### • 药材与资源 •

### 宽叶十万错叶绿体基因组比较分析和系统进化关系研究

赵艳林1,陈 渊1,沈晓婷1,安 昌2,王振华3,黄泽豪1,黄鸣清1,林彦翔1\*,郑燕芳1\*

1. 福建中医药大学药学院, 福建 福州 350122

2. 福建农林大学生命科学学院福建省海峡植物应用系统生物学重点实验室,福建 福州 350002

3. 宁德师范学院生物科学与工程学院,福建省特色药用植物工程技术研究中心,福建 宁德 352100

摘 要:目的 以药用植物宽叶十万错 Asystasia gangetica 为研究对象,解析其叶绿体基因组结构特征,探究宽叶十万错在 爵床族中的系统发育关系:方法 利用 Illumina NovaSeq 6000 测序平台对宽叶十万错叶绿体基因组进行测序,对其进行组 装、注释和特征序列分析,并基于 34 种爵床族物种构建系统发育树。结果 宽叶十万错叶绿体基因组为典型的四分体结构, 总序列长度为 150 307 bp,包括一个大单拷贝区(LSC, 82 332 bp),一个小单拷贝区(SSC, 17 425 bp)和一对反向重复区 (IRs, 25 275 bp);宽叶十万错叶绿体基因组共注释到光合作用基因、自我复制基因、其他基因和未知功能基因 4 类共 133 个基因,包括 88 个蛋白编码基因,37 个 tRNA 基因和 8 个 rRNA 基因;宽叶十万错叶绿体基因组共检测到 34 个 SSR 位点, SSR 类型以 A/T 为主,通过 REPuter 软件共检测到 33 个长重复序列;IR 区边界分析表明,宽叶十万错比其他属物种保守, 边界区域变异小;基于 8 个属 8 个物种全叶绿体基因组比较分析可知,基因间隔区 atpH-atpI、petN-trnD-GUC、rps4-trnT-UGU、trnF-GAA-ndhJ等的变异大于编码区;宽叶十万错密码子偏好性分析表明,甘氨酸是宽叶十万错使用频次最高的氨基 酸,以 A 或 U 结尾的密码子是宽叶十万错叶绿体基因组的偏好性密码子;ENC、PR2 和中性绘图分析表明,自然选择是影 响宽叶十万错密码子偏好性的主要原因;选择压力分析表明,有 9 个基因受到正选择,对宽叶十万错及其近缘属物种适应性 进化具有重要作用;系统发育分析表明,宽叶十万错所在的十万错属和钩粉草属形成姊妹类群,亲缘关系最近,并具有较高 的支持率,共同归为彩叶木亚族。结论 对宽叶十万错叶绿体基因组结构进行了解析,并探讨了爵床族物种系统发育关系, 丰富了爵床族药用植物遗传资源,为爵床族的物种分类提供了参考。

关键词:宽叶十万错;叶绿体基因组;密码子偏好性;选择压力;序列变异;系统发育
中图分类号: R286.12
文献标志码: A
文章编号: 0253 - 2670(2025)05 - 1731 - 16
DOI: 10.7501/j.issn.0253-2670.2025.05.022

# Comparative analysis and phylogenetic relationship of complete chloroplast genomes in *Asystasia gangetica*

ZHAO Yanlin<sup>1</sup>, CHEN Yuan<sup>1</sup>, SHEN Xiaoting<sup>1</sup>, AN Chang<sup>2</sup>, WANG Zhenhua<sup>3</sup>, HUANG Zehao<sup>1</sup>, HUANG Mingqing<sup>1</sup>, LIN Yanxiang<sup>1</sup>, ZHENG Yanfang<sup>1</sup>

- 1. College of Pharmacy, Fujian University of Traditional Chinese Medicine, Fuzhou 350122, China
- Fujian Key Laboratory of Haixia Plant Application Systems Biology, College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China
- 3. Engineering Technology Research Center of Characteristic Medicinal Plants of Fujian, School of Biological Science and Engineering, Ningde Normal University, Ningde 352100, China

**Abstract: Objective** To analyze the structural features of the chloroplast genome in medicinal plant *Asystasia gangetica* and explore its phylogenetic relationships in the tribe Justiceae. **Methods** The chloroplast genome sequence of *A. gangetica* was obtained by sequencing with Illumina NovaSeq 6000, and assembled, annotated and analyzed the characteristic sequences of the genome. A

基金项目:国家自然科学基金项目(32100168);福建省自然科学基金项目(2021J05061);福建中医药大学基础类学科科研提升计划项目 (XJC2023008);福建中医药大学校管课题(X2023025);福建中医药大学高层次人才科研启动资金项目(X2020012-人才) 作者简介:赵艳林(1998—),女,硕士研究生,研究方向为中药资源与鉴定。E-mail:3258174656@qq.com

\*通信作者:林彦翔(1991一),男,博士,副教授,主要从事中药资源学与分子生药学研究。E-mail: linyanxiang@fjtcm.edu.cn 郑燕芳(1978一),女,博士,副教授,主要从事中药药效物质基础及其作用机制研究。E-mail: yfzheng@fjtcm.edu.cn

收稿日期: 2024-08-09

phylogenetic tree was constructed based on 34 species of the tribe Justiceae. **Results** The chloroplast genome of A. gangetica exhibits a typical tetrad structure with a length of 150 307 bp, including a large single-copy region (LSC, 82 332 bp), a small single-copy region (SSC, 17 425 bp), and a pair of inverted repeat regions (IRs, 25 275 bp). The chloroplast genome of A. gangetica has been annotated with 133 genes in four categories: photosynthesis genes, self-replication genes, other genes and unknown functional genes, comprising 88 protein-coding genes, 37 tRNA genes and eight rRNA genes. A total of 34 SSR loci were detected in the chloroplast genome of A. gangetica, with A/T being the predominant SSR type; REPuter software identified 33 long repeat sequences; Analysis of IR boundaries indicated that A. gangetica was more conservative than other species, with minimal variation in the boundary region. Based on the comparative analysis of complete chloroplast genomes of eight species from eight genera, it can be concluded that the variation rates of atpH-atpI, petN-trnD-GUC, rps4-trnT-UGU, trnF-GAA-ndhJ etc in gene spacer regions were the highest. The codon usage bias analysis of A. gangetica indicated that glycine was the most frequently used amino acid in A. gangetica, and codons ending in A or U had a strong impact on codon perference. The ENC, PR2, and neutrality plot analyses indicated that natural selection was the primary factor influencing codon preference in A. gangetica. The selective pressure analysis revealed that nine genes were under positive selection, playing a crucial role in the adaptive evolution of A. gangetica and its closely related species. Phylogenetic analysis indicated that Asystasia and Pseuderanthemum form a sister group, with the closest genetic relationship and high support value, and were collectively classified as the subtribe Graptophyllinae. Conclusion This study analyzes the chloroplast genome structure of A. gangetica and explores the phylogenetic relationships within the tribe Justiceae, with view to enriching the genetic resources of medicinal plant in the tribe Justiceae and providing a reference for species classification.

Key words: Asystasia gangetica (L.) T. Anderson; chloroplast genome; codon preference; selection pressure; sequence variation; phylogeny

爵床科(Acanthaceae)作为热带至亚热带分布的植物大科,中草药资源十分丰富。《中国药典》 2020年版<sup>[1]</sup>收录的板蓝*Strobilanthes cusia*(Nees) Kuntze,其干燥根茎和根被称为"南板蓝根"<sup>[2]</sup>, 茎和叶经加工可得青黛,能够治疗病毒性肝炎、流 感、肺炎、毒蛇咬伤等<sup>[3]</sup>。穿心莲*Andrographis paniculata*(Burm. f.) Wall. ex Nees 因其富含二萜类抗 菌化合物,被誉为"中药抗生素"<sup>[4]</sup>,素有"中国的 草药之王"之称<sup>[5]</sup>。此外,尖药花*S. tomentosa*(Nees) J. R. I. Wood、枪刀药*Hypoestes purpurea*(L.) R. Br.、 九头狮子草*Peristrophe japonica*(Thunb.) Bremek.、 爵床*Justicia procumbens* L.等被收录于《全国中草 药汇编》和《中华本草》,都是具有潜在药用价值的 中草药<sup>[6-7]</sup>。

宽叶十万错 Asystasia gangetica (L.) T. Anderson 隶属于爵床科十万错属 Asystasia Blume,为多年生 草本,始载于《新华本草纲要》,别名赤道樱草、恒 河十万错。主要分布在印度、泰国、中南半岛至马 来半岛等地区<sup>[8]</sup>,在我国主要分布在云南、广东、 广西、福建和台湾等地<sup>[9]</sup>。由于其花色艳丽,可作 为室内花卉进行观赏<sup>[10]</sup>;在民间作药食两用植物, 称为"跌打草""盗偷草",可散瘀、止痛、止血、 通便、解毒<sup>[11]</sup>;现代药理研究表明,该植物含有类 固醇、氨基酸、生物碱等活性成分<sup>[12]</sup>,具有抗氧化、 抗炎、镇痛等多种生物活性<sup>[13-14]</sup>,可用于治疗支气 管炎、风湿病、毒虫咬伤等<sup>[15]</sup>,因此宽叶十万错是 一种潜力巨大的药用植物。然而,目前对该物种的 研究主要集中在药用和食用性质,宽叶十万错及其 所处的十万错属物种均无完整质体基因组发布,对 于该属系统发育研究较为有限。部分研究使用片段 构建系统发育树<sup>[16-17]</sup>,但分辨率较低,支持率不高, 物种系统关系问题仍未得到解决。

叶绿体是植物进行光合作用的主要场所。与核基 因组相比,叶绿体基因组具有母系遗传、结构高度保 守、全长序列短(120~210kb)等优点<sup>[18]</sup>,使其成为 植物种质资源鉴定、系统发育分析和分子育种等领域 研究的重要工具<sup>[19-20]</sup>。随着测序技术的进步和成本的 降低,叶绿体基因组在药用植物类群方面的研究日益 广泛,例如叶绿体基因组己应用于鸡柴骨 *Elsholtzia fruticose* (D. Don) Rehder <sup>[21]</sup>、珠子参 *Panax japonicus* var. *major* (Burkill) C. Y. Wu & K. M. Feng、羽叶三七 *P. japonicus* var. *bipinnatifidus* (Seem.) C. Y. Wu & K. M. Feng<sup>[22]</sup>、鸡屎藤 *Paederia foetida* L.<sup>[23]</sup>等药用植物的系 统发育研究。此外,叶绿体基因组也被用于不同产地 黄草乌 *Aconitum vilmorinianum* Kom.栽培品的鉴定, 以明确其栽培种源问题<sup>[24]</sup>。这些研究有助于解决系统 关系和物种鉴定问题。

叶绿体基因组的应用已经扩展到物种适应性 进化的研究领域。适应性进化是指物种在进化过程 中对环境变化的适应性提升,这一过程通常由自然 选择等进化力量驱动,与基因重组、基因突变等遗 传变异紧密相关<sup>[25]</sup>。金腰属 *Chrysosplenium* Tourn. ex L.物种的研究中已发现 15 个参与光合作用的基因正经历正选择,这有助于揭示它们对潮湿和阴凉栖息地环境的适应性进化<sup>[26]</sup>。

本研究通过对宽叶十万错高通量测序,组装并 注释质体基因组,明确其序列特征及基因结构,同 时也对其 GC 含量、简单重复序列、高变区、以及 密码子偏好性和选择压力进行分析,这些为宽叶十 万错的分子鉴定、DNA 条形码开发、分子育种、系 统发育关系和资源开发利用等研究提供参考。该研 究的目的和意义如下:(1)提供宽叶十万错完整的 叶绿体基因组,为该物种及其所属族的系统进化提 供一份基因资源;(2)明确宽叶十万错在爵床族中 的系统发育位置,丰富爵床族物种数据库;(3)检 测微卫星位点和高变区,提供可潜在开发为鉴定宽 叶十万错的 DNA 条形码。

### 1 材料

研究对象采自生长于福建省厦门市 (36°43′24.80′′N,101°44′54.11′′E),经福建中医药大 学林彦翔副教授鉴定为十万错属物种宽叶十万错*A.* gangetica (L.) T. Anderson。选择生长良好、干净的 叶片取样,在野外使用硅胶干燥保存,凭证标本存 放于福建中医药大学标本馆。后续分析使用的宽叶 十万错近缘物种质体基因组来源于 NCBI 数据库, 实验材料详细信息见表 1。

### 2 方法

### 2.1 DNA 提取与测序

采用快捷型植物基因组 DNA 提取系统试剂盒 (天根生物科技有限公司,北京,中国)提取基因

表 1 植物样品信息 Table 1 Plant samples used in this study

物种	属	GenBank 登录号
宽叶十万错 Asystasia gangetica	十万错属	PP680763
鸭嘴花 Justicia adhatoda	爵床属	MN848249
绿苞爵床 J. betonica	爵床属	MN848244
矮爵床 J. demissa	爵床属	MN885664
J. flava	爵床属	MK548577
小驳骨 J. gendarussa	爵床属	MN848252
大爵床 J. grossa	爵床属	MN848251
紫苞爵床 J. latiflora	爵床属	MN848246
南岭爵床 J. leptostachya	爵床属	MK389502
广东爵床 J. lianshanica	爵床属	MN885665
喀西爵床 J. mollissima	爵床属	MN848247
野靛棵 J. patentiflora	爵床属	MN848248
爵床 J. procumbens	爵床属	MN848245
杜根藤 J. quadrifaria	爵床属	MN848243
针子草 J. vagabunda	爵床属	MN848250
黑叶小驳骨 J. ventricosa	爵床属	MW580585
Dicliptera acuminata	狗肝菜属	MK830556
D. montana	狗肝菜属	MK833946
D. mucronata	狗肝菜属	MK848596
D. peruviana	狗肝菜属	MK833945
D. ruiziana	狗肝菜属	MK833947
D. tinctoria	狗肝菜属	OR063946
九头狮子草 Peristrophe japonica	观音草属	MW411448
白丝带木 Hypoestes forskaolii	枪刀药属	ON398071
孩儿草 Rungia pectinata	孩儿草属	MK946456
鳄嘴花 Clinacanthus nutans	鳄嘴花属	MH778102
海康钩粉草 Pseuderanthemum haikangense	山壳骨属	MT747169
金苞花 Pachystachys lutea	金苞花属	OP546128
绿虾花 Ecbolium viride	鹤扇花属	MW482858

组 DNA,超声波捣碎将基因组 DNA 打断成 350 bp 的 DNA 片段,并使用琼脂糖凝胶电泳和紫外分光 光度计检测 DNA 浓度和纯度。将合格样品进行基 因测序,通过 PCR 反应对 DNA 片段构建文库,并 利用二代测序平台 Illumina NovoSeq 6000 测序,获 取 150 bp 的双端测序短序列。测序由天津诺禾致源 生物信息技术有限公司完成。

### 2.2 叶绿体基因组的组装、注释及物理图谱绘制

为保证组装结果的准确性,在组装前,使用 Fastp v0.23.4<sup>[27]</sup>工具过滤原始数据中的低质量测序 数据,去除接头,滤去碱基 phred quality < 20 且占 比超过10%以及长度小于50 bp的双端测序短序列。 采用 Getorganelle v1.7.7.0<sup>[28]</sup>对预处理后的高质量测 序数据进行组装。选择宽叶十万错的近缘属物种绿 苞爵床 Justicia betonica(NC\_080235)作为参考基 因组,借助 CPGAVAS2<sup>[29]</sup>和 Geseq<sup>[30]</sup>在线网站进行 注释,对注释结果进行手动校正。最后,采用 OGDRAW 网站的 Chloroplot<sup>[31]</sup>工具绘制宽叶十万 错叶绿体基因组物理图谱。

### 2.3 重复序列鉴定和 SSR 分析

使用 Misa 软件鉴定宽叶十万错质体基因组 的简单重复序列(simple sequence repeat, SSR), 参数设置为单核苷酸设置为 10, 二核苷酸设置为 5, 三核苷酸设置为 4, 四、五、六核苷酸均设置 为 3。运行的结果含有较多复合型重复,因此我们 去掉可能出现的 c 和 c\*。使用 REPuter 在线网站 https://bibiserv.cebitec.unibielefeld.de/reputer?id=re puter\_view\_submission 对重复类型进行检测,包 括正向重复、反向重复、回文重复和互补重复序 列。序列之间的海明距离(hamming distance)填 3;最小重复片段大小(minimal repeat size)为 30, 最大计算序列数量(maximum computed repeats) 为 5000。

### 2.4 叶绿体基因组 IR 边界区分析

采用 IRscope 在线网站,对宽叶十万错、白丝 带木、九头狮子草、海康钩粉草、爵床、孩儿草、 D. acuminata、鳄嘴花 8 个物种的叶绿体基因组 4 个 区域的边界信息进行分析。

### 2.5 比较质体基因组分析

利用 mVISTA<sup>[32]</sup>软件,在 Shuffle-LAGAN 模式下,比较宽叶十万错及其近缘属物种质体全基因组序列。使用 MEGA7<sup>[33]</sup>对质体基因组序列进行比较,使用 DnaSP 5 <sup>[34]</sup>计算核苷酸多态性(*P<sub>i</sub>*)值,使用

常见参数设置: window length 设为 800 bp, step size 设为 200 bp。

### 2.6 密码子偏好性分析

同义密码子使用频率(Relative synonymous codon usage, RSCU),通过软件 Phylosuite v1.2.2<sup>[35]</sup> 提取宽叶十万错蛋白编码基因,过滤长度小于 300 bp 和重复序列;由于 RSCU=1 的密码子没有偏好性,终止密码子不编码任何氨基酸,因此从总的 CDS 中剔除;最后选择 CDS 以 ATG 为起始,以 TAA、TAG、TGA 为结尾的 CDS。将筛选后得到的 最终 CDS 序列导入 MEGA7.0<sup>[36]</sup>计算同义密码子使 用频率,最后通过 CondonW<sup>[37]</sup>在线工具分析宽叶十 万错使用同义密码子的情况。

有效密码子数(effective number of codon, ENCplot)分析能够直观反映物种基因的密码子使用模 式,用于衡量突变及自然选择对密码子偏好性的影 响程度<sup>[38]</sup>。统计了 GC<sub>3</sub>的含量,并使用 Emboss v6.6.0 软件的 Chips 工具计算 ENC 值<sup>[39]</sup>。以 GC<sub>3</sub>为 横坐标(*X*),ENC 为纵坐标(*Y*)绘制散点图。圆 点表示基因,如果基因接近标准曲线或位于标准曲 线上,则该基因的密码子使用偏好性主要受到突变 压力的影响<sup>[40]</sup>;相反,如果基因低于标准曲线,则 受到选择和其他因素的影响较大。

碱基奇偶偏移分析(PR2-bias plot analysis, PR2plot)用于分析各密码子第3位碱基A、T和G、C 之间的使用频率是否均衡<sup>[41]</sup>。使用软件 Phylosuite v1.2.2<sup>[35]</sup>提取宽叶十万错物种 CDS,剔除重复的、 序列长度小于 300 bp,并且选择具有4个碱并同义 密码子的氨基酸进行分析。分别以G<sub>3</sub>/(G<sub>3</sub>+C<sub>3</sub>)|4 为 横坐标,A<sub>3</sub>/(A<sub>3</sub>+T<sub>3</sub>)|4 为纵坐标绘制散点图<sup>[42]</sup>,通 过各点偏离中心点的距离来判断碱基偏移的程度 和方向。

中性绘图分析是衡量密码子偏性影响因素的 方法之一,同义密码子的突变通常发生第3位碱 基,而第1或第2位碱基的突变是非同义密码子 的突变,非同义密码子的突变率较低<sup>[43]</sup>。在中性 图中,以每个基因的 GC<sub>3</sub>为横坐标,GC<sub>12</sub>(GC<sub>1</sub> 和 GC<sub>2</sub>的平均值)为纵坐标绘制标准曲线图。图 中散点代表基因,回归系数表示密码子第1位、 第2位(GC<sub>12</sub>)含量平均值与密码子第3位 GC (GC<sub>3</sub>)含量的相关性,回归系数越接近于1,表 明 GC<sub>1</sub>、GC<sub>2</sub>和 GC<sub>3</sub>的碱基组成相似,密码子偏 好性受突变的影响更大<sup>[44]</sup>;相反,则表明3个位 置碱基组成差异大,密码子偏好性受自然选择的 影响更大。

### 2.7 选择压力分析

选择 2 种不同算法对与宽叶十万错亲缘关系最 近的 6 个爵床族物种叶绿体基因组进行了选择压力 分析。第 1 种采用蛋白编码基因串联序列计算成对 非同义(K<sub>a</sub>)与同义(K<sub>s</sub>)的比率(K<sub>a</sub>/K<sub>s</sub>)。使用 Phylosuite v1.2.2<sup>[35]</sup>将提取的 77 个蛋白编码基因串 联成 超级 阵列,利用 MAGA7.0<sup>[33]</sup>在 muscle (codons)模式下,对每个基因进行联合,最后利 用 K<sub>a</sub>K<sub>s</sub>\_Calculator v3.0.<sup>[45]</sup>计算各物种成对 K<sub>a</sub>/K<sub>s</sub>。 第 2 种是基于 YN 模型的 K<sub>a</sub>K<sub>s</sub>\_Calculator v3.0.<sup>[45]</sup>计 算非同义(K<sub>a</sub>)、同义(K<sub>s</sub>)和 K<sub>a</sub>/K<sub>s</sub>值,K<sub>a</sub>代表非 同义突变,K<sub>s</sub>代表同义突变。当 K<sub>a</sub>/K<sub>s</sub>小于 1 时, 说明 K<sub>s</sub>值更大,同义突变占主导,保留氨基酸功能; 如果 K<sub>a</sub>/K<sub>s</sub>大于 1 时,K<sub>a</sub>占主导,说明非同义突变 占主导<sup>[46]</sup>,自然选择对氨基酸产生正向选择。

### 2.8 系统发育分析

为明确宽叶十万错在爵床族中的系统发育 地位,从 NCBI 数据库下载了 34 种爵床科植物 的质体基因组序列和 6 个外类群,外类群包括 Echinacanthus attenuatus (MH045157),黄花恋岩 花 E. lofouensis (MF490441), 黑面将军 Strobilanthes crispa (MW599987), 板蓝 S. cusia (MG874806)<sup>[47]</sup>, Ruellia brittoniana (MW697905), 蓝花草 R. simplex (OQ579160)<sup>[48]</sup>。通过 CIPRES Science Gateway 网站中的 RAxML-HPC2 on ACCESS v8.2.12<sup>[49]</sup>软件, 采用最大似然法 (maximum likelihood, ML)构建系统发育树, bootstrap 重复次数设置为 1 000。

### 3 结果与分析

### 3.1 宽叶十万错叶绿体基因组基本特征

宽叶十万错叶绿体基因组是一个双链环状 DNA, 呈典型的四分体结构(图1),总长度是 150 307 bp, 包括一个长度为 82 332 bp 的大单拷贝区(large single-copy region, LSC),长度为 17 425 bp 的小单 拷贝区(small single-copy region, SSC)和一对长度 为 25 275 bp 的反向重复区(inverted repeat regions, IRa/IRb)。总的 GC(嘧啶)含量是 38.4%,AT(嘌 呤)的含量是 61.6%,说明有明显的 AT 偏向性。4 个区的 GC 含量和长度(表 2)。其中,IR 区的 GC 含量最高,达到 43.6%,其次是 LSC 区(36.4%)和 SSC 区(32.6%);造成 IR 区高 GC 含量的原因可能 是 IR 区含有高 GC 含量的 rRNA 基因。



图 1 宽叶十万错叶绿体基因组图谱 Fig. 1 Gene map of the chloroplast (cp) genome in *A. gangetica* 

Table 2     Base composition of choroplast genome in A. gangetica								
基因	T(U)/%	A/%	C/%	G/%	长度/bp	GC/%		
LSC	32.4	31.3	18.7	17.7	82 332	36.4		
SSC	33.6	33.8	17.3	15.3	17 425	32.6		
IRa	28.3	28.1	22.6	21.0	25 275	43.6		
IRb	28.1	28.3	21.0	22.6	25 275	43.6		
合计	31.1	30.5	19.6	18.8	150 307	38.4		

表 2 宽叶十万错叶绿体基因组碱基组成 able 2 Base composition of choroplast genome in *A. gangetic* 

# **3.2** 宽叶十万错叶绿体基因组的基因注释和类型分析

对宽叶十万错叶绿体基因组进行在线功能基因注释(表3),共注释到4类基因,包括光合作用基因、自我复制基因、其他基因和未知功能基因; 共注释得到133个基因,包括88个蛋白编码基因, 8个 rRNA 基因,37个 tRNA 基因。在这些基因中, 有25个基因含有内含子,其中有8个在IR 区,分 别是 ndhB(2)、rpl2(2)、trnI-GAU(2)、rps12(2);除 了 ndhA 在 SSC 区外,其余均在 LSC 区,值得注意 的是 ycf1 在 SSC 区和 IRb 区的边界重复。其中, ndhA、ndhB(2)、petB、petD、atpF、rpl16、rpl2(2)、 rps16、rpoC1、trnA-UGC(2)、trnG-UCC、trnI-GAU(2)、trnK-UUU、trnL-UAA、trnV-UAC、accD、 ycf1 基因含有 1 个内含子,而 rps12(2)、clpP、ycf3 基因含有 2 个内含子。

表 3	宽叶十万错叶绿体基因功能分类
-----	----------------

Table 3	Chloroplast gene functional classification of A. gangetica
---------	--

基因功能	基因类型	基因名
光合作用相关基因	亚光系统I	psaA, psaB, psaC, psaI, psaJ
	亚光系统 II	psbA、psbB、psbC、psbD、psbE、psbF、psbH、psbI、psbJ、psbK、
		psbL、psbM、psbN、psbT、psbZ
	NADH-脱氢酶亚基	$ndhA^*$ , $ndhB^*(2)$ , $ndhC$ , $ndhD$ , $ndhE$ , $ndhF$ , $ndhG$ , $ndhH$ , $ndhI$ ,
		ndhJ、 ndhK
	细胞色素复合物 b/f 亚基	petA、petB <sup>*</sup> 、petD <sup>*</sup> 、petG、petL、petN
	ATP 合酶亚基	$atpA_{\gamma} atpB_{\gamma} atpE_{\gamma} atpF^{*}_{\gamma} atpH_{\gamma} atpI$
	Rubisco 酶大亚基	rbcL
自我复制相关基因	核糖体蛋白大亚基	rpl14、rpl16 <sup>*</sup> 、rpl2 <sup>*</sup> (2)、rpl20、rpl22、rpl23(2)、rpl32、rpl33、rpl36
	核糖体蛋白小亚基	rps11、rps12**(2)、rps14、rps15、rps16*、rps18、rps19、rps2、 rps3、rps4、rps7(2)、rps8
	RNA 依赖核酸聚合酶	rpoA、rpoB、rpoC1 <sup>*</sup> 、rpoC2
	核糖体 RNA 基因	rrn16S (2)、rrn23S (2)、rrn4.5S (2)、rrn5S (2)
	转运 RNA 基因	trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-
		GCC、trnG-UCC <sup>*</sup> 、trnH-GUG、trnI-CAU(2)、trnI-GAU <sup>*</sup> (2)、trnK-
		UUU <sup>*</sup> 、trnL-CAA(2)、trnL-UAA <sup>*</sup> 、trnL-UAG、trnM-CAU、trnN-
		GUU(2)、trnP-UGG、trnQ-UUG、trnR-ACG(2)、trnR-UCU、trnS-
		GCU、trnS-GGA、trnS-UGA、trnT-GGU、trnT-UGU、trnV-
		$GAC(2)$ , $trnV-UAC^*$ , $trnW-CCA$ , $trnY-GUA$ , $trnfM-CAU$
其他基因	成熟酶	matK
	蛋白酶	$clp \mathtt{P}^{**}$
	包裹膜蛋白	cemA
	乙酰辅酶A羧化酶亚基	$acc D^*$
	C型细胞色素合成基因	ccsA
	转录起始因子	infA
	其他	-
未知功能基因	保守开放阅读框	ycf1*, ycf15(2), ycf2(2), ycf3**, ycf4

\*含有1个内含子;\*\*含有2个内含子。

\*Gene with one intron; \*\* Gene with one intron.

### 3.3 重复序列统计分析

共检测到宽叶十万错基因组存在 34 个 SSR 位点(表4),其中单核苷酸重复有 12 个,二核 苷酸重复有 6 个,三核苷酸重复有 4 个,四核苷 酸重复有 12 个,未发现五核苷酸和六核苷酸重 复。SSR 的类型以 A/T 为主,占总重复序列的 32.35%;其次为 AT/AT,占 17.65%,其余占比较 少。从分布区域来看,LSC 区 SSR 位点有 26 个, SSC 区位点有 9 个,在 IR 区没有 SSR 位点分布, 其中,有 8 个 SSR 位于内含子(intron),4 个 SSR 位于外显子(exon),18 个位于基因间隔区(IGS), 4 个 SSR 位于基因编码区。说明 SSRs 位点分布 是不均匀的。

使用 REPuter<sup>[50]</sup>在线工具在宽叶十万错的叶 绿体基因组中共检测到 33 个长重复序列(长度大 于 30 bp)(表 5),在 30~39 bp 内有 14 个正向重 复(forward repeats, F),占 42.42%;9 反向重复 (reverse repeats, R),占 27.27%,8 个回文重复 (palindromic repeats, P),占 24.24%。在 40~49 bp 范围内有 1 个正向重复,占 3.03%,一个回文 重复,占 3.03%。未检测到互补重复(complement repeats, C)。

### 表 4 宽叶十万错叶绿体基因组的 SSRs 类型及数量

Table 4	Simple sequence repeats of chloroplast genome in
	A. gangetica

		0 0		
核苷酸类型	SSR 重复序列	重复序列个数	占比/%	合计占比/%
单核苷酸	A/T	11	32.35	35.29
	C/G	1	2.94	
二核苷酸	AT/AT	6	17.65	17.65
三核苷酸	AAG/CTT	2	5.88	11.76
	AAT/ATT	2	5.88	
四核苷酸	AAAC/GTTT	1	2.94	35.28
	AAAG/CTTT	3	8.82	
	AAAT/ATTT	2	5.88	
	AATC/ATTG	3	8.82	
	AATT/AATT	3	8.82	

表 5 宽叶十万错的叶绿体基因组的长重复类型及数量 Table 5 Long repeats of chloroplast genome in *A*.

	gangetica		
重复长度/bp	重复类型	重复数量	占比/%
30~39	正向重复(F)	14	42.42
	反向重复(R)	9	27.27
	回文重复(P)	8	24.24
40~49	正向重复(F)	1	3.03
	回文重复(P)	1	3.03

### 3.4 IR 边界区分析

爵床科 8 个属 8 个物种叶绿体全基因组的 LSC/IRb、SSC/IRb、SSC/IRa 和 LSC/IRa 边界和基 因分布如图2所示。宽叶十万错和其余7个物种的 基因分布大致比较保守,在各边界区基因分布也基 本相同。但是也存在一些差异,对于LSC/IRb(JLB) 而言, 宽叶十万错、九头狮子草、爵床和 D. acuminata 的 rps19 基因在 LSC/IRb 的边界,并且九 头狮子草、爵床和 D. acuminate 的 rps19 基因扩张 长度相同,均为177 bp和102 bp;而海康钩粉草、 孩儿草和鳄嘴花的rps19基因在LSC/IRb的边界外; 另外,除了孩儿草外,其余7个物种的 rpl22 基因 和 rpl2 基因均分布在 JLB 的两侧。在 SSC/IRb(JSB) 边界,除鳄嘴花的 ndhF 基因分布在 JSB 的右侧外, 其余7个物种的 ndhF 基因均分布在 JSB 的边界, 基因全长在 2 291~2 303 bp。在 SSC/IRa (JSA) 边 界,除鳄嘴花的 ycfl 基因位于 JSA 边界的左侧外, 其余物种的 ycfl 基因均在 JSA 的边界, ycfl 基因全 长为5126~5441 bp,相对而言, 鳄嘴花在 JSB 和 JSA 边界的变异较其他物种大。在 LSC/IRa (JLA) 边界, 宽叶十万错、九头狮子草、爵床和 D. acuminate的 trnH基因在 JLA 边界, 而康海钩粉草、 孩儿草和鳄嘴花的 trnH 基因分布在 JLA 边界的右 侧;除了孩儿草外,其余7个物种的 rpl2 基因和 psbA 基因均分分布在 JLA 边界的两侧,相对而言, 白丝带木和孩儿草在 JLB 和 JLA 边界差异较大。 3.5 宽叶十万错叶绿体基因组比对及特异性 DNA 条码筛选

本研究通过 mVISTA 在线工具,并选择检测基 因重排和倒位的全局比对模式(Shuffle-LAGAN), 对宽叶十万错、海康钩粉草、鳄嘴花、孩儿草、爵 床、白丝带木、九头狮子草和 D. acuminate 的叶绿 体基因组全长序列进行比对分析(如图 3 所示)。从 图 3 可以看出,在物种水平上,鳄嘴花、孩儿草、 爵床、白丝带木、九头狮子草和 D. acuminate 的变 异比较大,并且非编码区的变异大于蛋白编码区, 而宽叶十万错较其他物种更为保守;从分布来看, LSC 区和 SSC 区的变异程度较大,IR 区变异程度 较小,比较保守。在 trnQ-UUG、atpF、rpoC1、rpl16、 trnC-GCA、ndhF、ndhA、ycfl 基因变异比较大;而 在基因间隔区 atpH-atpI、petN-trnD-GUC、rps4-trnT-UGU、trnF-GAA-ndhJ、atpB-rbcL、accD-psaI、ycf4cemA、petA-psbF 的变异大于基因编码区。



### 图 2 爵床科植物叶绿体全基因组 IR 边界的比较示意图

#### Fig. 2 Comparison of boundaries of eight Acanthaceae plants chloroplast genomes



图 3 爵床科 8 种植物叶绿体基因组全局比较分析

Fig. 3 Global alignment analysis of cp genomes in eight species of Acanthaceae

使用 DnaSP 5 软件对查找的高变区进行核苷酸 多态性计算。(如图 4 所示),核苷酸多态性( $P_i$ ) 的整体变化范围的 0~0.12,在图中标记了  $P_i$ >0.08 的高变区基因,分别是 LSC 区(trnT-GUU、trnC-GCA、petN)和 SSC 区(ycf1)。





## Fig. 4 Nucleotide diversity of Acanthaceae plants chloroplast genomes

3.6 宽叶十万错叶绿体基因组密码子偏好性分析

3.6.1 同义密码子分析 宽叶十万错氨基酸 RSCU 如表 6 和图 5 所示。删除 RSCU=1 和终止 密码子之后, 宽叶十万错共有 60 条密码子, 总使 用频次为15592次,使用次数最多的为GGA,共 使用 685 次, 占 4.39%, 编码的氨基酸为甘氨酸 (Gly), RSCU 为 1.52; 使用次数最少的是 UGC, 占 0.03%, 共编码 46 次, 编码的氨基酸为半胱氨 酸(Cys), RSCU 值为 0.5。 RSCU>1 的有 29 个, 其中,以A/U结尾的氨基酸有28个,占总数的 46.7%, 而以 G/C 结尾的有 1 个, 占总数的 1.7%; RSCU<1的共有31个,以G/C结尾的有28个, 占总数的 46.7%, 而以 A/T 结尾的有 3 个, 占总 数的 5.0%。综上所述,说明 RSCU>1 的氨基酸 中,多数以 A/U 结尾,说明这些是宽叶十万错叶 绿体基因组的偏好性密码子;而 RSCU<1 的氨基 酸中,多数以 G/C 结尾,说明这些是宽叶十万错 的非偏好性密码子。

**3.6.2** ENC-plot 分析 宽叶十万错有效密码子数 如表 7 和图 6-A 所示,从图中可以看出宽叶十万错 叶绿体基因组大部分的 ENC 大于 40,主要集中在 40~50,大多数基因位于标准曲线的下方,可知大 部分基因的 ENC 的理论值与实际值存在较大差异,表明自然选择是影响宽叶十万错叶绿体基因组密 码子使用偏性的重要原因,密码子使用偏性受突变 的影响较小。

表 6	宽叶十万错氨基酸的 RSCU
-----	----------------

Table 6 RSCU of amino acids of A. gangetica

氨基酸	密码子	RSCU	数量
Ala	GCU	1.87	471
	GCC	0.62	155
	GCA	1.08	272
	GCG	0.44	110
Arg	CGU	1.40	228
	CGC	0.45	74
	CGA	1.32	216
	CGG	0.48	79
	AGA	1.68	274
	AGG	0.67	109
Asn	AAU	1.55	548
<b>A</b> =	AAC	0.45	160
Asp	GAU	1.02	122
Cvs	UAC	0.58	125
Cys	UGU	0.50	46
Gln	CAA	1 49	434
Gill	CAG	0.51	150
	GAA	1.48	623
	GAG	0.52	219
Glv	GGU	1.33	310
	GGC	0.42	85
	GGA	1.52	685
	CCC	0.72	280
11.	GAU	0.75	280
H1S	CAU	1.57	397
	CAC	0.43	549
Ile	AUU	1.51	369
	AUC	0.62	375
	AUA	0.87	102
Leu	UUA	1.90	225
	UUG	1.27	117
	CUU	1.30	562
	CUC	0.35	191
	CUA	0.39	611
	CUA	0.70	200
т		0.40	309
Lys	AAA	1.49	268
	AAG	0.51	132
Phe	UUU	1.33	184
	UUC	0.67	109
Pro	CCU	1.55	334
	CCC	0.76	195
	CCA	1.06	226
	CCG	0.63	133
Ser	UCU	1.62	269
	UCC	0.95	78
	UCA	1.10	338
	UCG	0.65	160
	AGU	1.31	239
	AGC	0.38	84
Thr	ACU	1.65	485
	ACC	0.78	112
	ACA	1.16	347
	ACG	0.41	08
T.		1.60	250
I yr	UAU	1.02	550
<b>X</b> 7 1	UAC	0.38	119
val	GUU	1.52	403
	GUC	0.43	127
	GUA	1.53	460
	GUG	0.52	221



图 5 宽叶十万错叶绿体基因组相对同义密码子使用度

Fig. 5 Relative synonymous codon usage of cp genome in A. gangetica

Table 7	GC	proportion and ENC in chloroplast genomics of A. gangetica
表	ŧ7	宽叶十万错叶绿体基因组密码子 GC 含量及 ENC 值

基因 ENC		GC	2/%		基因 ENC		GC/%				
	ENC	GCall	$GC_1$	GC <sub>2</sub>	GC <sub>3</sub>		ENC	GCall	GC1	GC <sub>2</sub>	GC <sub>3</sub>
accD	52.594	38.32	41.29	40.85	32.81	psbI	39.097	36.94	51.35	29.73	29.73
atpA	45.810	39.83	55.12	39.57	24.80	psbJ	32.475	37.40	36.59	48.78	26.83
<i>atp</i> B	48.856	43.69	56.71	41.68	32.67	psbK	52.320	35.48	38.71	32.26	35.48
<i>atp</i> E	48.148	39.05	55.22	39.55	22.39	psbL	48.104	33.33	35.90	28.21	35.90
<i>atp</i> F	43.569	37.84	47.03	34.05	32.43	psbM	61.000	32.38	48.57	25.71	22.86
<i>atp</i> H	54.608	47.15	64.63	48.78	28.05	psbN	38.943	46.21	54.55	40.91	43.18
atpI	50.477	38.37	48.57	37.55	28.98	psbT	55.211	33.33	41.67	33.33	25.00
ccsA	51.559	32.29	34.89	36.14	25.86	psbZ	46.750	35.45	39.68	41.27	25.40
cemA	42.711	33.04	38.60	30.26	30.26	<i>rbc</i> L	48.751	43.99	56.91	43.71	31.34
clpP	51.794	41.79	57.87	37.56	29.95	rpl14	45.464	38.48	54.47	37.40	23.58
infA	57.686	38.46	50.00	34.62	30.77	<i>rpl</i> 16	43.657	41.67	51.47	50.74	22.79
matK	48.300	33.78	37.52	32.30	31.53	rpl2	53.282	44.12	50.91	48.36	33.09
ndhA	44.191	35.46	44.60	38.78	22.99	rpl20	42.052	34.88	37.98	42.64	24.03
ndhB	47.926	37.18	41.29	37.96	32.29	rpl22	44.928	34.79	41.88	39.38	23.12
ndhC	50.704	35.81	42.98	33.88	30.58	rpl23	47.515	37.59	42.55	41.49	28.72
ndhD	45.215	35.60	41.03	37.11	28.66	rpl32	36.900	33.33	28.85	50.00	21.15
ndhE	40.458	30.36	37.62	34.65	18.81	rpl33	46.840	37.31	32.84	37.31	41.79
ndhF	43.465	33.25	38.04	36.86	24.84	rpl36	19.942	39.47	39.47	50.00	28.95
ndhG	45.535	31.64	41.81	32.77	20.34	rpoA	52.088	35.31	44.97	32.84	28.11
ndhH	50.326	39.59	50.51	36.80	31.47	rpoB	49.507	39.53	50.61	38.56	29.41
ndhI	45.547	35.31	42.60	37.87	25.44	rpoC1	51.320	39.42	50.00	38.26	30.00
ndhJ	42.593	39.20	49.06	38.36	30.19	rpoC2	51.885	38.95	47.45	38.27	31.12
ndhK	44.602	38.35	43.81	44.25	26.99	rps11	53.451	47.48	54.68	57.55	30.22
petA	49.100	39.36	51.71	37.69	28.66	rps12	45.130	42.20	50.81	49.19	26.61
petB	45.535	40.28	47.69	41.67	31.48	rps14	35.295	39.93	42.57	48.51	28.71
petD	39.973	37.89	51.55	39.13	22.98	rps15	47.661	33.70	37.36	30.77	32.97
petG	50.500	36.84	47.37	31.58	31.58	rps16	43.933	33.33	50.56	30.34	19.10
petL	57.500	29.17	34.38	37.50	15.62	rps18	46.805	35.29	36.27	41.18	28.43
petN	31.511	43.33	50.00	43.33	36.67	rps19	36.825	31.90	45.16	35.48	15.05
psaA	49.535	42.92	52.46	43.28	33.02	rps2	51.826	39.38	44.73	43.04	30.38
psaB	47.319	40.54	48.84	42.99	29.80	rps3	52.157	35.59	46.40	31.98	28.38
psaC	59.163	42.68	45.12	53.66	29.27	rps4	51.709	38.94	51.49	37.13	28.22
psaI	46.429	36.94	51.35	29.73	29.73	rps7	44.343	39.96	52.56	44.87	22.44
psaJ	35.548	41.48	46.67	40.00	37.78	rps8	42.266	36.79	40.74	40.74	28.89
psbA	40.248	41.45	49.29	43.34	31.73	ycf1	48.387	31.30	37.77	29.09	27.06
psbB	50.464	44.14	54.22	46.37	31.83	ycf15	37.696	37.40	36.59	26.83	48.78
psbC	45.452	43.53	52.11	45.99	32.49	ycf2	53.596	38.71	42.15	35.77	38.21
psbD	43.088	42.00	51.69	43.22	31.07	ycf3	55.739	38.66	47.34	37.87	30.77
psbE	53.089	40.87	42.86	46.43	33.33	ycf4	54.649	39.46	45.41	42.16	30.81



A-ENC-plot 绘图分析; B-PR2-plot 分析; C-中性绘图分析。 A-ENC-plot analysis; B-Analysis of PR2-plot; C-Neutrality plot analysis.



3.6.3 PR2-plot 偏倚分析 宽叶十万错 PR2-plot 分析结果如图 6-B 所示。氨基酸密码子的第 3 位碱基 分布不均匀, G<sub>3</sub>/(G<sub>3</sub>+C<sub>3</sub>)|4 小于 0.5 的基因有 33 个,而 A<sub>3</sub>/(A<sub>3</sub>+T<sub>3</sub>)|4 大于 0.5 的基因有 11 个,说明 在总体上,第 3 位碱基的 A/T 或 G/C 使用不均等。由结果可知,4 个区域内的点分布不均匀,在垂直 方向上多数基因位于中线的下方;在水平方向上多数基因位于中线偏左侧,垂直方向上的基因分布差 异较大,表明了宽叶十万错植物多数基因受到自然 选择的影响。

**3.6.4** 中性分析 宽叶十万错中性分析如图 6-C 所示, GC<sub>12</sub>(GC<sub>1</sub>和 GC<sub>2</sub>的平均值)的取值在 0.3171~0.5671, GC<sub>3</sub>的取值在 0.1501~0.4878。*R*<sup>2</sup>=0.0159,可知 GC<sub>12</sub>与 GC<sub>3</sub>之间联系较低,说明宽叶十万错叶绿体基因组密码子的 3 个碱基在组成上存在较大差异。回归曲线的斜率为 0.1313,表明宽叶十万错进化中所受到的突变压力占比为 13.13%,占比小,而自然选择压力为 86.87%,占比大。因此,宽叶十万错叶绿体基因密码子偏好性主要影响来自于自然选择压力。

### 3.7 选择压力分析

本研究通过构建 77 个蛋白编码序列的超级矩阵,进行 K<sub>a</sub>/K<sub>s</sub> 计算来评估 6 个叶绿体基因组的选择压力。大部分 K<sub>a</sub>/K<sub>s</sub> 值在 0.2~0.3 (图 7-A),表明这些叶绿体基因组发生了纯化选择。为了研究爵床族物种在叶绿体基因组中的进化特征,对宽叶十万错叶绿体基因组及其 5 个近缘属物种的 77 个共有蛋白编码基因进行 K<sub>a</sub>/K<sub>s</sub> 计算。大部分与亚光系统、NADH-脱氢亚基、细胞色素复合物 b/f 亚基、ATP 合酶亚基等相关的基因,如 psaA、psaB、psbA、

psbB、ndhA、ndhB、petA、atpE、rbcL等基因的K<sub>a</sub>/K<sub>s</sub> 值均小于1(图7-B),表明这些关键基因可能受到 纯化选择。此外与核糖体蛋白大亚基、核糖体蛋白 小亚基、RNA 依赖核酸聚合酶、包裹酶蛋白、C型 细胞色素合成以及未知功能相关的基因,如rpl20、 rps8、rps15、rps16、rps18、rpoA、ccsA、cemA 和 ycf2 基因的K<sub>a</sub>/K<sub>s</sub>值均大于1,表明这些基因可能在 宽叶十万错及其近缘种的分化过程中起到了重要 作用。值得注意的是 rps16 基因在宽叶十万错和大 爵床的K<sub>a</sub>/K<sub>s</sub>为1.20095,但是在绿虾花和大爵床与 大爵床和金苞花的K<sub>a</sub>/K<sub>s</sub>分别达到了2.30577和 2.32737。该结果表明该基因对宽叶十万错和大爵床 物种的分化影响不大,但是在促进绿虾花与大爵床 和金苞花分化的过程中发挥了重要作用。

### 3.8 系统发育分析

为确定宽叶十万错在爵床族的系统发育位 置,本研究选取爵床族 12 个属 34 个物种,包括 爵床属 14 个物种,狗肝菜属 6 个物种,观音草 属、枪刀药属、孩儿草属、鳄嘴花属、山壳骨属、 金苞花属和鹤扇花属物种各 1 个,马蓝属和恋岩 花属物种各 2 个,并以 2 个芦莉草属物种作为外 类群。基于宽叶十万错和其余 34 个物种构建 ML 树,结果如图 8 所示。从图中可以看出,爵床族 主要被分成 7 个分支,4 个亚族。爵床属和孩儿 草属聚为一支,归为爵床亚族,说明两者的亲缘 关系较近,并具有较高的支持率;枪刀药属、观 音 草 属 和 狗 肝 菜 属 聚 为 一 支,共同属于 Diclipterinae 亚族,说明三者的亲缘关系较近。 而爵床亚族和 Diclipterinae 亚族聚为一个大的分 支,呈姐妹关系。塔鹃花亚族包含鹤扇花属、金





A-宽叶十万错近缘属物种的成对 K<sub>a</sub>/K<sub>s</sub>值; B-宽叶十万错近缘属物种不同基因的成对 K<sub>a</sub>/K<sub>s</sub>值。

A-Pairwise Ka/Ks ratios among closely related species of A. gangetica; B-Pairwise Ka/Ks ratios of different genes among closely related species of A. gangetica.





### 图 8 基于叶绿体全基因组构建爵床族的 ML 系统发育树 Fig. 8 Construction of ML phylogenetic tree of tribe Justiceae based on chloroplast genomes

苞花属、鳄嘴花属和爵床属;然而爵床属不仅在 爵床亚族,也在塔鹃花亚族,进一步说明了爵床 属是并系类群。宽叶十万错所在的十万错属和钩 粉草属形成姊妹类群并位于基部,亲缘关系最 近,并具有较高的支持率,共同归为彩叶木亚族。 值得注意的是, *D. tinctoria* 从 Diclipterinae 亚族 中分离出来,单独聚为一个分支,和狗肝菜属的 其他物种亲缘关系较远,值得进一步研究。

4 讨论

### 4.1 叶绿体基因组特征比较分析

宽叶十万错隶属于爵床科十万错属多年生草 本,具有较高的经济价值、药用价值和食源价值。

目前对它的研究很少,而完整的十万错属叶绿体基 因组尚处于空白状态。本研究对宽叶十万错叶绿体 基因组进行测序、组装和注释,并对其结构进行了 分析。研究结果表明, 宽叶十万错叶绿体基因组全 长为 150 307 bp, 总的 GC 含量为 38.4%, 4 个区域 的 GC 含量不同, IR 区的 GC 含量高于 LSC 区和 SSC 区,可能是由于 IR 区含有高含量的 rRNA 基 因,这与之前报道的爵床族物种爵床,九头狮子草, 孩儿草,绿虾花的结果保持一致[51-53]。此外,在功 能基因注释中,存在5个未知功能基因,包括 ycfl、 ycf15、ycf2、ycf3 和 ycf4。而 ycf15 为假基因,该基 因存在在大多数被子植物中,功能尚未清楚,在油 桐叶绿体基因组中也有报道[54]。

在叶绿体基因组中, IR 区边界的扩张和收缩普 遍存在。本研究基于爵床族 IR 边界分析结果表明, 宽叶十万错、九头狮子草、爵床和 D.acuminate 在 各个边界都较保守。在大多数被子植物中, JSA 边 界的 ycfl 基因假基因化, 三叶崖爬藤[55]和藜芦属药 用植物<sup>[56]</sup>中也报道过此现象。但是本研究中, ycfl 基因属于正常基因,未假基因化。mVISTA 分析发 现,爵床族物种在非编码区的变异大于蛋白编码 区,但是宽叶十万错相较其他属物种表现保守。通 过 DnaSP 5 筛选出 4 个高变区(trnT-GUU、trnC-GCA、petN、ycf1),这些高变区域为爵床族物种的 分子标记开发和种质资源评价提供科学依据。

SSR 标记由于具有多态性丰富、提供遗传信息 多和操作便利等优点,在许多遗传研究中被广泛使 用[57]。本研究通过分析宽叶十万错叶绿体基因组, 共检测到 34 个 SSR 位点,主要位于 LSC 区。以单 核苷酸 A/T 为主(32.53%),其次为二核苷酸 AT/AT (17.65%)。有明显的 AT 碱基偏向性, 这与尚明越 等[58]报道的濒危紫皮石斛叶绿体基因 SSRs 重复单 元主要以 A 和 T 碱基组成为主以及王星璐等<sup>[59]</sup>报 道的远志叶绿体基因组 A/T 含量远高于 G/C 含量 结果一致。这些重复系列和特征为十万错属的物种 鉴定、群体遗传等研究提供参考。

密码子偏好性分析有助于新基因的发现和对 基因功能的预测[60]。影响基因密码子偏好性包括碱 基组成差异、tRNA 分子丰度、mRNA 二级结构、 突变偏好性和翻译选择等多种因素[61]。其中碱基组 成是关键因素, GC 含量是衡量基因整体突变的重 要指标。本研究中宽叶十万错共编码 60 条密码子, 其中甘氨酸使用频次最高,对应的密码子是 GGA;

其中, RSCU>1 的有 29 个, 其中 28 个氨基酸以 A 或U结尾,这些氨基酸是宽叶十万错的偏好性密码 子。这与之前报道的黄瓜 Cucumis sativus L., 九翅 豆蔻 Amomum maximum Roxb., 卷丹百合 Lilium lancifolium Ker Gawl.等[62-64]密码子偏好性分析中偏 好性密码子主要以 A 或 U 结尾的结果一致。ENCplot 绘图显示,分布在标准曲线下方的基因偏多, 表明自然选择是宽叶十万错密码子偏好性的主要 原因; PR2-plot 绘图分析显示, 宽叶十万错叶绿体 基因分布不均匀,表明影响宽叶十万错密码子偏好 性的因素不止一个,还可能与自身基因突变有关; 中性绘图分析显示,宽叶十万错回归曲线斜率接近 零,这进一步支持了选择压力对其叶绿体基因密码 子偏好性有较大影响的观点。这些发现与大多数被 子植物如黄精属 Polygonatum Mill.物种,木薯 Manihot esculenta Crantz 等选择压力是影响其密码 子偏好性的原因的结果相一致[65-66]。

正选择被认为是生物体适应环境变化的关键 因素,而纯化选择是一种普遍存在的进化力量,它 在漫长的进化过程中维持基因组序列的稳定性。 K<sub>a</sub>/K<sub>s</sub>的比值已被广泛用作区分正选择和纯化选择 下的蛋白编码基因的重要指标[67]。本研究确定了 9 个携带正选择位点的基因,而这些基因大部分是编 码核糖体亚基蛋白的基因(rps8、rps15、rps16、 rps18),它们对叶绿体生物体发生合成和功能至关 重要,这意味着爵床族物种可能通过调节叶绿体中 核糖体亚基蛋白的编码来增强进化适应性。

### 4.2 系统发育分析

爵床族是爵床亚科最大的族,约有2000种, 由于其形态变异程度大,族内分类问题一直未得到 解决。最新的研究表明[68], 爵床族内被分成6个主 要分支。系统发育树揭示了爵床属是并系群,该属 物种同时出现在塔鹃花亚族和爵床亚族分支中。本 研究通过宽叶十万错完整叶绿体基因组构建 ML 树, 结果显示爵床族被分成7个分支,与Yaradua 等[53] 构建的树不同之处在于 D. tinctoria 从狗肝菜属中分 离出来,和爵床属其他物种聚为一支。Le 等[68]基于 D. tinctoria 完整叶绿体构建的系统发育树和本实验 结果保持一致。此外,在1820年的时候 D. tinctoria 曾被鉴定为 J. tinctoria, 两者属于同种异名的物种, 并且系统发育树显示 D. tinctoria 和其他爵床属的物 种聚在一起,由此推断 D. tinctoria 可能位于爵床亚 族II, 而不是位于 Diclipterinae III, 但是 D. tinctoria

的具体位置还需要进一步研究进行验证。

本研究也将爵床族的重要中药资源纳入系统 发育分析, 被誉为"中国草药之王"的板蓝和黑面 将军聚为一支,成姊妹关系,并具有很高的支持率; 具有潜在药用价值的中草药九头狮子草和爵床分 别与白丝带木和矮爵床聚为一支, 成姊妹关系, 这 对寻求新的可替代药源和促进重要资源的可持续 开发利用具有重要意义。此外本研究结果显示,塔 鹃花亚族包含鹤扇花属、金苞花属、鳄嘴花属和爵 床属,和 Yaradua 等<sup>[53]</sup>构建系统树分支Ⅱ的成员组 成一致。另外本课题组研究发现,彩叶木亚族包含 十万错属和钩粉草属,在先前的基础上,多了新的 成员十万错属,并有很高的支持率。这为爵床族的 分类提供了一份基因组资源,也为十万错属物种鉴 定提供了参考,但是由于该研究仅测序并组装了十 万错属一个物种,未来研究还需要测序更多十万错 属和钩粉草属物种进行验证。另外,系统发育树显 示, 宽叶十万错位于爵床族的基部, 代表该族比较 古老的类群,研究基部类群利于对原始祖先和起源 分化关系进行推断。

利益冲突 所有作者均声明不存在利益冲突

#### 参考文献

- [1] 中国药典 [S]. 一部. 2020: 1902.
- [2] 肖启蒙. 板蓝与葡萄、漾濞槭基因组的共线性分析 [J]. 南方农业, 2023, 17(15): 12-15.
- [3] Gu W, Zhang Y, Hao X J, *et al*. Indole alkaloid glycosides from the aerial parts of *Strobilanthes cusia* [J]. *J Nat Prod*, 2014, 77(12): 2590-2594.
- [4] 魏思敏,刘强,苏思琪,等.穿心莲纳米银制备及抗氧 化和抑菌活性研究 [J].中草药,2024,55(19):6508-6518.
- [5] 江楠,马欣.近 20 年穿心莲内酯的药理作用研究进展[J].辽宁中医药大学学报,2024,26(12): 203-208.
- [6] 国家中医药管理局《中华本草》编委会.中华本草(7)[M].上海:上海科学技术出版社,1999:1020.
- [7] 《全国中草药汇编》编写组编. 全国中草药汇编(上)[M]. 北京:人民卫生出版社, 1975: 1009.
- [8] 谢宇,周重建.中国中草药彩色图鉴大全集 [M].长沙:湖南科学技术出版社,2018:693.
- [9] 中国科学院中国植物志编辑委员会编著.中国植物志(第70卷)[M].北京:科学出版社, 2016: 397.
- [10] 刘兴剑. 宽叶十万错 [J]. 花木盆景: 花卉园艺, 2017(6): 39.
- [11] 陆洁梅,杜怡然,王奕恒,等.有一定开发潜力的野生 蔬菜: 宽叶十万错 [J].蔬菜,2020(9):75-76.

- [12] Ungay Barbaza M Y, De Castro-Cruz K A, Hsieh C L, et al. Determination of the chemical constituent contents and antioxidation properties of Asystasia gangetica [J]. Indian J Pharm Educ Res, 2021, 55(3): 863-871.
- [13] Adeyemi O O, Aigbe F R, Uyaiabasi N G. Analgesic and anti-inflammatory activities of the aqueous stem and leaf extract of *Asystasia gangetica* (Linn) T. Anderson [J]. *Nig Q J Hosp Med*, 2011, 21(2): 129-134.
- [14] Akah P A, Ezike A C, Nwafor S V, et al. Evaluation of the anti-asthmatic property of *Asystasia gangetica* leaf extracts [J]. *J Ethnopharmacol*, 2003, 89(1): 25-36.
- [15] 江纪武. 世界药用植物速查辞典 [M]. 北京: 中国医药科技出版社, 2015: 1378.
- [16] McDade L A, Daniel T F, Kiel C A, et al. Phylogenetic relationships among acantheae (Acanthaceae): Major lineages present contrasting patterns of molecular evolution and morphological differentiation [J]. 2005, 30(4): 834-862.
- [17] Kiel C A, McDade L A, Daniel T F, et al. Phylogenetic delimitation of isoglossinae (Acanthaceae: Justicieae) and relationships among constituent Genera [J]. Taxon, 2006, 55(3): 683-694.
- [18] Daniell H, Lin C S, Yu M, et al. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. Genome Biol, 2016, 17(1): 134.
- [19] Asaf S, Waqas M, Khan A L, et al. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species [J]. Front Plant Sci, 2017, 8: 304.
- [20] Liu X, Zhou B Y, Yang H Y, et al. Sequencing and analysis of Chrysanthemum carinatum schousb and Kalimeris indica. the complete chloroplast genomes reveal two inversions and rbcL as barcoding of the vegetable [J]. Molecules, 2018, 23(6): 1358.
- [21] 余潇, 宋雨茹, 赵振宁. 鸡骨柴及香薷属药用植物叶绿 体基因组比较分析 [J]. 西南农业学报, 2024, 37(7): 1420-1434.
- [22] 彭毅丹,马楠,叶奕含,等.秦巴山区三种人参属药用 植物叶绿体基因组特征分析[J]. 广西植物,2024,44(6): 1138-1150.
- [23] 吴民华, 吴子健, 叶晓霞, 等. 鸡矢藤叶绿体基因组分析 [J]. 河南农业科学, 2024, 53(1): 70-77.
- [24] 施晓静,程子丹,张颖敏,等.基于全叶绿体基因组分 析的栽培黄草乌基源研究 [J]. 广西植物,2023,43(10): 1892-1906.
- [25] Scott-Phillips T C, Laland K N, Shuker D M, et al. The niche construction perspective: A critical appraisal [J]. Evolution, 2014, 68(5): 1231-1243.

• 1745 •

- [26] Wu Z H, Liao R, Yang T G, et al. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae) [J]. *BMC Genomics*, 2020, 21(1): 621.
- [27] Chen S F, Zhou Y Q, Chen Y R, et al. Fastp: An ultra-fast all-in-one FASTQ preprocessor [J]. Bioinformatics, 2018, 34(17): i884-i890.
- [28] Jin J J, Yu W B, Yang J B, et al. GetOrganelle: A fast and versatile toolkit for accurate *de novo* assembly of organelle genomes [J]. *Genome Biol*, 2020, 21(1): 241.
- [29] Shi L C, Chen H M, Jiang M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. *Nucleic Acids Res*, 2019, 47(W1): W65-W73.
- [30] Tillich M, Lehwark P, Pellizzer T, et al. GeSeq versatile and accurate annotation of organelle genomes [J]. Nucleic Acids Res, 2017, 45(W1): W6-W11.
- [31] Zheng S Y, Poczai P, Hyvönen J, et al. Chloroplot: An online program for the versatile plotting of organelle genomes [J]. Front Genet, 2020, 11: 576124.
- [32] Poliakov A, Foong J, Brudno M, et al. GenomeVISTA: An integrated software package for whole-genome alignment and visualization [J]. *Bioinformatics*, 2014, 30(18): 2654-2655.
- [33] Tamura K, Stecher G, Peterson D, et al. MEGA6: Molecular evolutionary genetics analysis version 6.0 [J]. Mol Biol Evol, 2013, 30(12): 2725-2729.
- [34] Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data [J]. *Bioinformatics*, 2009, 25(11): 1451-1452.
- [35] Zhang D, Gao F L, Jakovlić I, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies [J]. Mol Ecol Resour, 2020, 20(1): 348-355.
- [36] Kumar S, Stecher G, Peterson D, et al. MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis [J]. Bioinformatics, 2012, 28(20): 2685-2686.
- [37] Fuglsang A. Codon optimizer: A freeware tool for Codon optimization [J]. Protein Expr Purif, 2003, 31(2): 247-249.
- [38] Chen X H, Zhao Y D, Xu S H, et al. Analysis of Codon usage bias in the plastid genome of Diplandrorchis sinica (Orchidaceae) [J]. Curr Issues Mol Biol, 2024, 46(9): 9807-9820.
- [39] Li G L, Pan Z L, Gao S C, et al. Analysis of synonymous Codon usage of chloroplast genome in Porphyra umbilicalis [J]. Genes Genomics, 2019, 41(10): 1173-1181.

- [40] 邢朝斌,曹蕾,周秘,等. 刺五加叶绿体基因组密码子的用法分析 [J]. 中国中药杂志, 2013, 38(5): 661-665.
- [41] Guo X N, Wang Y L, Wang S M. Complete Chloroplast Genome Sequences from Yellowhorn (Xanthoceras sorbifolia) and Evolution Analysis Based on Codon Usage Bias [J]. *International J Agric Biol*, 2020, 24(4): 676-684.
- [42] Chi X F, Zhang F Q, Dong Q, et al. Insights into comparative genomics, Codon usage bias, and phylogenetic relationship of species from Biebersteiniaceae and Nitrariaceae based on complete chloroplast genomes [J]. Plants, 2020, 9(11): 1605.
- [43] Tang D F, Wei F, Cai Z Q, et al. Analysis of Codon usage bias and evolution in the chloroplast genome of Mesona chinensis Benth [J]. Dev Genes Evol, 2021, 231(1/2): 1-9.
- [44] Xu C, Cai X N, Chen Q Z, et al. Factors affecting synonymous Codon usage bias in chloroplast genome of Oncidium Gower Ramsey [J]. Evol Bioinform Online, 2011, 7: 271-278.
- [45] Zhang Z. KaKs\_Calculator 3.0: Calculating selective pressure on coding and non-coding sequences [J]. *Genomics Proteomics Bioinformatics*, 2022, 20(3): 536-540.
- [46] Huh M K. The rates of synonymous and nonsynonymous substitutions in *Sorbus aucuparia* using nuclear and chloroplast genes [J]. *J Life Sci*, 2010, 20(4): 481-486.
- [47] Chen H M, Shao J J, Zhang H, et al. Sequencing and analysis of Strobilanthes cusia (nees) kuntze chloroplast genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in angiosperm [J]. Front Plant Sci, 2018, 9: 324.
- [48] Siyi F, Liu Z J, Hui T. The complete chloroplast genome of ornamental plant *Ruellia simplex* C.Wright (Acanthaceae)
  [J]. *Mitochondrial DNA B Resour*, 2021, 6(7): 2017-2018.
- [49] Miller M A, Schwartz T, Pickett B E, et al. A RESTful API for access to phylogenetic tools via the CIPRES science gateway [J]. Evol Bioinform Online, 2015, 11: 43-48.
- [50] Kurtz S, Choudhuri J V, Ohlebusch E, et al. REPuter: The manifold applications of repeat analysis on a genomic scale [J]. Nucleic Acids Res, 2001, 29(22): 4633-4642.
- [51] Wang W, Xu T, Lu H B, et al. Chloroplast genome of Justicia procumbens: Genomic features, comparative analysis, and phylogenetic relationships among justicieae species [J]. J Appl Genet, 2024, 65(1): 31-46.
- [52] Chen J X, Wang L, Zhao Y C, et al. The complete chloroplast genome of a Chinese medicinal plant, *Peristrophe japonica* (Thunb.) Bremek. (Lamiales: Acanthaceae) from Nanjing, China [J]. *Mitochondrial DNA B Resour*, 2021, 6(7): 1888-1889.

- [53] Yaradua S S, Yessoufou K. Chloroplast genome of *Ecbolium viride* (forssk.) alston: Plastome evolution and phylogenomics of justiceae (Acanthaceae, acanthoideae) [J]. *Genome*, 2024, 67(8): 267-280.
- [54] 李泽. 油桐光合生理特性及叶绿体基因组的研究 [D]. 长沙: 中南林业科技大学, 2017.
- [55] 蒋明, 王军峰, 应梦豪, 等. 三叶崖爬藤叶绿体基因组的组装与序列分析 [J]. 中草药, 2020, 51(2): 461-468.
- [56] 田星, 刘莹莹, 张颖敏, 等. 藜芦属药用植物的叶绿体 基因组比较分析和系统发育研究 [J]. 中草药, 2022, 53(4): 1127-1137.
- [57] Zhang Z C, Hou X L. Strategies for development of SSR molecular markers [J]. *Yi Chuan*, 2004, 26(5): 763-768.
- [58] 尚明越, 王嘉乐, 周莹, 等. 濒危紫皮石斛叶绿体基因 组结构及系统发育分析 [J]. 中草药, 2023, 54(19): 6424-6433.
- [59] 王星璐,李慧娟,赵伟,等.远志叶绿体基因组序列特征与系统发育分析 [J]. 中草药, 2023, 54(11): 3655-3665.
- [60] 罗永坚, 王茹, 赵仁菲, 等. 珙桐叶绿体基因组同义密码子使用偏好性分析 [J]. 北京林业大学学报, 2024, 46(3): 8-16.
- [61] Wang Y Z, Jiang D C, Guo K, et al. Comparative analysis of Codon usage patterns in chloroplast genomes of ten Epimedium species [J]. BMC Genom Data, 2023, 24(1): 3.

- [62] Niu Y, Luo Y Y, Wang C L, et al. Deciphering Codon usage patterns in genome of Cucumis sativus in comparison with nine species of Cucurbitaceae [J]. Agronomy, 2021, 11(11): 2289.
- [63] 傅武祥, 郭雨桐, 张雪梅, 等. 九翅豆蔻叶绿体基因组 特征及其密码子偏好性分析 [J]. 基因组学与应用生 物学, 2024, 43(1): 17-30.
- [64] 邢传吉,魏蕾,陈绪清,等.卷丹百合(Lilium lancifolium)叶绿体基因组密码子偏好性分析 [J].分子 植物育种, 2023, 32: 1-13.
- [65] Shi N X, Yuan Y W, Huang R J, et al. Analysis of Codon usage patterns in complete plastomes of four medicinal Polygonatum species (Asparagaceae) [J]. Front Genet, 2024, 15: 1401013.
- [66] Geng X S, Huang N, Zhu Y L, et al. Codon usage bias analysis of the chloroplast genome of cassava [J]. S Afr N J Bot, 2022, 151: 970-975.
- [67] Zhou T, Chen C, Wei Y, et al. Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* species [J]. Front Plant Sci, 2016, 7: 1512.
- [68] Le T T N, Vu M T, Do H D K. The complete chloroplast genome of *Dicliptera* tinctoria (Nees) Kostel. and comparative analysis of chloroplast genomes in Acanthaceae [J]. *Genet Mol Biol*, 2024, 47(2): e20230297.

[责任编辑 时圣明]