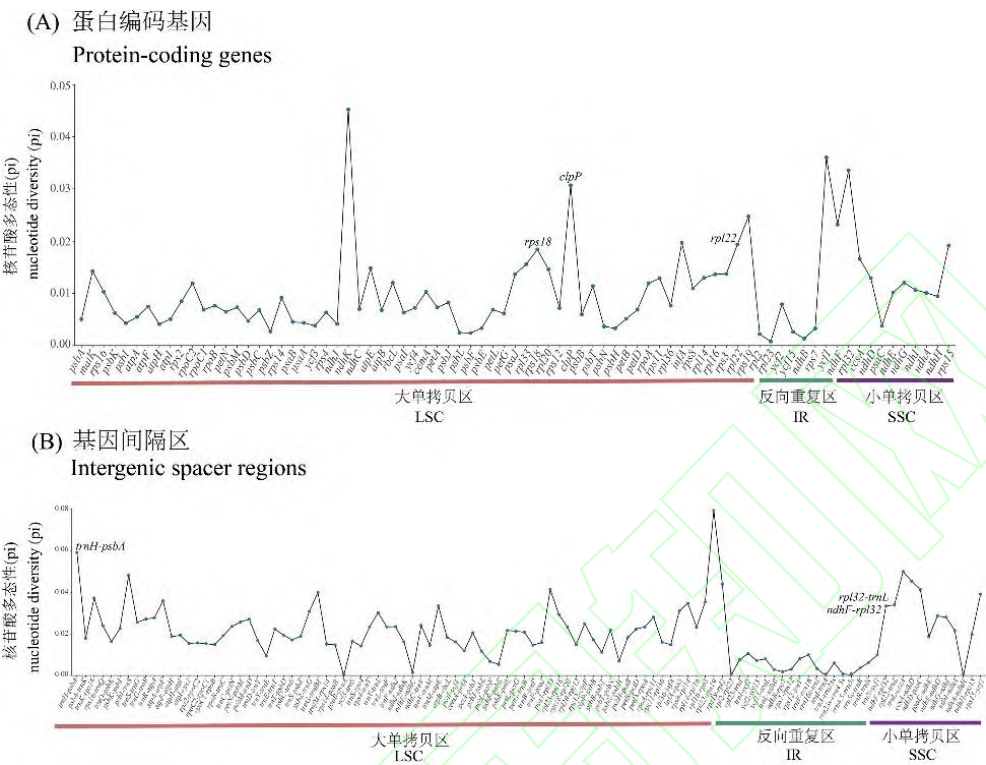


图 4 27 个木犀科物种蛋白编码基因和基因间隔区核苷酸多态性
Figure 4 The nucleotide diversity of protein-coding genes and intergenic spacer regions among 27 species of Oleaceae
表 2 木犀科蛋白编码基因与基因间隔区的高变异片段
Table 2 Highly variable fragments of protein-coding genes and intergenic spacer regions in Oleaceae

基因 Genes	比对长度 Aligned length (bp)	变异位点 No. of variable sites	信息位点 No. of parsimony informative sites	变异位点占比 (%) % of variable sites	信息位点占比 (%) % of parsimony informative sites
<i>clpP</i>	675	176	28	0.261	0.041
<i>rpl22</i>	489	80	19	0.164	0.039
<i>rps18</i>	464	70	9	0.151	0.019
<i>trnH-psbA</i>	586	182	76	0.311	0.13
<i>ndhF-rpl32</i>	547	135	47	0.247	0.086
<i>rpl32-trnL</i>	940	211	79	0.224	0.084

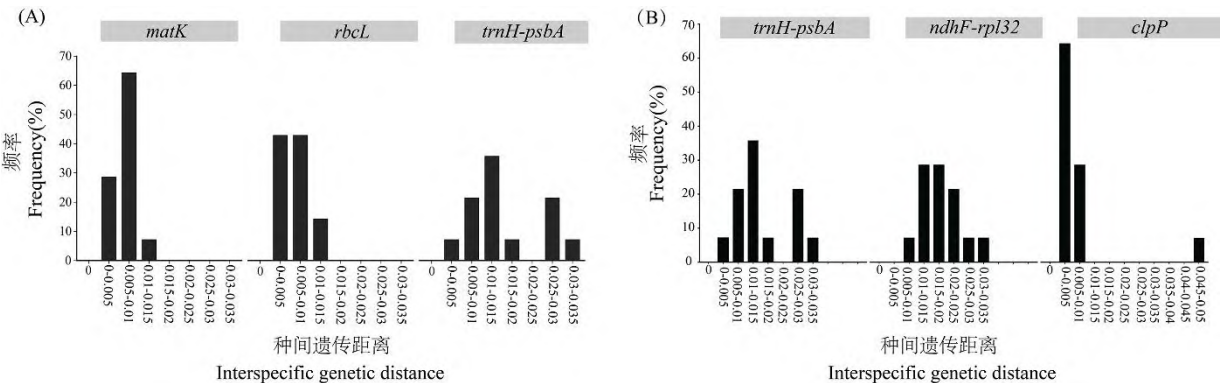


图 5 木犀科国际通用条形码和高变异片段种间遗传距离分布
Figure 5 Distribution of interspecific genetic distances for international universal barcodes and highly variable fragments of Oleaceae

1.3 基于叶绿体基因组木犀科系统发育关系

本研究分别基于 27 个木犀科植物完整的叶绿体基因组 (179 165 bp)、79 个共有蛋白编码基因 (66 672 bp)、106 个基因间隔区 (36 736 bp)、3 个国际通用条形码 (3 355 bp)、3 个高变异片段 (1 449 bp) 共 5 个数据集 (表 3)，分别基于最大似然法 (Maximum likelihood, ML) 和贝叶斯法 (Bayesian inference, BI) 构建系统发育树，结果发现，基于国际通用条形码和高变异片段构建的系统发育树对木犀科植物的鉴定能力分别为 48.15% (13/27) (图 6A) 和 59.26% (16/27) (图 6B)，但高变异片段构建的系统发育树在木犀榄族 (*Oleeae*) 内 (*Olea europaea* subsp. *europaea*、*Olea lancea*、*Olea paniculata*、*Priogymnanthus hasslerianus*、*Haenianthus salicifolius*) 仍未能得到较好的解析，支持率均低于 50%。基于蛋白编码基因及基因间隔区构建的系统发育树的鉴定能力均为 88.89% (24/27) (图 7A; 图 7B)，而基于叶绿体基因组对木犀科植物的鉴定成功率高达 100% (27/27) (图 8)。

表 3 木犀科 5 种数据集的建树参数
Table 3 The parameters for constructing phylogenetic trees in five data sets of Oleaceae

数据集 Data sets	长度 Length (bp)	方法 Method	分区 Partition	最优模型分 值 Best model score	方法 Method	分区 Partition	最优模型分 值 Best model score
27 个叶绿体基因组 27 chloroplast genomes	179 165	最大似然法 Maximum likelihood	未分区 No partition	-456950.857	贝叶斯法 Bayesian inference	未分区 No Partition	-458679.761
79 个蛋白编码基因 79 protein coding genes	66 672	最大似然法 Maximum likelihood	11	-158831.733	贝叶斯法 Bayesian inference	11	-159063.035
106 个基因间隔区 106 intergenic spacer Regions	36 736	最大似然法 Maximum likelihood	7	-117418.503	贝叶斯法 Bayesian inference	9	-117880.886
3 个国际通用条形码 Three international universal barcodes	3 355	最大似然法 Maximum likelihood	3	-9827.932	贝叶斯法 Bayesian inference	3	-9883.736
3 个高变异片段 Three highly variable fragments	1 449	最大似然法 Maximum likelihood	2	-5939.762	贝叶斯法 Bayesian inference	2	-5997.468

(A) 基于国际通用条形码 (*matK+rbcl+trnH-psbA*) 构建木犀科系统发育树
Phylogenetic tree of Oleaceae based on international universal barcodes



(B) 基于高变异片段 (*trnH-psbA+ndhF-rpl32+clpP*) 构建木犀科系统发育树
Phylogenetic tree of Oleaceae based on highly variable fragments



图 6 基于最大似然法 (ML) 和贝叶斯法 (BI) 构建木犀科国际通用条形码和高变异片段系统发育树
注: ★代表 bootstrap<60%, ★代表 bootstrap: 60%-100%; 未写中文名称表示该物种在国外分布

Figure 6 Phylogenetic trees of Oleaceae species using Maximum likelihood (ML) and Bayesian inference (BI) methods based on the international universal barcodes and highly variable fragments

Note: The blue asterisk represents bootstrap<60%, the red asterisk represents bootstrap: 60%-100%; if the species name is not marked in Chinese, it means that the species is distributed abroad

(A) 基于79个蛋白编码基因构建木犀科系统发育树
Phylogenetic tree of Oleaceae based on 79 protein-coding genes



(B) 基于106个基因间隔区构建木犀科系统发育树
Phylogenetic tree of Oleaceae based on 106 intergenic spacer regions

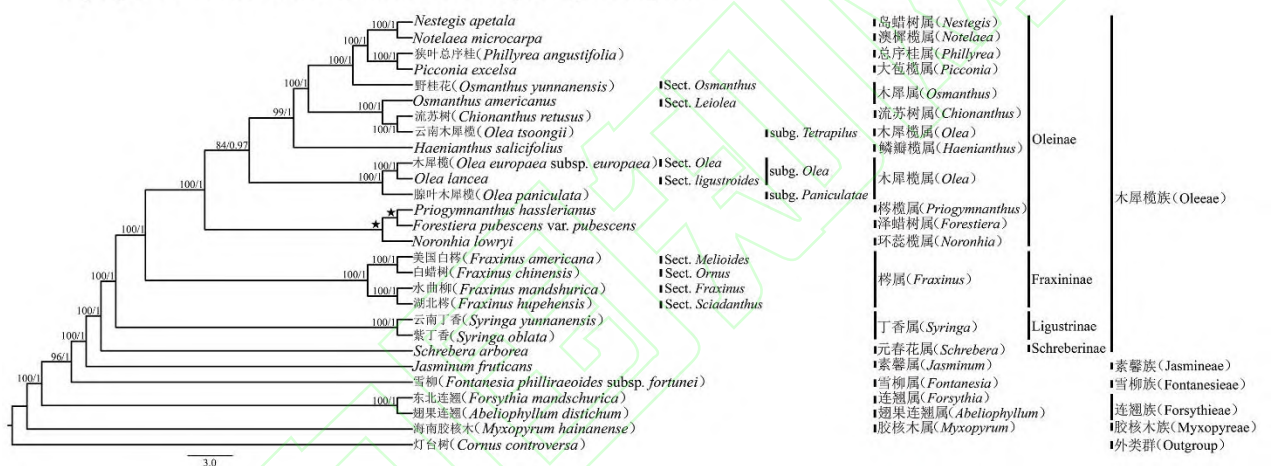


图7 基于最大似然法 (ML) 和贝叶斯法 (BI) 构建木犀科 79 个蛋白编码基因和 106 个基因间隔区系统发育树

注: ★表示最大似然法 (ML) 与贝叶斯法 (BI) 构建的系统发育树之间存在不一致; 未写中文名称表示该物种在国外分布
Figure 7 Phylogenetic tree of Oleaceae species using Maximum likelihood (ML) and Bayesian inference (BI) based on 79 protein-coding genes and 106 intergenic spacer regions

Note: The black asterisk represents incongruence for phylogenetic tree of Maximum likelihood (ML) and Bayesian inference (BI); if the species name is not marked in Chinese, it means that the species is distributed abroad

以叶绿体基因组构建的系统发育树为例, 结果支持将 27 个木犀科植物划分为 5 个族 4 个亚族。其中支持胶核木族 (Myxopyreae) 位于基部分支, 并与 Forsythiae 的亲缘关系最近 (图 8)。同时, 我们的结果也支持 Oleaceae 和 Jasmineae 的亲缘关系较近 (BS=100%, PP=1)。在木犀榄族 (Oleaceae) 中, Schreberinae 为最早分化出来的亚族并与 Ligustrinae 的亲缘关系较近 (BS=100%, PP=1)。其中, 在 Oleinae 中包含有 11 个属, Noronhia 为最早分化出来的属。同时在 Oleinae 中, Osmanthus 中的 Osmanthus yunnanensis 和 Osmanthus americanus 应聚为一支, 但 Osmanthus americanus 与 Chionanthus retusus+Olea tsoongii 聚为一支; 同为 Olea 的 (Olea europaea subsp. europaea+Olea lancea+Olea paniculata) 单独聚为一支, 而 Olea tsoongii 与 Chionanthus retusus 聚为一支。在 Fraxininae 中仅有一个栲属 (Fraxinus), 该属下的 4 个物种聚为一支, 构成单系群 (BS=100%, PP=1)。此外, 我们的系统发育树结果强烈支持 Forsythia 和 Abeliophyllum 聚为一支 (BS=100%, PP=1) 且与 Fontanesia 的亲缘关系较近 (图 8)。



图 8 基于最大似然法 (ML) 和贝叶斯法 (BI) 构建木犀科叶绿体基因组系统发育树

注：未写中文名称表示该物种在国外分布

Figure 8 Phylogenetic tree of Oleaceae species using Maximum likelihood (ML) and Bayesian inference (BI) methods based on the chloroplast genomes

Note: If the species name is not marked in Chinese, it means that the species is distributed abroad

1.4 重复序列分析

对木犀科叶绿体基因组中 SSR 的分布类型及数量进行统计，结果发现在雪柳族 (Fontanesieae) 的 *Fontanesia philliraeoides* subsp. *fortunei*、连翘族 (Forsythieae) 的 *Forsythia mandschurica*、木犀榄族 (Oleaceae) 的 *Syringa yunnanensis*、胶核木族 (Myxopyreae) 的 *Myxopyrum hainanense*、素馨族 (Jasmineae) 的 *Jasminum fruticans* 中包含有 6 种类型 SSR (单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸) (图 9A)。在 *Fontanesia philliraeoides* subsp. *fortunei* 中，总核苷酸数量为 52 个，其中单核苷酸数量最多 (46, 88.46%)，二核苷酸数量最少 (2, 3.85%)，未发现有三核苷酸、五核苷酸和六核苷酸类型。在 *Forsythia mandschurica* 中，总核苷酸数量为 63 个，其中单核苷酸数量最多 (49, 77.78%)，三核苷酸和五核苷酸数量均较少 (1, 1.59%)，未发现六核苷酸类型。在 *Syringa yunnanensis* 中，总核苷酸数量为 71 个，其中单核苷酸数量最多 (53, 74.65%)，六核苷酸数量最少 (2, 2.82%)。在 *Myxopyrum hainanense* 中，总核苷酸数量为 79 个，其中单核苷酸数量最多 (64, 81.01%)，三核苷酸数量最少 (1, 1.27%)，未发现有六核苷酸类型。在 *Jasminum fruticans* 中，总核苷酸数量为 79 个，其中单核苷酸数量最多 (46, 58.23%)，五核苷酸最少 (1, 1.27%)。在 *Jasminum fruticans* 和 *Myxopyrum hainanense* 中的总核苷酸数量最多，均为 79 个。在 *Fontanesia philliraeoides* subsp. *fortunei* 中含有的总核苷酸数量最少，共 52 个，仅在 *Syringa yunnanensis* 和 *Jasminum fruticans* 中发现含有六核苷酸类型 (图 9A)。SSR 在 LSC、SSC 及 IR 区均有分布，但在 *Forsythia mandschurica* 中，未发现在 IR 区有 SSR 分布 (图 9B)。

选择以上木犀科 5 个族的叶绿体基因组，对其重复序列进行检测，共检测到 4 种重复序列 (图 9C)。在 *Syringa yunnanensis* 中，有 4 种重复序列，其中正向重复序列最多 (59, 52.68%)，互补重复序列最少 (1, 1.69%)。在 *Myxopyrum hainanense* 中，只有正向重复序列和回文重复序列，其中正向重复序列 (15, 42.86%)，回文重复序列 (20, 57.14%)。在 *Jasminum fruticans* 中，有 3 种重复序列，其中正向重复序列最多 (73, 62.39%)，其次是回文重复序列 (42, 35.9%)，反向重复序列最少 (2, 1.71%)。在 *Fontanesia philliraeoides* subsp. *fortunei* 中，有 4 种重复序列类型，其中正向重复序列最多 (25, 2.04%)，其次是回文重复序列 (24, 47.06%)，反向重复序列和互补重复序列均为 1 (1, 1.96%)。在 *Forsythia mandschurica* 中，只有 2 种重复序列，其中回文重复序列最多 (24, 52.17%)，其次是正向重复序列 (22, 47.83%)，未发现反向重复序列和互补重复序列 (图 9C)。5 个木犀科叶绿体基因组重复序列的长度大多集中在 30~34 bp，少量重复序列长度在 50~54 bp，重复序列长度 >60 bp 区域，*Jasminum fruticans* 的重复序列数量最多 (图 9D)。

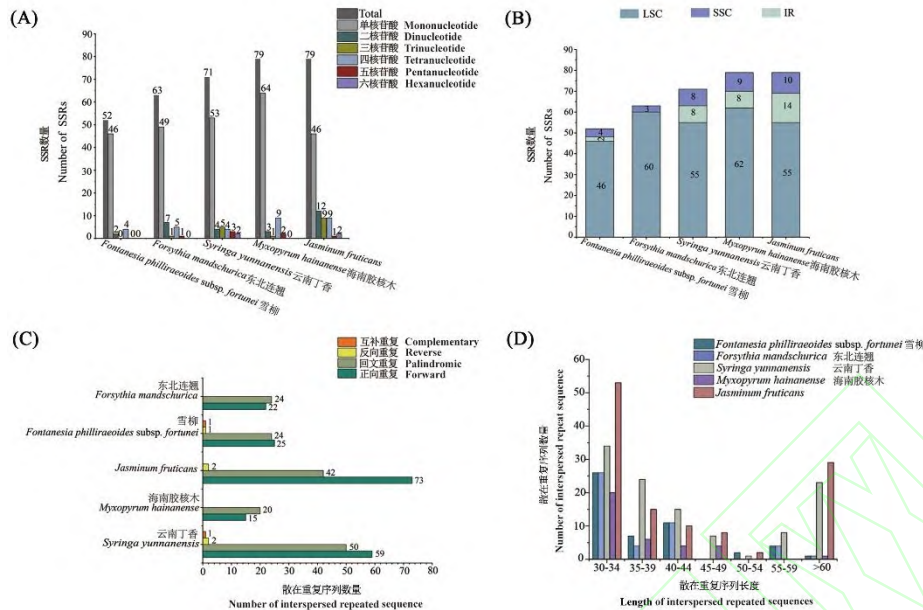


图9 木犀科5个族叶绿体基因组重复序列分析

注：(A) 简单重复序列类型；(B) 简单重复序列分布；(C) 散在重复序列数量；(D) 散在重复序列长度；未写中文名称表示该物种在国外分布

Figure 9 Analysis of repeat sequences of five tribes chloroplast genomes in Oleaceae

Note: (A) Type of simple sequence repeats; (B) Distribution of simple sequence repeats; (C) Number of interspersed repeated sequences; (D) Length of interspersed repeated sequences; if the species name is not marked in Chinese, it means that the species is distributed abroad

2 讨论

2.1 叶绿体基因组结构比较

本研究整合已发表的27个木犀科叶绿体基因组，代表了现有木犀科5个族4个亚族19个属。对其叶绿体基因组进行结构比较分析及系统发育关系重建，结果发现木犀科植物的叶绿体基因组结构较保守，为典型的四分体结构，基因组大小在154 kb-159 kb，表明木犀科叶绿体基因组与前人研究的被子植物的长度范围一致 (Zhang et al., 2012)。然而，木犀科叶绿体基因组结构在族间和属间存在显著的结构变异 (图1)。首先在基因数目上，素馨族 (Jasmineae) 中存在 *accD* 基因的缺失，明显区别于其他的族。在被子植物豆科 (Fabaceae)、禾本科 (Poaceae)、桔梗科 (Campanulaceae) 及菖蒲科 (Acoraceae) 中均发现 *accD* 基因丢失现象 (Jansen et al., 2007) 且研究表明 *accD* 基因在维持烟草叶片发育中发挥重要的作用 (Kode et al., 2005)。因此，推测素馨族 (Jasmineae) 中 *accD* 基因的缺失可能是该族物种进化的一种表现。此外，也有研究发现 *accD* 基因的丢失也可能与重排事件有关 (Lee et al., 2007)，我们进一步进行共线性分析发现在素馨族 (Jasmineae) 中，LSC 区域的 *ycf4-ycf3* 之间发生了~20 kb 的倒位，验证了这一可能。

我们进一步比较其四分体结构的变异，结果发现木犀科植物在 LSC 与 IRb 以及 SSC 与 IRa 区域的边界基因均分别为 *rpl2* 和 *ycf1*。但在不同族间连接处的位置具有明显的差别，比如在 IRb 和 SSC 区，相较于木犀科其他族的 *ycf1* 向 SSC 偏移了 (78 bp-1 001 bp)，Trib. Fontanesieae (*Fontanesia philliraeoides* subsp. *fortunei*) 的 *ycf1* 仅向 SSC 偏移了 59 bp (图1)，研究表明 IR 与 SC 边界区的收缩与扩张会导致植物叶绿体基因组长度的变化 (Kim and Lee, 2004)。比如李连星等 (2022) 对长爪柃[木衣]叶绿体基因组结构变异进行分析，结果发现 IR/SC 的边界区域的扩张收缩可能导致了长爪柃[木衣]叶绿体基因组的长度相较于其他近缘种要短。因此，在本研究木犀科5个族中，Trib. Fontanesieae (*Fontanesia philliraeoides* subsp. *fortunei*) 的叶绿体基因组长度 (155 992 bp) 要短于其他4个族的物种 (156 064 bp-159 404 bp) 可能是由于 IR 区和 SC 区域边界基因的收缩与扩张导致的。

2.2 DNA 条形码研究

DNA 条形码技术是一种新型的物种鉴定方法，主要是利用生物体内一段相对较短、易扩增且具足够变异的一段 DNA 序列来对物种进行准确鉴定 (Hebert et al., 2003)。2009 年，CBOL 植物条形码研究小组提出将叶绿体基因组中的 *matK*、*rbcL* 基因以及基因间隔区的 *trnH-psbA* 片段作为植物标准的 DNA 条形码 (CBOL Plant Working Group, 2009)。前人已基于以上片段做了一些研究，比如，Kim 等 (2014) 基于叶绿体

片段对韩国兰科 (Orchidaceae) 植物进行鉴定, 结果发现 *atpF-atpH+psbK-psbI+trnH-psbA* 组合片段可以对该科植物进行准确鉴定 (鉴定率为 98.8%)。此外, 徐素素等 (2021) 发现以 *ITS* 和 *ITS2* 为主, *matK*、*psbA-trnH* 为辅可实现对药用石斛及其混伪品准确快速的鉴定。本研究基于 5 个数据集构建系统发育树对木犀科植物的鉴定能力进行评估, 结果发现筛选出来的高变异片段对物种的鉴定能力为 59.26% (16/27), 明显高于国际通用条形码 48.15% (13/27)。此外, 由于叶绿体基因组能够提供更多的变异位点和信息位点, 可以对物种进行准确的鉴定。前人也基于叶绿体基因组对物种进行了 DNA 条形码的研究。比如 Fu 等 (2022) 基于 8 个数据集对杜鹃属 (*Rhododendron*) 植物进行 DNA 条形码的研究, 结果发现叶绿体基因组对杜鹃属植物 52% 的鉴定率要远高于 *ITS+matK+rbcL+trnH-psbA* (33%)、*ITS+matK+trnH-psbA* (29%)、*matK+rbcL+trnH-psbA* (24%)、nrDNA (19%) 的鉴定率; Song 等 (2020) 基于叶绿体基因组及其片段对沙拐枣属 (*Calligonum*) 植物进行 DNA 条形码的研究, 结果发现相较于国际通用条形码 *matK+rbcL+trnH-psbA* (33%) 的鉴定率, 叶绿体基因组的鉴定率可以达到 (78%); 本研究对木犀科进行 DNA 条形码的研究, 结果发现相较于高变异片段 59.26% (16/27) 的鉴定能力, 基于叶绿体基因组构建的系统发育树对物种的鉴定能力高达 100% (27/27)。综合以上研究结果均表明了叶绿体基因组可用于对木犀科植物 DNA 条形码的研究, 有利于促进木犀科种质资源的准确鉴定和优良品种的改良。

2.3 系统发育分析

基于 5 个不同的矩阵对木犀科系统发育关系进行重建, 进一步证实了 Wallander 和 Albert (2000) 对木犀科 5 个族 4 个亚族的分类体系。首先, 我们支持胶核木族 (Myxopyreae) 位于系统发育树的基部, 为木犀科中较早分化出来的较原始类群 (图 8), 该结果与 Dong 等 (2022) 基于叶绿体基因组及核 SNP 构建的系统发育树的结果一致。同时在形态上, 胶核木族 (Myxopyreae) 的胚珠不同于木犀科其他属为下垂的胚珠, 胶核木族 (Myxopyreae) 为向上的胚珠 (Wallander and Albert, 2000), 可见, 形态和分子的证据均支持将其单独划分出来。同时我们发现支持 Jasmineae 和 Oleaceae 的亲缘关系最近, 通过比较二者的化学成分发现木犀科中仅在 Jasmineae 和 Oleaceae 中发现含有木犀榄苷 (oleoside), 进一步支持这 2 个族的亲缘关系最近 (Jensen, 1992), 并与近期的系统发育结果一致 (Dong et al., 2022)。基于形态数据和分子证据发现在 Forsythieae 和 Fontanesieae 2 个族中, 相较于 *Forsythia* 开裂的蒴果, *Abeliophyllum* 和 *Fontanesia* 的果实均为翅果, 因此认为 *Abeliophyllum* 和 *Fontanesia* 的亲缘关系较近 (Taylor, 1945), 但基于染色体证据和化学成分的研究表明 *Abeliophyllum* 和 *Forsythia* 的染色体基数均为 14 并含有棘木甙 (cornoside), 而 *Fontanesia* 的染色体基数为 13 且不含有棘木甙 (cornoside), 因此认为 *Abeliophyllum* 和 *Forsythia* 的亲缘关系最近 (Taylor, 1945; Jensen, 1992), 同时 Dupin 等 (2020) 基于 80 个叶绿体基因、37 个线粒体基因、核基因分别构建木犀科系统发育树, 结果发现均支持 *Abeliophyllum* 和 *Forsythia* 聚为一支且归属于连翘族 (Forsythieae), 本研究的结果与 Dupin 等 (2020) 研究结果一致。在 Oleaceae 内, Oleinae、Fraxininae、Ligustrinae 及 Schreberinae 均具有相同的染色体基数 ($x=23$)、韧型纤维及化学物质 (均含有黄酮苷 flavone glycosides) (Taylor, 1945; Wallander and Albert, 2000)。其中, 在 Oleaceae 中, Schreberinae 是其最早分化出来的亚族, 其中 Oleinae 与 Fraxininae 的亲缘关系较近且在形态上二者的木材均有边缘薄壁组织。Fraxininae 中的 4 个桤属 (*Fraxinus*) 物种聚为一支, 前人的研究结果发现其染色体基数均为 23, 支持其构成单系类群, 本研究构建的系统发育树也支持该结论 (Taylor, 1945) (图 8)。

综上, 本研究对木犀科植物叶绿体基因组结构变异及进化规律进行研究, 并重建了木犀科叶绿体基因组系统发育树, 进一步探讨了其物种进化关系, 为木犀科植物种质资源的鉴定、利用与保护提供理论依据。

3 材料与方法

3.1 数据来源

本研究中的 27 个木犀科叶绿体基因组序列, 均下载自 NCBI (<https://www.ncbi.nlm.nih.gov/>) 数据库中, 其代表了木犀科 5 个族 (5/5), 4 个亚族 (4/4), 19 个属 (19/27), 27 种 (27/700) 木犀科植物。其物种名称、序列登录号及基因组的详细信息见表 1。

3.2 木犀科叶绿体基因组比较分析

利用 Geneious v. 9.0 软件对已经发表的木犀科 27 个叶绿体基因组的序列进行校正, 同时对边界基因的长度、基因数量及 GC 含量等信息进行统计 (Kearse, 2012)。由于木犀科植物叶绿体基因组在族水平的保守性, 因此木犀科叶绿体基因组比较分析时选取了代表木犀科 5 个族的叶绿体基因组 *Forsythia mandschurica* (Forsythieae)、*Myxopyrum hainanense* (Myxopyreae)、*Fontanesia philliraeoides* subsp. *fortunei* (Fontanesieae)、*Syringa yunnanensis* (Oleaceae)、*Jasminum fruticans* (Jasmineae) 进行分析, 基于 IRscope (<http://irscope.shinyap>

ps.io/irapp)在线网站 (Amiryousefi et al., 2018) 绘制木犀科叶绿体基因组的 IR 边界图, 对其四分体结构变异及边界基因的收缩与扩张现象进行分析。同时, 以 *Forsythia mandschurica* (MH559275) 为参考序列, 分别使用 mVISTA (<https://genome.lbl.gov/vista/mvista/submit.shtml>) 在线网站中的 Shuffle-LAGAN 模型及 MAUV E 软件中的 Align sequences 对木犀科叶绿体基因组的结构变异及重排现象进行检测 (Frazer et al., 2004; Darling et al., 2004)。

3.3 通用条形码的评估及高变异片段的筛选

利用 PhyloSuite v. 1.2.2 软件分别对木犀科 27 个叶绿体基因组中 79 个共有的蛋白编码基因和 106 个基因间隔区进行提取及比对 (Zhang et al., 2020)。同时, 基于 DnaSP v. 6.0 软件分别计算其核苷酸多态性 (pi), 统计变异位点 (Variable sites) 和简约信息位点 (Parsimony informative sites) 从而筛选获得 3 个高变异片段 (Rozas et al., 2017)。为了评估国际通用条形码对木犀科植物的鉴定能力, 利用 TaxonDNA v. 1.0 软件并基于“Pairwise Summary”计算国际通用条形码和高变异片段在木犀科中的种间遗传距离并绘制种间遗传距离频率分布图 (Meier et al., 2006)。

3.4 系统发育分析

首先利用 PhyloSuite v. 1.2.2 软件提取得到 5 个不同的矩阵用于木犀科植物系统发育重建, 分别为木犀科完整的 27 个叶绿体基因组, 79 个蛋白编码基因, 106 个基因间隔区, 3 个国际通用条形码, 3 个高变异片段 (Zhang et al., 2020)。我们以山茱萸科 (Cornaceae) 山茱萸属 (*Cornus*) 的灯台树 *Cornus controversa* (NC_030260) 作为外类群 (Lee et al., 2007), 使用 PhyloSuite v. 1.2.2 软件中的 MAFFT 插件对 5 个数据集分别比对后串联成矩阵 (Zhang et al., 2020), 根据贝叶斯信息准则 Bayesian information criterion (BIC), 利用 ModelFinder 对以上数据集分别计算最佳核苷酸替代模型和最优分区策略, 分别利用最大似然法 (Maximum likelihood, ML) 和贝叶斯法 (Bayesian inference, BI) 构建木犀科系统发育树。其中利用 MrBayes v. 3.2.7 软件构建贝叶斯系统发育树, 马尔可夫链 (MCMC) 长设置为 2 000 000 代, Sampling Freq 设置为 100, Burnin Fraction 设置为 0.25, Number of Runs 设置为 2, Number of Chains 设置为 4 (Ronquist et al., 2012)。利用 Tracer v. 1.7.2 软件对所有参数的收敛情况 (Efficient Sampling Size, ESS) 进行检验 (Rambaut et al., 2018)。基于最大似然法利用 IQ-Tree 软件构建木犀科系统发育树, Bootstrap 设置为 1000 次。最后, 利用 Figtree v. 1.4.3 (<https://github.com/rambaut/figtree/releases>) 软件对获得的树文件进行可视化。

3.5 重复序列与 SSR 分析

为了探究木犀科植物中的重复序列情况, 利用 REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) (Kurtz et al., 2001) 在线网站选择以上 5 个木犀科叶绿体基因组 (*Forsythia mandschurica*、*Myxopyrum hainanense*、*Fontanesia philliraeoides* subsp. *fortunei*、*Syringa yunnanensis*、*Jasminum fruticans*) 识别鉴定其正向重复序列 (Forward repeat, F)、反向重复序列 (Reverse repeat, R)、互补重复序列 (Complement repeat, C) 和回文重复序列 (Palindromic repeat, P)。参数设置为 Hamming distance: 3, Maximum computed repeats: 5000, Minimal repeat size: 30。同时使用 MISA 在线网站 (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) (Thiel et al., 2003) 对木犀科叶绿体基因组中的简单重复序列 (Simple sequence repeat, SSR) 进行统计, 参数设置依次为单核苷酸 (Mono-nucleotide): 10, 二核苷酸 (Di-nucleotide): 5, 三核苷酸 (Tri-nucleotide): 4, 四核苷酸 (Tetra-nucleotide): 3, 五核苷酸 (Penta-nucleotide): 3, 六核苷酸 (Hexa-nucleotide): 3。

作者贡献

刘蕊是本研究的实验设计者和实验研究的执行人, 完成数据分析, 论文初稿的写作; 王贝、郭朋、刘燕培和王一涵参与实验设计和试验结果分析; 尚富德和林楠是项目的构思者及负责人, 指导实验设计、数据分析、论文写作与修改。全体作者都阅读并同意最终的文本。

致谢

本研究由河南省高等学校重点科研项目(23A180011)、河南省重大公益专项(201300110900)、河南省重点研发与推广专项(科技攻关)项目(202102110004) 和河南农业大学大学生创新训练计划项目共同资助。

参考文献

- Amiryousefi A., Hyvönen J., and Poczar P., 2018, IRscope: an online program to visualize the junction sites of chloroplast genomes, *Bioinformatics*, 34(17): 3030-3031.
- CBOL Plant Working Group., 2009, A DNA barcode for land plants, *Proceedings of the National Academy of Sciences of the United States of America*, 106(31): 12794-12797.
- Darling A.C.E., Mau B., Blattner F.R., and Perna N.T., 2004, Mauve: multiple alignment of conserved genomic sequence with rearrangements, *Genome Research*, 14(7): 1394-1403.
- Dong W.P., Li E., Liu Y.L., Xu C., Wang Y.S., Liu K.J., Cui X.Y., Sun J.H., Suo Z.L., Zhang Z.X., Wen J., and Zhou S.L., 2022, Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family, *BMC Biology*, 20 (1): 92.

- Dupin J., Raimondeau P., Hong-Wa C., Manzi S., Gaudeul M., and Besnard G., 2020, Resolving the phylogeny of the olive family (Oleaceae): confronting information from organellar and nuclear genomes, *Genes* (Basel), 11(12): 1508.
- Flora of China Editorial Board., eds., 1992, *Flora of China*, Science Press, Beijing, China, pp.2. (中国植物志编辑委员会, 编著, 1992, 中国植物志, 科学出版社, 中国, 北京, pp.2.)
- Frazer K.A., Pachter L., Poliakov A., Rubin E.M., and Dubchak I., 2004, VISTA: computational tools for comparative genomics, *Nucleic Acids Research*, 32(W): 273-279.
- Fu C.N., Mo Z.Q., Yang J.B., Cai J., Ye L.J., Zou J.Y., Qin H.T., Zheng W., Hollingsworth P.M., Li D.Z., and Gao L.M., 2022, Testing genome skimming for species discrimination in the large and taxonomically difficult genus *Rhododendron*, *Molecular Ecology Resources*, 22(1): 404-414.
- Green P.S., 2004, Oleaceae, In: Kubitzki K., and Kadereit J.W. (eds.), *The families and genera of vascular plants, Flowering plants, dicotyledons*, Springer, New York, pp. 296-306.
- Hebert P.D.N., Cywinska A., Ball S.L., and deWaard J.R., 2003, Biological identifications through DNA barcodes, *Proceedings of the Royal Society B: Biological Sciences*, 270(1512): 313-321.
- Jansen R.K., Cai Z.Q., Raubeson L.A., Daniell H., Depamphilis C.W., Leebens-Mack J., Müller K.F., Guisinger-Bellian M., Haberle R.C., Hansen A.K., Chumley T.W., Lee S.B., Peery R., McNeal J.R., Kuehl J.V., and Boore J.L., 2007, Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns, *Proceedings of the National Academy of Sciences of the United States of America*, 104(49): 19369-19374.
- Jensen S.R., 1992, Systematic implications of the distribution of iridoids and other chemical compounds in the Loganiaceae and other families of the Asteridae, *Annals of the Missouri Botanical Garden*, 79(2): 284-302.
- Johnson L.A.S., 1957, A review of the family Oleaceae, *Contributions from the New South Wales National Herbarium*, 2: 395-418.
- Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., and Drummond A., 2012, Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics*, 28(12): 1647-1649.
- Kim H.M., Oh S.H., Bhandari G.S., Kim C.S., and Park C.W., 2014, DNA barcoding of Orchidaceae in Korea, *Molecular Ecology Resources*, 14(3): 499-507.
- Kim K.J., and Lee H.L., 2004, Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants, *DNA Research*, 11(4): 247-261.
- Kode V., Mudd E.A., Lamtham S., and Day A., 2005, The tobacco plastid *accD* gene is essential and is required for leaf development, *The Plant Journal*, 44(2): 237-244.
- Kurtz S., Choudhuri J.V., Ohlebusch E., Schleiermacher C., Stoye J., and Giegerich R., 2001, REPuter: the manifold applications of repeat analysis on a genomic scale, *Nucleic Acids Research*, 29(22): 4633-4642.
- Lee H.L., Jansen R.K., Chumley T.W., and Kim K.J., 2007, Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions, *Molecular Biology and Evolution*, 24(5): 1161-1180.
- Li D.Z., and Zeng C.X., 2015, Prospects for plant DNA barcoding, *Shengwu Duoyangxing* (Biodiversity Science), 23(3): 297-298. (李德铎, 曾春霞, 2015, 植物 DNA 条形码研究展望, 生物多样性, 23(3): 297-298.)
- Li L.X., Peng J.Y., Wang D.W., and Duan A.A., 2022, Chloroplast genome phylogeny and codon preference of *Docynia longiunguis*, *Shengwu Gongcheng Xuebao* (Chinese Journal of Biotechnology), 38(1): 328-342. (李连星, 彭劲谕, 王大玮, 段安安, 2022, 长爪核[木衣]叶绿体基因组特征系统发育及密码子偏好性分析, 生物工程学报, 38(1): 328-342.)
- Meier R., Shiyang K., Vaidya G., and Ng P.K.L., 2006, DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success, *Systematic Biology*, 55(5): 715-728.
- Rambaut A., Drummond A.J., Xie D., Baele G., and Suchard M.A., 2018, Posterior summarization in Bayesian phylogenetics using Tracer 1.7, *Systematic Biology*, 67(5): 901-904.
- Ronquist F., Teslenko M., Mark P.V.D., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., and Huelsenbeck J.P., 2012, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, *Systematic Biology*, 61(3): 539-542.
- Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J.C., Guirao-Rico S., Librado P., Ramos-Onsins S.E., and Sánchez-Gracia A., 2017, DnaSP 6: DNA sequence polymorphism analysis of large data sets, *Molecular Biology and Evolution*, 34(12): 3299-3302.
- Song F., Li T., Burgess K.S., Feng Y., and Ge X.J., 2020, Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L. (Polygonaceae) in China, *BMC Plant Biology*, 20(1): 261.
- Taylor H., 1945, Cyto-taxonomy and phylogeny of the Oleaceae, *Brittonia*, 5(4): 337-367.
- Thiel T., Michalek W., Varshney R.K., and Graner A., 2003, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.), *Theoretical and Applied Genetics*, 106(3): 411-422.
- Wallander E., and Albert V.A., 2000, Phylogeny and classification of Oleaceae based on *rps16* and *trnL-F* sequence data, *American Journal of Botany*, 87(12): 1827-1841.
- Xu S.S., Gao J., Chen J.W., He J.S., Liu X.Y., and Zhang J.L., 2021, Molecular identification of medicinal *Dendrobium* and its adulterants based on DNA barcoding, *Yunnan Nongye Daxue Xuebao* (Journal of Yunnan Agricultural University), 36(5): 862-871. (徐素素, 高静, 陈军文, 赫京生, 刘祥宇, 张敬丽, 2021, 药用石斛及其混伪品的 DNA 条形码分子鉴定, 云南农业大学学报, 36(5): 862-871.)
- Zhang D., Gao F.L., Jakovčić I., Zou H., Zhang J., Li W.X., and Wang G.T., 2020, PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies, *Molecular Ecology Resources*, 20(1): 348-355.
- Zhang T.W., Fang Y.J., Wang X.M., Deng X., Zhang X.W., Hu S.N., and Yu J., 2012, The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes, *PLoS One*, 7(1): e30531.