生物多样性 BIODIVERSITY SCIENCE

2026, 34 (5): 25475, pages 1–17
doi: 10.17520/biods.2025475
cstr: 32101.14.biods.2025475

•遗传多样性及保护专题•

基于叶绿体基因组构建的系统发育关系应慎用于物种鉴定和生物地理学研究：以枫杨属为例

王姗姗[1,2], 刘琦[2,3], 李玉[2], 王天瑞[2], 戴锡玲[3], Gregor Kozlowski[2,4,5], 许瑾[1*], 宋以刚[2*]

1. 上海应用技术大学城市建设与生态技术学部, 上海 201418, 中国; 2. 上海辰山植物园华东植物迁地保护与利用国家林业和草原局重点实验室, 上海 201602, 中国; 3. 上海师范大学生命科学学院, 上海 200234, 中国; 4. Department of Biology and Botanic Garden, University of Fribourg, Fribourg 1700, Switzerland; 5. Natural History Museum Fribourg, Fribourg 1700, Switzerland

摘要: 叶绿体基因组因其结构保守、母系单亲遗传以及拷贝数高等特点, 曾被广泛应用于植物物种的识别、界定以及系统发育关系的重建。然而, 叶绿体基因组无法有效反映花粉介导的基因流, 对遗传漂变较为敏感, 且容易受到不完全谱系分选和杂交事件的影响。因此, 在植物鉴定和种群演化历史研究中, 如果仅依赖叶绿体基因组证据, 可能导致结论存在偏差或不够全面。本研究对枫杨属(*Pterocarya*)的7个个体进行了叶绿体全基因组测序、组装与注释, 并结合已发表数据开展了比较基因组与系统发育分析。结果显示, 叶绿体基因组长度介于160,176–160,318 bp, 均具典型四分体结构, 编码131个基因(86个蛋白编码基因、8个rRNA和37个tRNA)。简单重复序列(SSRs)位点数量为87–96个, 以单核苷酸重复为主。比较基因组分析结果表明, 枫杨属物种间的叶绿体基因组整体呈现高度保守性, 其中, 反向重复区和编码区的序列保守性高于单拷贝区和非编码区。值得注意的是, 在编码区中, *ndhF*基因的序列变异程度显著高于其他蛋白编码基因。系统发育分析结果显示, 基于不同数据分区(全基因组、蛋白质编码序列、高变区与滑窗)构建的系统树呈现2–5分支的多种拓扑结构。枫杨(*P. stenoptera*)、甘肃枫杨(*P. macroptera* var. *macroptera*)和华西枫杨(*P. insignis*)等物种个体在不同分析中均出现谱系混杂与跨分支分布现象。这些结果表明, 叶绿体基因组在解析枫杨属物种鉴定及生物地理历史方面存在明显局限, 单纯依赖其数据易导致系统性偏差。因此, 未来的植物系统发育研究有必要整合核基因组数据并采用多物种溯祖分析方法, 以实现更准确的物种鉴定与演化关系重建。

关键词: 叶绿体基因组; 枫杨属; 系统发育分析; 生物地理学

王姗姗, 刘琦, 李玉, 王天瑞, 戴锡玲, Gregor Kozlowski, 许瑾, 宋以刚 (2026) 基于叶绿体基因组构建的系统发育关系应慎用于物种鉴定和生物地理学研究: 以枫杨属为例. 生物多样性, 34, 25475. doi: 10.17520/biods.2025475; cstr: 32101.14.biods.2025475.
Wang SS, Liu Q, Li Y, Wang TR, Dai XL, Kozlowski G, Xu J, Song YG (2026) Cautions on using chloroplast genome-based phylogeny for species identification and biogeography: A case study of *Pterocarya*. Biodiversity Science, 34, 25475. doi: 10.17520/biods.2025475; cstr: 32101.14.biods.2025475.

Cautions on using chloroplast genome-based phylogeny for species identification and biogeography: A case study of *Pterocarya*

Shanshan Wang[1,2], Qi Liu[2,3], Yu Li[2], Tianrui Wang[2], Xiling Dai[3], Gregor Kozlowski[2,4,5], Jin Xu[1*], Yigang Song[2*]

1 *School of Urban Construction and Ecological Technology, Shanghai Institute of Technology, Shanghai 201418, China*
2 *Key Laboratory of National Forestry and Grassland Administration on East China Plant Conservation and Utilization, Shanghai Chenshan Botanical Garden, Shanghai 201602, China*
3 *College of Life Sciences, Shanghai Normal University, Shanghai 200234, China*
4 *Department of Biology and Botanic Garden, University of Fribourg, Fribourg 1700, Switzerland*
5 *Natural History Museum Fribourg, Fribourg 1700, Switzerland*

ABSTRACT

Aim: The chloroplast genome has been widely used for plant species identification, delimitation, and phylogenetic

收稿日期 Received: 2025-12-12; 接受日期 Accepted: 2026-03-12
基金项目: 上海市绿化和市容管理局科研专项(G212406; G242414; G242416)和上海应用技术大学中青年教师科技人才发展基金(ZQ2022-17)
Supported by the Scientific Research Special Project of Shanghai Landscaping & City Appearance Administrative Bureau (G212406; G242414; G242416) and Science and Technology Talent Development Fund for Young and Middle-aged Teachers of Shanghai Institute of Technology (ZQ2022-17)
* 共同通讯作者 Co-authors for correspondence. E-mail: ygsong@cemps.ac.cn; xujin@sit.edu.cn

reconstruction due to its structural conservation, maternal uniparental inheritance, and high copy number. However, the chloroplast genome cannot effectively reflect pollen-mediated gene flow, is sensitive to genetic drift, and is susceptible to incomplete lineage sorting and hybridization events. Therefore, in plant identification and population evolutionary history studies, relying solely on chloroplast genome evidence may lead to biased or incomplete conclusions.
Methods: This study performed whole-chloroplast genome sequencing, assembly, and annotation for seven *Pterocarya* individuals, and conducted comparative genomic and phylogenetic analyses incorporating published data.
Results: The chloroplast genome ranged in length from 160,176 to 160,318 bp and exhibited a typical quadripartite structure, encoding 131 genes (86 protein-coding genes, 8 rRNAs, and 37 tRNAs). The number of simple sequence repeats (SSRs) loci ranged from 87 to 96, dominated by mononucleotide repeats. Comparative genomic analysis indicated that chloroplast genomes among *Pterocarya* species were highly conserved overall. The sequence conservation of the inverted repeat regions and coding regions was higher than that of the single-copy regions and non-coding regions. Notably, within the coding regions, the *ndhF* gene exhibited significantly higher sequence variation than other protein-coding genes. The results of phylogenetic analysis showed that the phylogenetic trees constructed based on different data partitions (whole genome, protein-coding sequence, hypervariable regions, and sliding windows) exhibited various topological structures with two to five major clades. Individuals of species such as *P. stenoptera*, *P. macroptera* var. *macroptera*, and *P. insignis* consistently showed lineage admixture and cross-clade distribution across different analyses.
Conclusion: The chloroplast genome exhibits significant limitations in resolving species identification and biogeographic history within the genus *Pterocarya*, and relying solely on its data may lead to systematic bias. Therefore, future studies on plant phylogenetics should integrate nuclear genomic data and employ multispecies coalescent methods to achieve more accurate species identification and evolutionary relationship reconstruction.
Key words: chloroplast genome; *Pterocarya*; phylogenetic analysis; biogeography

叶绿体基因组凭借其结构高度保守、通常呈母系单亲遗传、细胞内拷贝数极高且易于获取等独特优势, 无可争议地成为重建植物生命之树, 尤其是解决属及以上阶元系统发育关系的经典工具(Zhang et al., 2022)。在物种鉴定领域, 叶绿体基因组中的多个质体区域(如*trnH-psbA*、*matK*、*rbcL*等)已被确立为核心植物DNA条形码, 为物种, 特别是近缘种和隐存种的快速、准确鉴别提供了关键分子证据(Chase et al., 2007; Kress & Erickson, 2007; Zhang et al., 2012)。基于叶绿体基因组的系统发育与谱系地理分析, 为追溯类群起源中心、推断大陆扩散路线等宏观进化历史研究提供了关键证据(Chen et al., 2025; Li PW et al., 2025)。

随着植物系统学与生物地理学研究焦点逐渐从宏观历史格局下移至物种内部的种群动态与近期分化过程(Cao et al., 2022; Dong et al., 2024), 叶绿体基因组的应用场景也发生了深刻变化。在过去二三十年间, 利用叶绿体片段或完整基因组开展种群水平的遗传学研究已成为一种广泛范式(Avise, 2000; Yan et al., 2023)。这类研究试图通过母系遗传的叶绿体谱系来描绘种群的遗传结构、历史动态和基因流模式(刘家奇等, 2021; 张健和张宏祥, 2024; Liu et al., 2025)。这种研究范式的广泛应用在本质上预设了一个关键前提: 即适用于解析长期、深度分化历史的叶绿体基因组, 其进化历史同样能够准确反映近期分化且可能存在复杂基因流的种群整体历史。

然而, 将叶绿体基因组谱系简单等同于种群历史的研究方法正面临日益增多的理论质疑与实证挑战(Wu et al., 2014; Arredondo et al., 2021)。叶绿体基因组因其固有的生物学特性, 在揭示物种进化历史方面存在本质性局限: 首先, 在被子植物中叶绿体基因组大多呈现母系遗传特性, 其基因流仅能通过种子传播实现, 无法捕获由花粉介导的父系基因流(胡颖等, 2019; Xu et al., 2020), 导致基于叶绿体基因组推断的基因流格局和地理隔离模式存在严重偏差, 往往显著低估种群间的实际遗传连通性(Petit et al., 2005)。其次, 叶绿体基因组的单倍体与非重组性质使其有效种群大小(effective population size, Ne)理论上仅为核基因组的1/4, 这一特征使叶绿体基因组更容易受遗传漂变影响, 因此其显示出的种群结构可能只是遗传漂变造成的假象, 而非真实的进化历史(付涛等, 2018; 胡颖等, 2019)。此外, 在近期经历快速分化的类群中, 叶绿体基因组与核基因组谱系之间常因不完全谱系分选(incomplete lineage sorting, ILS)而出现不一致(O'Donnell et al.,

2025; Shen et al., 2025), 这不仅会影响系统发育关系的重建, 也会增加单纯依赖叶绿体基因组进行物种鉴定的误判风险(Wu et al., 2025)。而叶绿体基因组通过杂交导致的跨物种水平转移(即“细胞质捕获”)更可能进一步引入误差, 从而得出误导性的系统发育关系与生物地理推论(王茜等, 2019; Shen et al., 2025)。

枫杨属(*Pterocarya*)是胡桃科中阐释上述叶绿体基因组局限性与推断偏差的经典范例。该属包含8个分类群, 其现代地理分布呈现出典型的东亚至地中海东缘的间断性格局(匡可任等, 1979)。其中, 东亚地区有枫杨(*P. stenoptera*)、越南枫杨(*P. tonkinensis*)、甘肃枫杨(*P. macroptera* var. *macroptera*)、华西枫杨(*P. macroptera* var. *insignis*)、云南枫杨(*P. macroptera* var. *delavayi*)、湖北枫杨(*P. hupehensis*)和水胡桃(*P. rhoifolia*), 而地中海东缘(以高加索地区为核心)仅有高加索枫杨(*P. fraxinifolia*) 1个物种(Kozlowski et al., 2018)。化石记录表明枫杨属在新生代曾经广泛分布在北半球, 但随着气候变冷(尤其是第四纪末次冰盛期的到来), 枫杨属在北美、西伯利亚和欧洲大陆地区陆续消失, 最终形成了当前这一新生代孑遗植物属的间断性分布格局(Song et al., 2020, 2021)。早期研究多依赖少数叶绿体片段(如*trnH-psbA*), 导致部分节点支持率较低(Song et al., 2020)。近期研究通过整合化石校准与转录组数据, 将枫杨属的起源时间推至始新世早期, 并揭示其现存东亚–地中海间断分布格局与渐新世气候变冷和青藏高原隆升相关(Yan et al., 2024)。近年来, 随着枫杨、湖北枫杨、甘肃枫杨、华西枫杨、云南枫杨、越南枫杨以及高加索枫杨的叶绿体基因组相继发表(Mu et al., 2020; Yan et al., 2021; Zhou et al., 2021), 该属的系统发育研究取得了新进展。然而枫杨属是一类典型的风媒异花授粉植物(Lu et al., 2024), 这意味着单独使用叶绿体基因组构建的系统发育树可能无法准确揭示其演化历史与分布格局的成因。

本研究基于高通量测序技术, 对采自不同种群的水胡桃、甘肃枫杨、华西枫杨和云南枫杨的多个个体进行了叶绿体全基因组测序, 旨在通过获取枫杨属代表性物种的完整叶绿体基因组, 达成以下两个目的: 首先, 在方法学层面, 系统评估叶绿体基因组数据在枫杨属物种水平上的系统发育分辨率及其应用局限; 其次, 在进化生物学层面, 基于所获数据深入探讨可能造成其谱系历史复杂性的关键进化机制。具体而言, 本研究将依次解决以下3个关键科学问题: (1)解析枫杨属叶绿体基因组的结构与序列变异特征; (2)揭示基于叶绿体基因组的枫杨属系统发育拓扑结构, 并评估其与现有分类框架和核基因系统发育的异同; (3)探讨上述发现对理解枫杨属物种分化模式与演化历史复杂性的启示。

1 材料与方法

1.1 植物样品采集、DNA提取与测序

分别从野外自然分布区域内采集水胡桃、甘肃枫杨、华西枫杨和云南枫杨的个体标本与叶片材料(表1)。在每个个体样本上采集完整的新鲜成熟叶片, 用硅胶迅速干燥后带回实验室储存备用, 凭证标本存放于上海辰山植物园标本馆(CSH)。

利用改良的CTAB法(Doyle & Doyle, 1987)提取植物叶片的总DNA, 检测合格后进行机械打断和纯化处理。将纯化后的DNA片段进行末端修饰、接头连接、连接产物片段分选纯化和PCR扩增。然后, 利用全基因组鸟枪法(Pachter et al., 1999)构建平均长度为350 bp的DNA文库。最终, 在武汉贝纳科技有限公司使用Illumina NovaSeq6000 (Illumina, San Diego)平台进行定量DNA的双端测序, reads长度为150 bp, 样本数据量为5 GB。

1.2 叶绿体基因组组装、注释与获取

高通量测序得到的原始图像数据文件经碱基识别(base calling)分析(Ewing & Green, 1998)转化为原始数据(raw data), 使用过滤软件SOAPnuke Toolkit v.1.3.0进行低质量数据过滤以获得有效数据(clean data) (Chen et al., 2018)。基于有效数据, 使用GetOrganelle v.1.7.6.1软件进行叶绿体基因组组装(Jin et al., 2020)。使用美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)的BLAST+ v.2.13.0在线软件(https://blast.ncbi.nlm.nih.gov/Blast.cgi)比对组装序列, 获得相似度最高的基因组序列。以此为参考基因组, 利用专门针对叶绿体基因组的在线注释软件CPGAVAS2 (http://47.96.249.172:16019/analyzer/annotate)进行叶

表1 枫杨属植物材料野外采样的基本信息
Table 1 Basic information on field sampling of plant materials of *Pterocarya*

物种 Species	采集地 Location	缩写 Abbreviation	纬度 Latitude	经度 Longitude	海拔 Altitude (m)	标本号 Specimen no.	登录号 GenBank no.	来源 Source
水胡桃 *P. rhoifolia*	Niigata University Forest, Sado City, Niigata, Japan	PR_NS	38.23° N	138.40° E	370	NUF3	PV130126	自测 This study
甘肃枫杨 *P. macroptera* var. *macroptera*	甘肃省天水市党川乡 Dangchuan Township, Tianshui City, Gansu Province, China	PMM_TS	34.27° N	106.06° E	1,503	DM15793-2	PV130121	自测 This study
甘肃枫杨 *P. macroptera* var. *macroptera*	四川省冕宁县森荣乡 Senrong Township, Mianning County, Sichuan Province, China	PMM_MN	28.35° N	101.98° E	2,620	DM17327	PV130122	自测 This study
甘肃枫杨 *P. macroptera* var. *macroptera*	四川省峨眉山景区 Emei Mountain Scenic Area, Sichuan Province, China	PMM_EM	29.54° N	103.34° E	2,055	LXC00013	PV130125	自测 This study
华西枫杨 *P. macroptera* var. *insignis*	浙江省西天目山 West Tianmu Mountain, Zhejiang Province, China	PMI_TM	28.35° N	118.87° E	950	SYG00006-3	PV130127	自测 This study
云南枫杨 *P. macroptera* var. *delavayi*	云南省迪庆藏族自治州维西傈僳族自治县 Weixi Lisu Autonomous County, Diqing Tibetan Autonomous Prefecture, Yunnan Province, China	PMD_DQ	27.76° N	99.02° E	2,240	DM21298	PV130123	自测 This study
云南枫杨 *P. macroptera* var. *delavayi*	云南省怒江傈僳族自治州贡山独龙族怒族自治县 Gongshan Derung and Nu Autonomous County, Nujiang Lisu Autonomous Prefecture, Yunnan Province, China	PMD_NJ	28.07° N	98.74° E	2,899	DM21322	PV130124	自测 This study

表格中的“缩写”由物种和地名缩写组成, 例如, PR_NS中的“PR”代表物种, “NS”代表地名。
Abbreviations in the table consist of species and place name. For example, in PR_NS, ‘PR’ represents the species and ‘NS’ represents a place name.

绿体基因注释(Shi et al., 2019)。随后, 使用OrganellarGenomeDRAW (OGDRAW) v1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html)绘制叶绿体基因组谱图(Lohse et al., 2013)。组装注释好的7个叶绿体基因组序列信息已上传至NCBI数据库, GenBank登录号为PV130121–PV130127。

1.3 重复序列与SSRs分析

从在线软件MIcroSAtellite (MISA, https://webblast.ipk-gatersleben.de/misa/) (Beier et al., 2017)下载其本地运行脚本, 并在适当修改后对17个叶绿体基因组中的简单重复序列(simple sequence repeats, SSRs)进行批量分析。MISA配置文件中的参数设置为: 单核苷酸重复不少于10次, 二核苷酸不少于5次, 三核苷酸不少于4次, 四、五和六核苷酸重复均不少于3次(即1-10、2-5、3-4、4-3、5-3和6-3) (Zhao et al., 2019)。

利用在线工具Tandem Repeats Finder (TRF, https://tandem.bu.edu/trf/trf.html) (Benson, 1999)检测17个枫杨属叶绿体基因组中长度不少于10 bp的串联重复序列(tandem repeat sequences)。Match、Mismatch和Indel 3个对齐参数分别设置为2、7、7, 最低对齐分数和最大周期分别设置为80和500。

利用Vmatch v2.3.1软件(http://www.vmatch.de/)预测叶绿体基因组的散在重复序列(dispersed repeat sequences), 包括正向重复序列(direct repeat sequences, D)和回文重复序列(palindromic repeat sequences, P) (Kurtz, 2003)。在分析过程中, 最小重复长度阈值设定为30 bp (-l 30), 允许的最大汉明距离为3 (-h 3) (Liang et al., 2019)。此外, 保留每个重复序列的最佳匹配结果(-best 50), 并输出其在基因组中的绝对位置信息(-absolute), 以便后续统计与可视化分析。

1.4 叶绿体基因组结构比较与序列差异分析

从NCBI数据库下载10个已发表的枫杨属物种个体的叶绿体基因组序列, 包括枫杨(MN262640.1)、湖北枫杨(MH188293.1)、高加索枫杨(MH188291.1)和越南枫杨(MH188288.1) 4个物种(附录1)。结合本研究测序并组装的4个物种(水胡

桃、甘肃枫杨、华西枫杨和云南枫杨), 共8个分类群, 对整个枫杨属进行比较基因组学分析, 探讨其物种间叶绿体基因组的变异关系。

使用在线工具IRscope (https://irscope.shinyapps.io/irapp/) (Amiryousefi et al., 2018)可视化叶绿体基因组连接位点边界上的基因, 分析4个边界区域IRa (反向重复区a, inverted repeat a)/LSC (大单拷贝区, large single copy) (JLA)、LSC/IRb (反向重复区b, inverted repeat b) (JLB)、IRb/SSC (小单拷贝区, small single copy) (JSB)和SSC/IRa (JSA)的基因分布, 观察IR的扩张与收缩情况, 比较叶绿体基因组结构的差异性。

利用比较基因组学工具mVISTA (https://genome.lbl.gov/vista/mvista/submit.shtml) (Frazer et al., 2004), 以枫杨的叶绿体基因组为参考序列, 采用Shuffle-LAGAN模型进行对齐, 与其余几个物种的叶绿体全基因组序列进行可视化比对分析, 评估各基因组序列之间的整体相似性和差异性。

通过MAFFT v7.487软件(Katoh & Standley, 2013)对所有物种的叶绿体全基因组序列进行多序列比对, 然后使用DnaSP v.6.12.03软件(Rozas et al., 2017)计算其单核苷酸多态性(single nucleotide polymorphism, SNP)和插入/删除(insertion/deletion, indel)位点。再进一步进行滑动窗口分析(窗口长度为800 bp, 步长为200 bp), 计算叶绿体基因组之间的核苷酸多样性(nucleotide diversity, Pi), 综合研究物种之间的高变异热点区域(Gou et al., 2020)。

1.5 系统发育分析

为探究枫杨属植物的系统亲缘关系, 本研究整合了所有已公布的枫杨属叶绿体基因组(附录1)及新测序数据。基于4类数据集——全基因组、蛋白质编码序列(coding sequence, CDS)、高变区及滑动窗口分析数据——采用最大似然法(maximum likelihood method, ML)构建系统发育树。使用马尾树(*Rhoiptelea chiliantha*)和杨梅(*Morella rubra*)作为外类群为该系统发育树定根。CDS数据集通过自定义Python脚本(基于Biopython库)构建, 该脚本从GenBank格式文件中提取所有CDS特征的核苷酸序列并记录基因名称, 确保数据提取的标准化。利用IQtree v.2.1.3软件构建系统发育树, 碱基替换模型选择K3Pu + F + I + G4, bootstrap抽样重复1,000次(Nguyen et al., 2015)。最后用FigTree v.1.4.3软件对系统发育树进行可视化绘图。

2 结果

2.1 枫杨属叶绿体基因组基本特征

对高通量测序得到的序列进行低质量数据过滤后获得有效数据, 然后对每个叶绿体基因组进行组装注释, 获得水胡桃(PR_NS)、甘肃枫杨(PMM_TS、PMM_MN、PMM_EM)、华西枫杨(PMI_TM)和云南枫杨(PMD_DQ、PMD_NJ) 7个个体的叶绿体全基因组。与典型被子植物叶绿体基因组相似, 这7个个体的叶绿体基因组均为闭合环状的四分体结构, 由1个LSC、1个SSC和2个IR (IRa和IRb)构成(图1)。枫杨属这7个个体的叶绿体基因组高度保守, 总长度相差较小, 介于160,176 bp (PMD_DQ)到160,318 bp (PMM_EM)之间, 仅相差142 bp。其中LSC长度在89,701–89,810 bp之间; SSC长度在18,412–18,459 bp之间; IR长度在26,014–26,031 bp之间(表2)。除PMM_EM叶绿体基因组的GC含量为36.10%外, 其他个体的叶绿体基因组GC含量均为36.20%。其IR的GC含量在42.54%–42.57%之间, 均显著高于LSC (最高为33.79%)和SSC (最高为29.89%) (表2)。

本研究中, 这7个个体的叶绿体基因组注释到的基因总数均为131个, 包括86个蛋白质编码基因(protein-coding genes, PCGs)、37个转运RNA基因(tRNAs)和8个核糖体RNA基因(rRNAs) (图1, 附录2)。其中有20个基因存在两个拷贝, 包括8个蛋白质编码基因(*rps*7、*rps*8、*rps*12、*ndhB*、*rpl*2、*rpl*23、*ycf*1和*ycf*2)、8个tRNA基因(*trnA-UGC*、*trnI-CAU*、*trnI-GAU*、*trnL-CAA*、*trnN-GUU*、*trnR-ACG*、*trnS-GCU*和*trnV-GAC*)和4个rRNA基因(*rrn*4.5*S*、*rrn*5*S*、*rrn*16*S*和*rrn*23*S*) (附录2)。此外, 有17个基因含有内含子, 其中14个只含有1个内含子, 包括8个蛋白质编码基因(*rps*16、*rpl*2、*rpl*16、*atpF*、*petB*、*petD*、*ndhA*和*ndhB*)和6个tRNA基因(*trnA-UGC*、*trnG-GCC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA*和*trnV-UAC*); 而其余3个含有2个内含子(*rps*12、*clpP*和*ycf*3)。

2.2 重复序列

利用MISA软件在这17个叶绿体基因组中检测

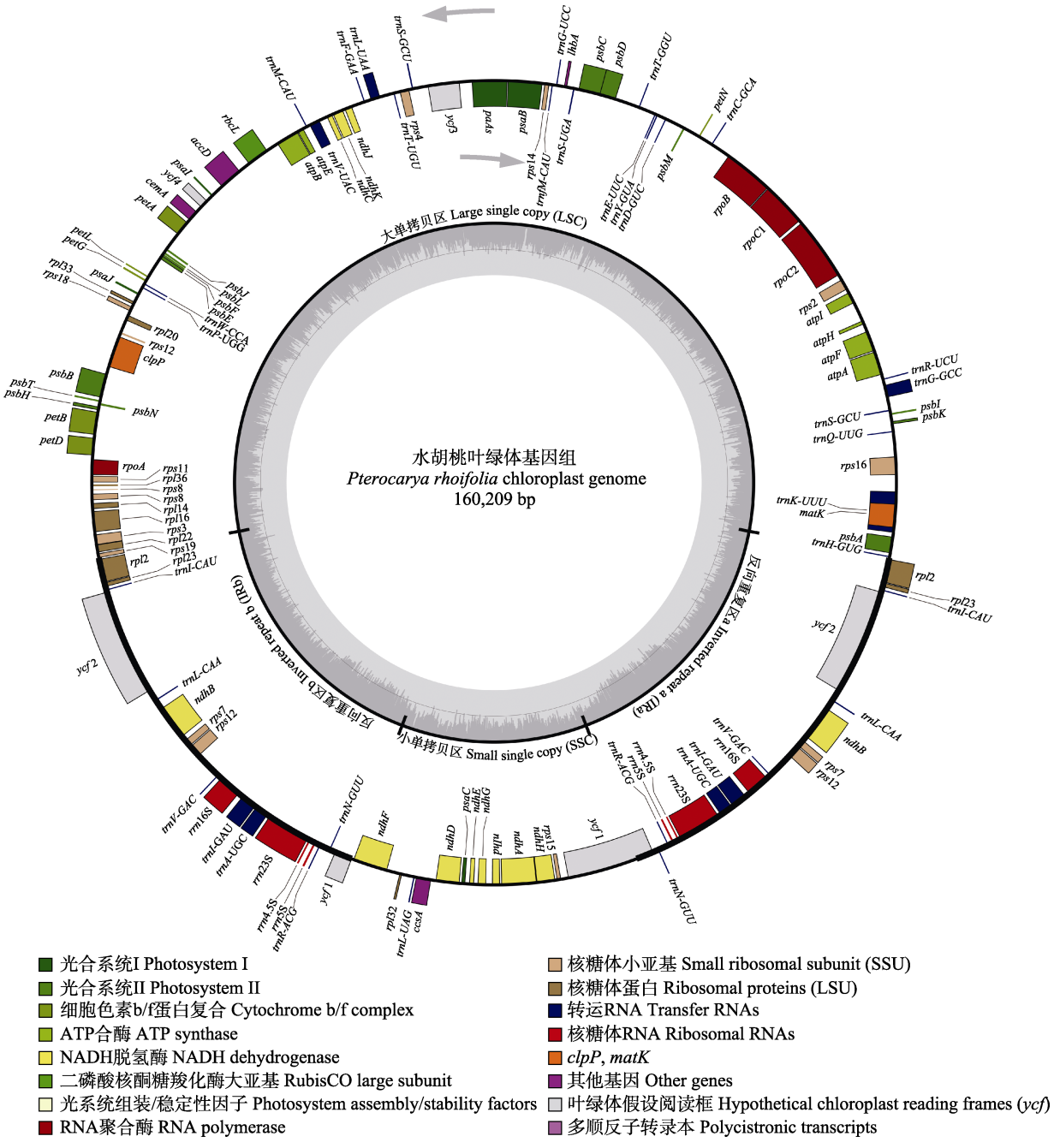

图1　以水胡桃为例展示枫杨属植物叶绿体基因组的基因组图谱(其他6个个体的叶绿体基因组图谱相似)。外圈显示质体上的基因结构。圆圈外的基因是逆时针转录，而圆圈内的基因是顺时针转录。基因功能分类由不同颜色区分。

Fig. 1　Genomic map of the *Pterocarya rhoifolia* chloroplast genome (the chloroplast genome maps of other six individuals are similar). The outer circle shows the gene structure on the plastome. Genes outside the circle are transcribed counterclockwise, while genes inside the circle are transcribed clockwise. Genes are color-coded according to their functional categories.

到87–96个SSRs位点，其中枫杨和华西枫杨的SSRs位点数最多(均为96个)，而湖北枫杨的SSRs位点数最少(87个)。在所有的SSRs位点中，均包括单核苷酸(数量最多)、二核苷酸、三核苷酸、四核苷酸类型，而五核苷酸和六核苷酸类型仅在个别物种中出现(图2A)。SSRs绝大部分由腺嘌呤(A)和胸腺嘧啶(T)两种碱基构成，具有较强的A/T碱基偏向性。SSRs位点主要分布于LSC（占总数的

表2　本研究新测序的7个个体的叶绿体基因组特征比较

Table 2　Comparison of chloroplast genomic characteristics of seven newly sequenced individuals in this study

基因组特征 Genome characteristics	水胡桃 PR_NS	甘肃枫杨 PMM_TS	甘肃枫杨 PMM_MN	甘肃枫杨 PMM_EM	华西枫杨 PMI_TM	云南枫杨 PMD_DQ	云南枫杨 PMD_NJ
序列全长 Genome size (bp)	160,209	160,210	160,238	160,318	160,236	160,176	160,232
大单拷贝区长度 Large single copy (LSC) length (bp)	89,708	89,731	89,770	89,810	89,768	89,701	89,757
小单拷贝区长度 Small single copy (SSC) length (bp)	18,459	18,451	18,412	18,446	18,412	18,443	18,443
反向重复区长度 Inverted repeat (IR) length (bp)	26,021	26,014	26,028	26,031	26,028	26,016	26,016
GC含量 GC content (%)	36.20	36.20	36.20	36.10	36.20	36.20	36.20
大单拷贝区GC含量 GC content of LSC (%)	33.77	33.77	33.78	33.72	33.72	33.77	33.79
小单拷贝区GC含量 GC content of SSC (%)	29.86	29.86	29.86	29.85	29.85	29.89	29.87
反向重复区GC含量 GC content of IR (%)	42.57	42.57	42.56	42.57	42.57	42.54	42.55
A碱基占比 A base content (%)	31.60	31.50	31.60	31.60	31.60	31.50	31.60
T碱基占比 T base content (%)	32.30	32.30	32.30	32.30	32.30	32.30	32.30
C碱基占比 C base content (%)	18.40	18.40	18.40	18.40	18.40	18.40	18.40
G碱基占比 G base content (%)	17.80	17.80	17.80	17.80	17.80	17.80	17.80
基因总数 No. of total genes	131	131	131	131	131	131	131
蛋白质编码基因数 No. of protein coding genes	86	86	86	86	86	86	86
tRNA基因数 No. of tRNA genes	37	37	37	37	37	37	37
rRNA基因数 No. of rRNA genes	8	8	8	8	8	8	8

物种缩写含义见表1。The abbreviations of species see Table 1.

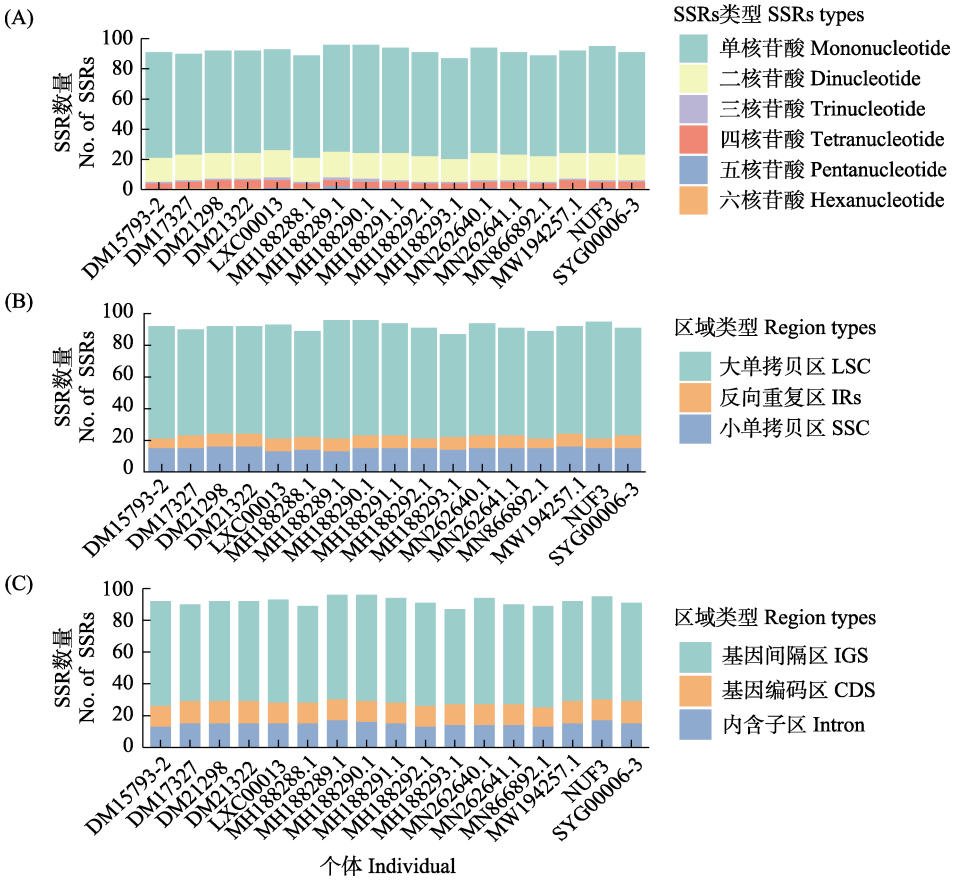

图2　17个个体的叶绿体基因组简单重复序列(SSRs)分析。(A)不同SSRs类型的数量; (B)大单拷贝区(LSC)、反向重复区(IRs)和小单拷贝区(SSC)的SSRs数量; (C)基因间隔区(IGS)、基因编码区(CDS)及内含子区的SSRs数量。

Fig. 2　Simple sequence repeat (SSRs) analysis of the chloroplast genomes of 17 individuals. (A) Number of SSRs in different types; (B) Number of SSRs in the large single copy (LSC), inverted repeats (IRs) and small single copy (SSC); (C) Number of SSRs in the intergenic spacer (IGS), coding sequence (CDS) and intron regions.

76%), 其次为SSC (占16%), IR最少(占8%) (图2B)。另外, SSRs位点大部分分布于叶绿体基因组的非编码区上, 且基因间隔区(intergenic spacer, IGS)分布最多(占70%), 其次为内含子(占16%), 而在基因编码区分布最少(占14%) (图2C)。

通过Tandem Repeats Finder软件鉴定的串联重复序列在6–13个(图3A), 长度在5–47 bp之间, 主要集中在20–30 bp (图3B)。17个个体的叶绿体基因组的串联重复序列绝大多数都分布于基因间隔区, 占总数的70%以上, 而其余串联重复序列都分布在*ycf*2基因区中。

利用Vmatch v2.3.1软件鉴定出的散在重复序列在32–47个, 其中回文重复数量最多(17–27个), 正向重复其次(13–20个) (图3A)。散在重复序列长度在30–54 bp之间, 且主要分布于30–45 bp (图3C)。

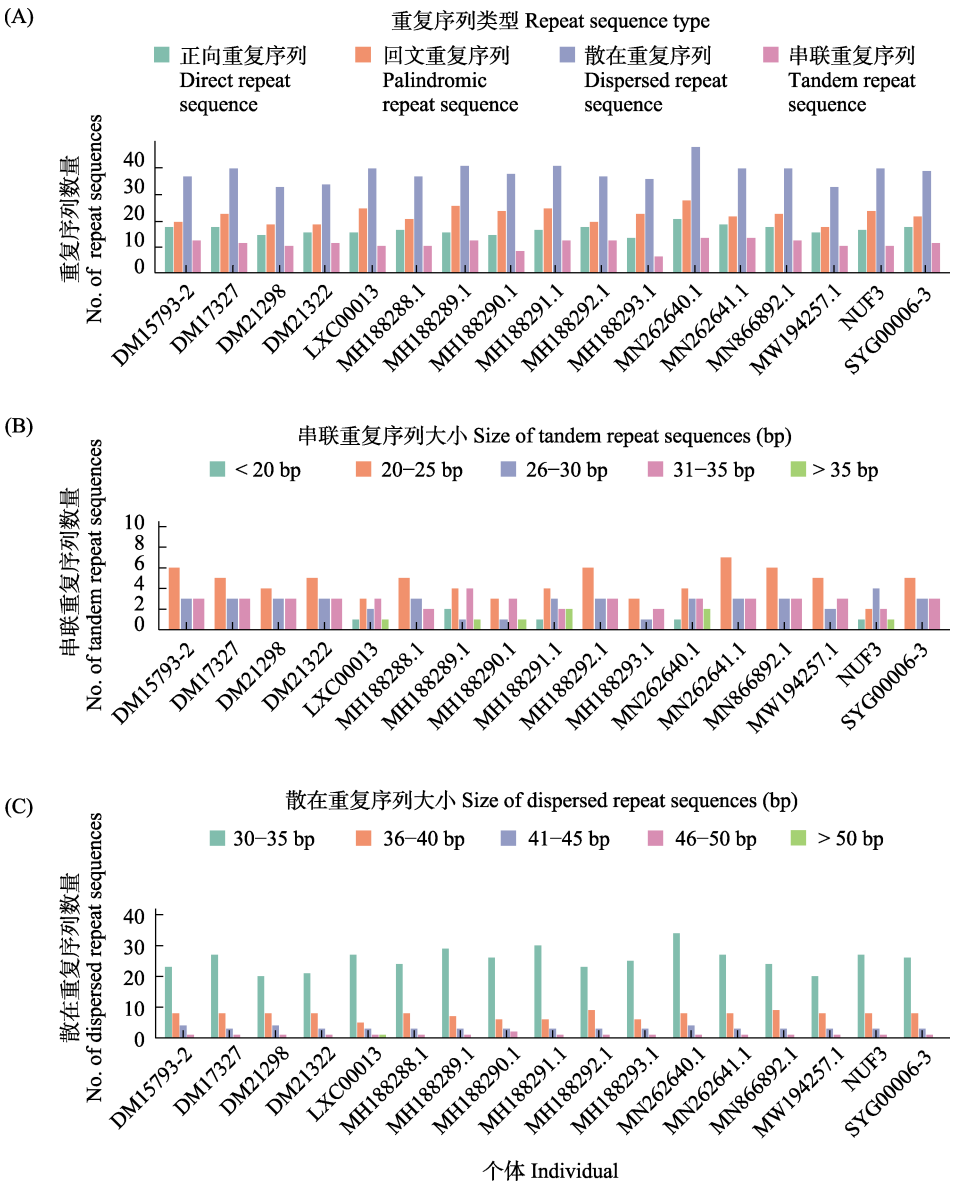

图3　17个个体的叶绿体基因组重复序列分析(包括串联重复和散在重复)。(A)不同重复序列类型的分布数量; (B)不同长度分布范围内串联重复序列的数量; (C)不同长度分布范围内散在重复序列的数量。

Fig. 3　Analysis of repeat sequences in the chloroplast genomes of 17 individuals, including tandem and dispersed repeats. (A) Distribution and counts of different repeat sequence types; (B) Counts of tandem repeats across different length ranges; (C) Counts of dispersed repeats across different length ranges.

2.3 枫杨属叶绿体基因组的结构比较与序列差异

2.3.1 IR的扩张与收缩

使用在线软件IRscope比较枫杨属8个分类群的叶绿体基因组IR的扩张与收缩情况(图4)。在8个分类群的叶绿体基因组中, 分布于各边界附近的基因基本一致, 但距边界的距离存在差异。其中*rps*19基因位于JLB附近, 在水胡桃、甘肃枫杨、华西枫杨和云南枫杨中距边界均收缩1 bp, 在枫杨中收缩7 bp, 而在高加索枫杨、湖北枫杨和越南枫杨中距边界扩张了3 bp。*trnH*基因在JLA附近, 除甘肃枫杨距边界的距离为47 bp以外, 其余物种均为41 bp。*ndhF*基因在枫杨、高加索枫杨和越南枫杨的JSB附近被检测到; 而湖北枫杨、云南枫杨、华西枫杨、甘肃枫杨及水胡桃未在JSB附近检测到*ndhF*基因。值得注意的是, 在越南枫杨中, *ndhF*基因跨越了IRb和SSC两个区域。高等植物叶绿体基因组中的重复基因*ycf*1横跨SSC和IRa、IRb (图1)。在枫杨属中, *ycf*1基因位于IRa和IRb的重复序列长度在1,137–1,443 bp之间, 但在JSB附近的*ycf*1基因剩余3–27 bp, 而在JSA附近的*ycf*1基因剩余4,206–4,512 bp。

2.3.2 叶绿体基因组序列差异可视化

选取枫杨的叶绿体基因组作为参考序列, 利用mVISTA对枫杨属8个分类群的叶绿体全基因组序列进行可视化比较分析(图5)。结果表明, 枫杨属植物的叶绿体基因组拥有高度的序列相似性, 具有高度保守性。非编码区相比于编码区具有更高的分化变异水平; 单拷贝区的分化变异水平显著高于反向重复区。4个rRNA基因最为保守, 几乎没有发生变异。编码区中*ndhF*基因发生了较大变异, 而非编码区变异较大的主要有*ndhC–trnV-UAC*和*petA–psbJ*等基因间隔区。这些区域具有高度可变性, 为枫杨属的遗传进化分析提供了有用的数据信息。

2.3.3 叶绿体基因组高变区

利用软件DnaSP计算枫杨属8个分类群叶绿体基因组序列之间的核苷酸多态性并进行统计分析

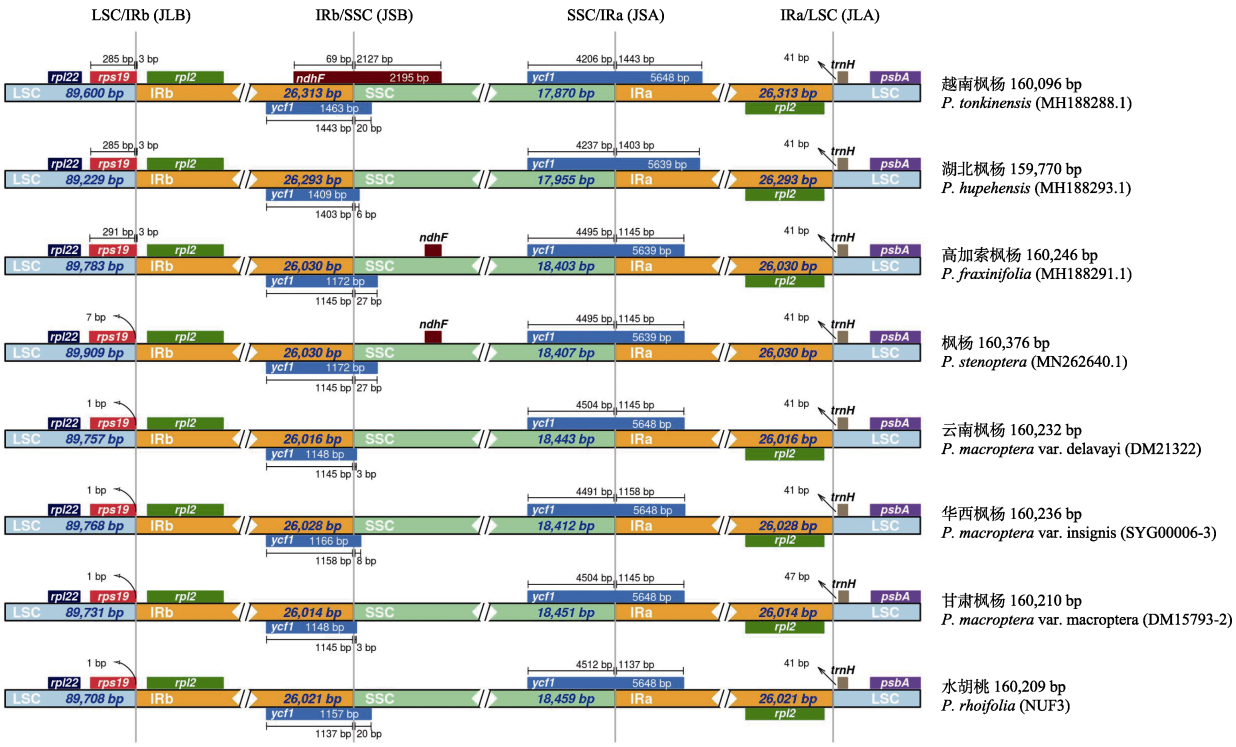

图4 枫杨属叶绿体基因组反向重复区(IR)边界的扩张与收缩分析: IRa/LSC (JLA)、LSC/IRb (JLB)、IRb/SSC (JSB)和SSC/IRa (JSA)。LSC: 大单拷贝区; SSC: 小单拷贝区; IRa: 反向重复区a; IRb: 反向重复区b。

Fig. 4 Analysis of the contraction and expansion of the inverted repeat (IR) region boundaries in the chloroplast genomes of *Pterocarya* species: IRa/LSC (JLA), LSC/IRb (JLB), IRb/SSC (JSB), SSC/IRa (JSA). LSC, Large single copy; SSC, Small single copy; IRa, Inverted repeat a; IRb, Inverted repeat b.

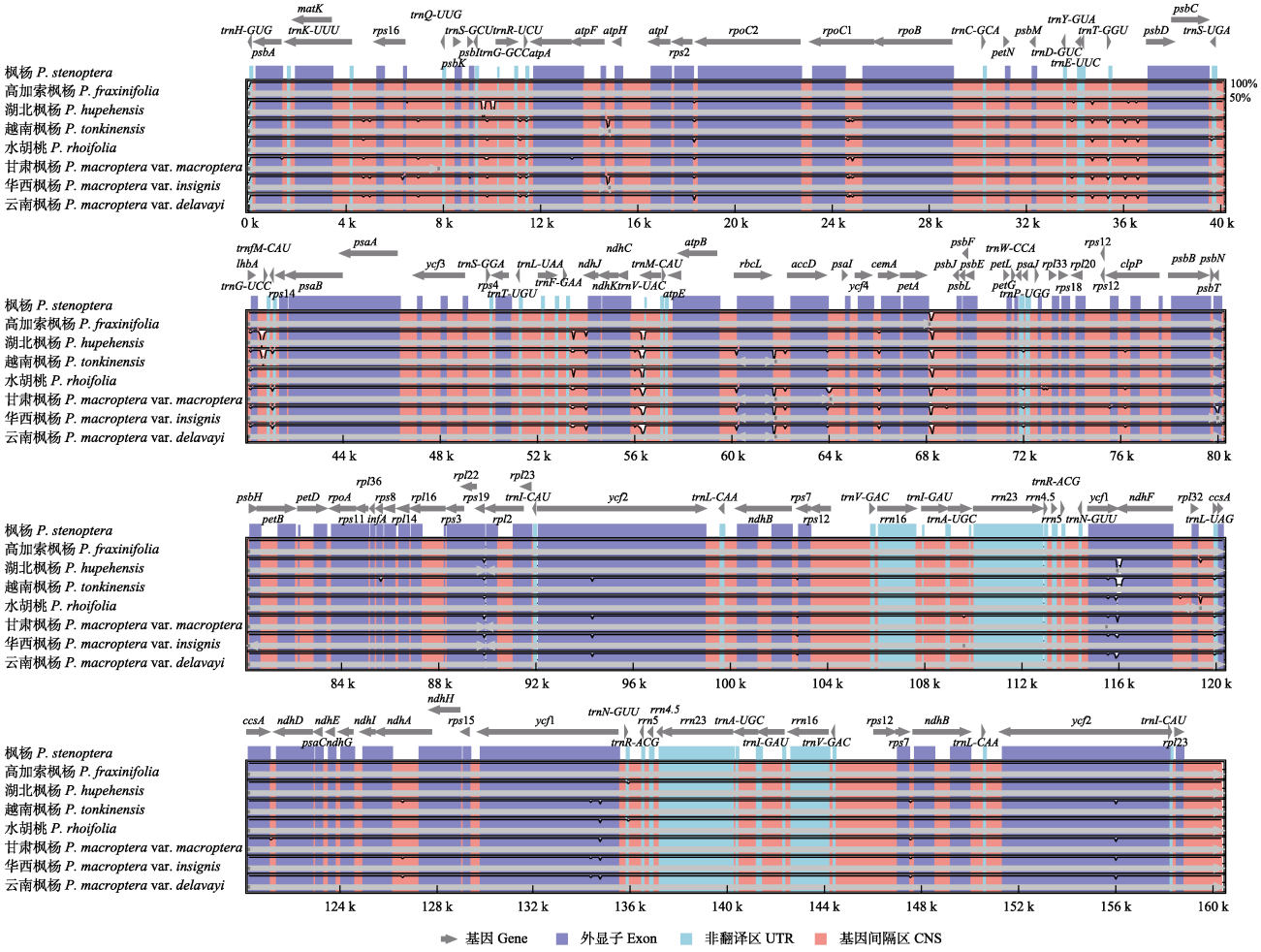

图5　以枫杨(*Pterocarya stenoptera*)为参考序列的枫杨属(*Pterocarya*) 8个分类群的叶绿体基因组序列差异可视化比对。上面的灰色箭头显示了参考序列基因的位置，方向为正向或反向。对齐相似度百分比显示在图的右侧(垂直轴)。

Fig. 5　Visualization of the eight chloroplast genomes of *Pterocarya* with *Pterocarya stenoptera* as a reference using mVISTA. The gray arrows above show the position of the reference sequence genes, in either the forward or reverse direction. Alignment similarity percentages are shown on the right side of the figure (vertical axis).

发现，全基因组序列中共检测到467个变异多态性位点(variable polymorphic sites)，包括202个单核苷酸多态性位点和265个简约信息位点(parsimony informative sites)，还检测到120个indels事件。基因组IR的核苷酸多态性显著低于LSC和SSC。对叶绿体基因组序列进一步进行滑动窗口分析，发现枫杨属叶绿体基因组核苷酸多样性值(Pi值)为0–0.04196，平均值为0.00120。通过将基因的分布位置与滑动窗口分析的结果进行匹配，筛选出一个具有较大Pi值的蛋白质编码基因，为位于SSC区域的*ndhF*基因(Pi > 0.01) (图6，附录3)。

2.4　叶绿体基因组系统发育分析

基于枫杨属17个个体的叶绿体全基因组、CDS、高变区(*ndhF*)及10 kb滑动窗口的系统发育分析表明，各分支节点均具有较高的分辨率(图7，附录4)。这些系统发育树主要呈现4种拓扑结构: (1)叶绿体全基因组、CDS及14个滑动窗口(Window_2–4、6–12、14–17)的分析结果将全部个体划分为两个明显支系; (2)高变区(*ndhF*)和Window_13的分析结果呈现3个支系; (3)滑动窗口Window_5的分析结果显示出4个支系; (4)滑动窗口Window_1的分析结果则形成5个支系。在系统发育关系中，云南枫杨的3个个体在多数分析中(除CDS、*ndhF*及Window_2、3、6、8、11、14、16外)能稳定聚为一个单系群。相比之下，采自不同种群的3个枫杨、3个甘肃枫杨和4个华西枫杨个体都没有形成一个单系，同一物种的

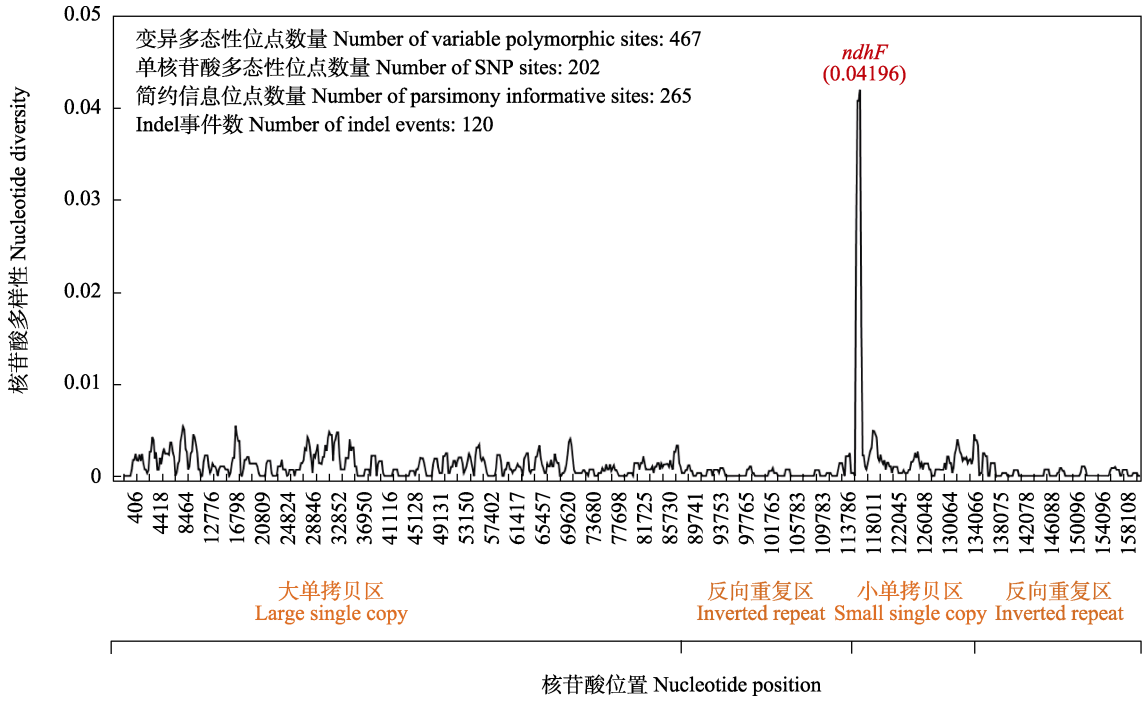

图6　枫杨属叶绿体基因组的滑动窗口分析。X轴表示核苷酸位置, Y轴表示每个窗口的核苷酸多样性(Pi)值。SNP: 单核苷酸多态性。

Fig. 6　Sliding window analysis of chloroplast genome of *Pterocarya*. The X-axis represents the nucleotide positions and the Y-axis represents the value of nucleotide diversity (Pi) per window. SNP, Single nucleotide polymorphism.

不同个体分散在不同的支系中。对于仅有1个个体的物种, 湖北枫杨和水胡桃在大多数分析中位于上部支系(除Window_5中水胡桃单独成支, 以及滑动窗口Window_3和Window_17中位于下部支系外), 而越南枫杨则稳定地位于下部支系。

3　讨论

3.1　枫杨属及胡桃科的叶绿体基因组结构变异

本研究对枫杨属的叶绿体基因组进行了测序、注释及比较分析。结果显示, 枫杨属的叶绿体基因组与大多数陆生植物的叶绿体基因组一样, 呈现出高度的保守性, 这体现在其大小、四分体结构以及GC含量的稳定性上(Ruhlman & Jansen, 2021)。枫杨属在基因总数和独特基因数上与胡桃科其他属存在细微差异(Mo et al., 2020; Luo et al., 2022)。综合多篇文献的注释结果, 枫杨属表现出以下可能特征: 与近缘属相比, 其多次被注释出含有一个额外的*rps*8基因(Hu et al., 2017b; Geng et al., 2019; Luo et al., 2022), 而缺失*infA*和*ycf*15基因(Yan et al., 2017; Huang et al., 2024; 李廷章等, 2024)。在tRNA基因方面, 枫杨属均被注释为缺少*trnS-GGA*但拥有*trnG-UCC*, 此模式与化香树属(*Platycarya*)、山核桃属(*Carya*)及胡桃属(*Juglans*)的注释结果相异(Hu et al., 2017a; Yan et al., 2017; Mo et al., 2020)。这些基于已有注释的发现为枫杨属的独特性提供了线索, 但其最终确认需依赖未来更一致的数据分析方法。

根据多种叶绿体基因组分析结果, 重复序列在诱导插入和替换过程中扮演着不可或缺的角色(Yan et al., 2019)。这些序列不仅在叶绿体基因组序列的重排与稳定性维护中发挥着关键作用, 而且还能够影响相似物种乃至不同物种间拷贝数的差异(Wang et al., 2020)。在对枫杨的研究中, 我们观察到正向重复序列与回文重复序列的含量较为丰富, 这一发现与山核桃(*Carya cathayensis*)中的研究结果相吻合(Shen et al., 2022)。此外, 枫杨的SSRs中, 所有重复单元的A和T碱基含量显著较高, 这表明枫杨对AT碱基组合具有强烈的偏好性, 这与其他物种叶绿体基因组的研究结果相一致(Lee et al., 2025; Zhang et al., 2025)。这些质体SSRs有望成为枫杨属的候选分子标记, 对于推动群体遗传学和进化研究具有重要意义。

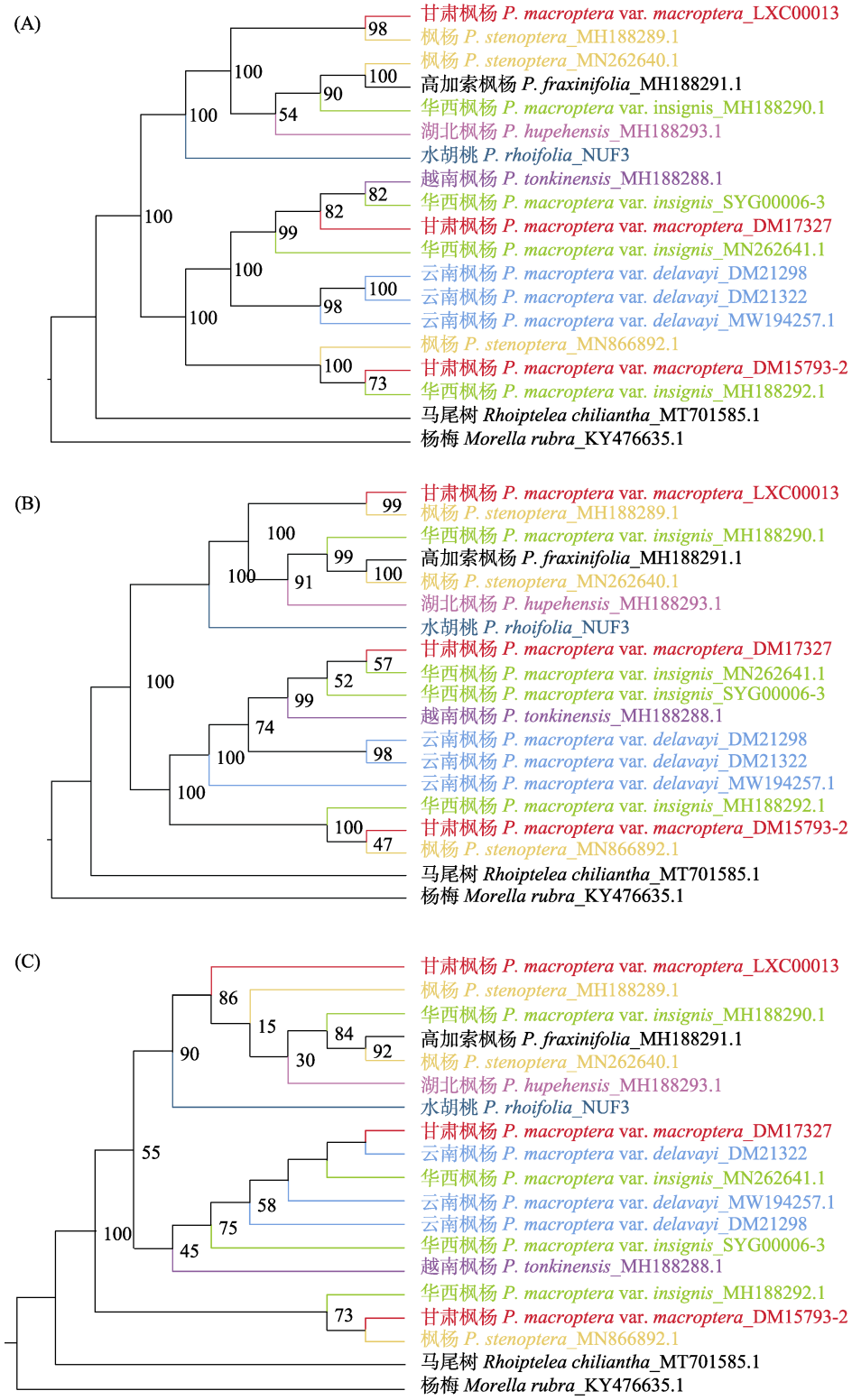

图7　基于不同数据集构建的枫杨属最大似然法系统发育树。(A)基于17个枫杨属个体的叶绿体全基因组序列; (B)基于17个枫杨属个体的蛋白质编码序列(CDS); (C)基于17个枫杨属个体的高变区*ndhF*序列。自展支持率(BS)值标记于进化树的分支。

Fig. 7　Maximum likelihood phylogenetic trees of *Pterocarya* based on different datasets. (A) Based on the complete chloroplast genomes of 17 individuals of *Pterocarya*; (B) Based on the coding sequences (CDS) of 17 individuals of *Pterocarya*; (C) Based on the highly variable *ndhF* region sequences of 17 individuals of *Pterocarya*. The bootstrap support values (BS) are labeled at branches in the evolutionary tree.

3.2 枫杨属的叶绿体比较基因组学

叶绿体序列之间的边界位置是叶绿体基因组进化研究的关键组成部分, 其中IR边界的收缩与扩张被视为被子植物叶绿体基因组变异的主要驱动力(Hansen et al., 2007; Huang et al., 2014)。尽管枫杨属的叶绿体基因组结构和基因排列顺序高度保守, 但在IRb/SSC (JSB)区域却观察到了一定的差异性。具体而言, 在越南枫杨、高加索枫杨和枫杨中检测到*ndhF*基因的存在, 而在其他枫杨种类中未检测到该基因。这一发现与胡桃科其他属不一致, 因为在已发表的研究中, 青钱柳属(*Cyclocarya*)、山核桃属、化香树属的JSB区域均能检测到*ndhF*基因的存在(李廷章等, 2024)。

利用mVISTA和DnaSP v6.12.03软件进行分析发现, 枫杨属叶绿体基因组中的LSC和SSC序列多样性显著高于IR, 这符合叶绿体基因组的普遍规律(Han et al., 2016; He et al., 2017)。序列比对进一步揭示了基因间隔区(如*ndhC–trnV-UAC*和*petA–psbJ*)以及编码区(如*ndhF*)存在显著变异。然而, 与通常观察到的多个高变区不同, 本研究中*ndhF*基因表现出尤为突出的高变异性(Pi > 0.01), 成为一个显著的变异热点。鉴于枫杨属为喜光植物, *ndhF*基因的主要功能与植物的光合作用有关(李景剑等, 2016), 因此我们推测*ndhF*基因可能是枫杨属在叶绿体水平进化适应过程中的重要遗传基础(陈淑瑾, 2015), 其在理论上具有成为DNA条形码的潜力, 然而本研究结果显示, 单独使用*ndhF*基因序列构建的系统发育树难以有效区分枫杨属的不同物种。这一结果符合植物分子系统学的普遍经验, 即多位点组合通常优于单基因标记(Fazekas et al., 2008; Jiang et al., 2022; Ma et al., 2024)。为此, 我们进一步将*ndhF*与本研究中筛选出的其他高变区(如*ndhC–trnV-UAC*、*petA–psbJ*)进行联合分析(附录5), 但其物种区分能力依然有限。令人深思的是, 即使我们采用完整的叶绿体基因组数据进行系统发育重建, 依然未能有效区分枫杨属的不同物种。这一系列结果共同指向一个结论: 对于枫杨属而言, 叶绿体基因组的整体进化速率可能不足以记录其近期物种分化的历史, 其序列变异未能积累出足够用于物种鉴定的系统发育信号。因此, 未来解析该属的系统发育关系, 可能需要转向核基因组标记, 以获取更高分辨率的进化信息。

3.3 叶绿体系统发育信号的复杂性及其对推断演化的启示

叶绿体基因组因其易于获取、处理便捷且蕴含丰富系统发育信息的特点, 在植物系统基因组学中得到广泛应用(如重建系统发育关系、推断网状进化及界定分类群系统位置等) (Dong et al., 2023)。然而, 受其自身遗传特性的限制, 叶绿体基因组在系统发育推断中能力有限, 有时甚至会导致结论偏差, 这一点在枫杨属系统发育研究中表现得尤为明显。近年来, 叶绿体基因组虽为探讨该属的核质冲突和生物地理学问题提供了关键证据(Yang et al., 2021, 2023; Zhou et al., 2021; Yan et al., 2024), 但本研究基于叶绿体全基因组及其分区序列的系统发育分析, 却揭示了其推断结果的高度复杂性: 全基因组与多数滑动窗口支持二分支结构, 而*ndhF*高变区与Window_13形成3分支, Window_5显示4分支, Window_1甚至解析出5分支。值得注意的是, 除云南枫杨在多数分析中保持单系性外, 其他物种如枫杨和甘肃枫杨的个体均分散在不同支系中。这一复杂的拓扑结构与基于核基因组的清晰结论形成了直接冲突(附录6) (Song et al., 2020)——后者将枫杨属清晰地划分为两个组: 水胡桃组(Section *Platyptera*), 包括水胡桃、甘肃枫杨和云南枫杨; 枫杨组(Section *Pterocarya*), 包括高加索枫杨、湖北枫杨和枫杨。显然, 本研究的叶绿体系统发育树未能完全支持现有的分类体系。

上述叶绿体系统发育拓扑结构的不一致性与核基因框架的冲突表明, 枫杨属的演化历史可能比现有分类体系所反映的更为复杂。在植物系统学中, 此类核质冲突常被归因于不完全谱系分选或由杂交导致的叶绿体捕获等进化机制(Rieseberg & Soltis, 1991; Degnan & Rosenberg, 2009; Li XP et al., 2025; Wu et al., 2025; Zhang et al., 2026)。枫杨属作为风媒异花授粉植物(Lu et al., 2024), 其花粉易于远距离传播, 为种间基因流提供了条件, 这从生物学特性上增加了上述机制发生的可能性。例如, 本研究中湖北枫杨与枫杨在叶绿体树中聚为一支而在核基因树中分离的模式, 就与壳斗科栎属(*Quercus*)等类群中因历史基因流导致的核质冲突案例相似(Hipp et al., 2020; Zhou et al., 2022)。

然而, 需要指出的是, 仅凭本研究的叶绿体数据可能无法区分这些潜在机制的具体贡献。这一系列复杂的信号恰恰说明, 依赖单一的、母系遗传的叶绿体基因组来推断枫杨属的物种分化历史存在固有风险。其系统发育树可能更多地反映了通过种子传播的母系谱系历史, 而无法捕捉由花粉流介导的或物种快速分化过程中产生的完整进化图景。因此, 本研究最重要的启示在于: 对于枫杨属这类可能存在复杂进化动态的类群, 基于叶绿体基因组的系统发育分析应被视为一个初步的、需要谨慎解读的视角, 而非定论。未来研究应整合核基因组数据(如SNPs或低拷贝基因)以提高系统发育分辨率(Hollingsworth et al., 2016), 并结合溯祖模型(Pei et al., 2020)来准确解析不完全谱系分选、基因流等过程的相对作用, 从而重建该属可靠的物种树并准确推断其生物地理历史。这一整合性策略不仅是解决枫杨属系统发育难题的关键, 也对研究其他具有类似生物学特性的植物类群具有重要的方法论意义。

ORCID

宋以刚 https://orcid.org/0000-0003-2584-2338

参考文献

Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: An online program to visualize the junction sites of chloroplast genomes. Bioinformatics, 34, 3030–3031.

Arredondo A, Mourato B, Nguyen K, Boitard S, Rodríguez W, Mazet O, Chikhi L (2021) Inferring number of populations and changes in connectivity under the n-island model. Heredity, 126, 896–912.

Avise JC (2000) Phylogeography: The History and Formation of Species. Harvard University Press, Harvard.

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: A web server for microsatellite prediction. Bioinformatics, 33, 2583–2585.

Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Research, 27, 573–580.

Cao DL, Zhang XJ, Qu XJ, Fan SJ (2022) Plastid phylogenomics sheds light on divergence time and ecological adaptations of the tribe Persicarieae (Polygonaceae). Frontiers in Plant Science, 13, 1046253.

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, Seberg O, Jørgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M (2007) A proposal for a standardised protocol to barcode all land plants. Taxon, 56, 295–299.

Chen SJ (2015) Analysis of causes of failure of introduction and cultivation of *Pterocarya stenoptera* C. DC in Zhongwei municipality. Ningxia Journal of Agriculture and Forestry Science and Technology, 56(8), 26–27. (in Chinese with English abstract) [陈淑瑾 (2015) 中卫市引种栽培枫杨失败原因分析. 宁夏农林科技, 56(8), 26–27.]

Chen T, Xiang TT, Zheng Q, Kuang XK, Kong QB, Zhou JS, Wang H, Zhou LJ, Feng SL, Yuan M, Yang HY, Ding CB (2025) Deciphering the evolutionary imprints of *Camellia oleifera* Abel.: Delineating its distinct phylogeographic structure and demographic history through microsatellite and plastid fragment. BMC Plant Biology, 25, 526.

Chen YX, Chen YS, Shi CM, Huang ZB, Zhang Y, Li SK, Li Y, Ye J, Yu C, Li Z, Zhang XQ, Wang J, Yang HM, Fang L, Chen Q (2018) SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience, 7, gix120.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution, 24, 332–340.

Dong Q, Zhang YL, Zhong SF, Zhang QY, Yang H, Yang H, Yi XX, Tan FQ, Chen C, Luo PG (2024) Conserved DNA sequence analysis reveals the phylogeography and evolutionary events of *Akebia trifoliata* in the region across the eastern edge of the Tibetan Plateau and subtropical China. BMC Ecology and Evolution, 24, 52.

Dong WP, Gao LM, Xu C, Song Y, Poczai P (2023) Editorial: Rise to the challenges in plastome phylogenomics. Frontiers in Plant Science, 14, 1200302.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19, 11–15.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using *Phred*. II. Error probabilities. Genome Research, 8, 186–194.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE, 3, e2802.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: Computational tools for comparative genomics. Nucleic Acids Research, 32, W273–W279.

Fu T, Wang ZL, Lin LJ, Lin L, Li W, Yuan DM (2018) Molecular phylogenetic analysis of wild *Cerasus* plants in South China. Journal of Nuclear Agricultural Sciences, 32, 2126–2134. (in Chinese with English abstract) [付涛, 王志龙, 林乐静, 林立, 李文, 袁冬明 (2018) 我国南方野生樱属植物的分子系统发育分析. 核农学报, 32, 2126–2134.]

Geng YF, Hu GX, Wang S, Xu JC (2019) Complete chloroplast

genome of the threatened *Rhoiptelea chiliantha* (Juglandaceae s.l.). Conservation Genetics Resources, 11, 317–319.

Gou W, Jia SB, Price M, Guo XL, Zhou SD, He XJ (2020) Complete plastid genome sequencing of eight species from *Hansenia*, *Haplosphaera* and *Sinodielsia* (Apiaceae): Comparative analyses and phylogenetic implications. Plants, 9, 1523.

Han YW, Duan D, Ma XF, Jia Y, Liu ZL, Zhao GF, Li ZH (2016) Efficient identification of the forest tree species in Aceraceae using DNA barcodes. Frontiers in Plant Science, 7, 1707.

Hansen DR, Dastidar SG, Cai ZQ, Penaflor C, Kuehl JV, Boore JL, Jansen RK (2007) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). Molecular Phylogenetics and Evolution, 45, 547–563.

He L, Qian J, Li XW, Sun ZY, Xu XL, Chen SL (2017) Complete chloroplast genome of medicinal plant *Lonicera japonica*: Genome rearrangement, intron gain and loss, and implications for phylogenetic studies. Molecules, 22, 249.

Hipp AL, Manos PS, Hahn M, Avishai M, Bodénès C, Cavender-Bares J, Crowl AA, Deng M, Denk T, Fitz-Gibbon S, Gailing O, González-Elizondo MS, González-Rodríguez A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C, Rodríguez-Correa H, Schulze ED, Simeone MC, Sork VL, Valencia-Avalos S (2020) Genomic landscape of the global oak phylogeny. New Phytologist, 226, 1198–1212.

Hollingsworth PM, Li DZ, van der Bank M, Twyford AD (2016) Telling plant species apart with DNA: From barcodes to genomes. Philosophical Transactions of the Royal Society B: Biological Sciences, 371, 20150338.

Hu Y, Wang Q, Zhang XX, Zhou W, Chen XY, Hu XS (2019) Advancing phylogeography with chloroplast DNA markers. Biodiversity Science, 27, 219–234. (in Chinese with English abstract) [胡颖, 王茜, 张新新, 周玮, 陈晓阳, 胡新生 (2019) 叶绿体DNA标记在谱系地理学中的应用研究进展. 生物多样性, 27, 219–234.]

Hu YH, Woeste KE, Zhao P (2017a) Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. Frontiers in Plant Science, 7, 1955.

Hu YH, Yan J, Feng XJ, Dang M, Woeste KE, Zhao P (2017b) Characterization of the complete chloroplast genome of wheel wingnut (*Cyclocarya paliurus*), an endemic in China. Conservation Genetics Resources, 9, 273–275.

Huang H, Shi C, Liu Y, Mao SY, Gao LZ (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. BMC Evolutionary Biology, 14, 151.

Huang Y, Jin XJ, Zhang CY, Li P, Meng HH, Zhang YH (2024) Plastome evolution of *Engelhardia* facilitates phylogeny of Juglandaceae. BMC Plant Biology, 24, 634.

Jiang S, Chen FL, Qin P, Xie H, Peng G, Li YL, Guo XH (2022) The specific DNA barcodes based on chloroplast genes for species identification of Theaceae plants. Physiology and Molecular Biology of Plants, 28, 837–848.

Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology, 21, 241.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution, 30, 772–780.

Kozlowski G, Bétrisey S, Song YG (2018) Wingnuts (*Pterocarya*) and Walnut Family: Relict Trees: Linking the Past, Present and Future. Natural History Museum Fribourg, Fribourg, Switzerland.

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE, 2, e508.

Kuang KR, Zheng SX, Li PQ, Lu AM (1979) Angiospermae, Dicotyledoneae: Myricaceae, Juglandaceae, Betulaceae. In: Flora Reipublicae Popularis Sinicae, Tomus 21 (ed. Delectis Florae Reipublicae Popularis Sinicae Agendae Academiae Sinicae Edita). Science Press, Beijing. (in Chinese) [匡可任, 郑世铨, 李沛琼, 路安民 (1979) 杨梅科、胡桃科、桦木科. 见: 中国植物志(第21卷) (中国科学院中国植物志编辑委员会). 科学出版社, 北京.]

Kurtz S (2003) The Vmatch large scale sequence analysis software. Computer Program, 4–12. http://www.vmatch.de/. (accessed on 2026-05-20)

Lee WS, Donrung W, Hung BM, Muhamad Rusly NH, Lee SY, Tanee T (2025) The complete plastid genome of *Citrus hystrix* DC. 1813 (Rutaceae) and its phylogenetic analysis. Mitochondrial DNA Part B, 10, 94–98.

Li JJ, Liu HX, Mao SZ, Zhao B, Huang SX (2016) Adaptive evolution of the *ndhF* gene in the genus *Rheum* (Polygonaceae). Guihaia, 36, 101–106. (in Chinese with English abstract) [李景剑, 刘合霞, 毛世忠, 赵博, 黄仕训 (2016) 大黄属(蓼科)植物*ndhF*基因的适应性进化. 广西植物, 36, 101–106.]

Li PW, Lu YB, Qin XM, Zhang Q (2025) Plastome phylogenomics unravels the evolutionary relationships and biogeographic history of Chloranthaceae. BMC Plant Biology, 25, 543.

Li TZ, Zhang SY, Wu SD, Liao SY, Tang CQ, Li YL (2024) Chloroplast genome and characteristic analysis of *Cyclocarya paliurus* (batalin) iljinsk. Journal of Guangxi Academy of Sciences, 40, 407–423. (in Chinese with English abstract) [李廷章, 章双玉, 吴书德, 廖水意, 汤楚

乔, 李玉玲 (2024) 青钱柳叶绿体基因组特征分析. 广西科学院学报, 40, 407–423.]

Li XP, Xu H, Yu JY, Han Y, Han S, Niu Y, Zhang FQ (2025) Comparative and phylogenetic analysis of *Potentilla* and *Dasiphora* (Rosaceae) based on plastid genome. BMC Plant Biology, 25, 176.

Liang CL, Wang L, Lei J, Duan BZ, Ma WS, Xiao SM, Qi HJ, Wang Z, Liu YQ, Shen XF, Guo S, Hu HY, Xu J, Chen SL (2019) A comparative analysis of the chloroplast genomes of four *Salvia* medicinal plants. Engineering, 5, 907–915.

Liu JQ, Fu MJ, Wu HY, Tian B (2021) Phylogeography of the Loess Plateau endemic plant *Prinsepia uniflora* inferred from the chloroplast DNA. Guihaia, 41, 396–403. (in Chinese with English abstract) [刘家奇, 付孟娇, 吴海洋, 田斌 (2021) 基于叶绿体DNA的黄土高原特有植物蔷核的谱系地理学. 广西植物, 41, 396–403.]

Liu KJ, Cui HX, Wu LD, Yang HM, Cao LL, Sun JH, Li B, Dong WP, Wang YH (2025) Chloroplast genome-based genetic variation and genetic diversity of *Ulmus pumila* (Ulmaceae) germplasm. Industrial Crops and Products, 233, 121452.

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research, 41, W575–W581.

Lu ZJ, Wang TR, Zheng SS, Meng HH, Cao JG, Song YG, Kozlowski G (2024) Phylogeography of *Pterocarya hupehensis* reveals the evolutionary patterns of a Cenozoic relict tree around the Sichuan Basin. Forestry Research, 4, e008.

Luo J, Chen JH, Guo WL, Yang ZF, Lim KJ, Wang ZJ (2022) Reassessment of *Annamocarya sinesis* (*Carya sinensis*) taxonomy through concatenation and coalescence phylogenetic analysis. Plants, 11, 52.

Ma WB, Ou YF, Dayananda B, Ji HJ, Yu T (2024) The complete chloroplast genome of *Rhododendron ambiguum* and comparative genomics of related species. Comparative Cytogenetics, 18, 143–159.

Mo ZH, Lou WR, Chen YQ, Jia XD, Zhai M, Guo ZR, Xuan JP (2020) The chloroplast genome of *Carya illinoinensis*: Genome structure, adaptive evolution, and phylogenetic analysis. Forests, 11, 207.

Mu XY, Tong L, Sun M, Zhu YX, Wen J, Lin QW, Liu B (2020) Phylogeny and divergence time estimation of the walnut family (Juglandaceae) based on nuclear RAD-Seq and chloroplast genome data. Molecular Phylogenetics and Evolution, 147, 106802.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32, 268–274.

O'Donnell RP, Wong DCJ, Phillips RD, Peakall R, Linde CC (2025) Discordance down under: Combining phylogenomics and fungal symbioses to detangle difficult nodes in a diverse tribe of Australian terrestrial orchids. Systematic Biology, 74, 434–452.

Pachter L, Batzoglou S, Spitkovsky VI, Beebee Jr WS, Lander ES, Berger B, Kleitman DJ (1999) A dictionary based approach for gene annotation. Proceedings of the Third Annual International Conference on Computational Molecular Biology, 6, 419–430.

Pei JW, Zhang YM, Nielsen R, Wu YF (2020) Inferring the ancestry of parents and grandparents from genetic data. PLoS Computational Biology, 16, e1008065.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Molecular Ecology, 14, 689–701.

Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. American Journal of Botany, 5, 65–84.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution, 34, 3299–3302.

Ruhlman TA, Jansen RK (2021) Plastid genomes of flowering plants: Essential principles. Methods in Molecular Biology, 2317, 3–47.

Shen JS, Li XQ, Chen X, Huang XL, Jin SH (2022) The complete chloroplast genome of *Carya cathayensis* and phylogenetic analysis. Genes, 13, 369.

Shen Z, Zhou BF, Liang YY, Wang JS, Yu RX, Shi Y, Ling SJ, Luo WJ, Lin QQ, Niu JW, Qiao LJ, Manos PS, Wang BS (2025) Teasing apart the sources of phylogenetic tree discordance across three genomes in the oak family (Fagaceae). BMC Plant Biology, 25, 919.

Shi LC, Chen HM, Jiang M, Wang LQ, Wu X, Huang LF, Liu C (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Research, 47, W65–W73.

Song YG, Li Y, Meng HH, Fragnière Y, Ge BJ, Sakio H, Yousefzadeh H, Bétrisey S, Kozlowski G (2020) Phylogeny, taxonomy, and biogeography of *Pterocarya* (Juglandaceae). Plants, 9, 1524.

Song YG, Walas Ł, Pietras M, Sâm HV, Yousefzadeh H, Ok T, Farzaliyev V, Worobiec G, Worobiec E, Stachowicz-Rybka R, Boratyński A, Boratyńska K, Kozlowski G, Jasińska AK (2021) Past, present and future suitable areas for the relict tree *Pterocarya fraxinifolia* (Juglandaceae): Integrating fossil records, niche modeling, and phylogeography for conservation. European Journal of Forest Research, 140, 1323–1339.

Wang Q, Cheng X, Zhou W, Zhang XX, Hu Y, Chen XY, Hu XS (2019) Assessing the ecological and evolutionary processes underlying cytonuclear interactions. Scientia

Sinica Vitae, 49, 951–964. (in Chinese with English abstract) [王茜, 程祥, 周玮, 张新新, 胡颖, 陈晓阳, 胡新生 (2019) 细胞核质互作形成的生态与进化过程分析. 中国科学: 生命科学, 49, 951–964.]

Wang W, Yang T, Wang HL, Li ZJ, Ni JW, Su S, Xu XQ (2020) Comparative and phylogenetic analyses of the complete chloroplast genomes of six almond species (*Prunus* spp. L.). Scientific Reports, 10, 10137.

Wu Q, Zheng PP, Hu YB, Wei FW (2014) Genome-scale analysis of demographic history and adaptive selection. Protein & Cell, 5, 99–112.

Wu Y, Liu R, Wang WJ, Li DZ, Burgess KS, Yu WB, Wang H (2025) High species discrimination in *Pedicularis* (Orobanchaceae): Plastid genomes and traditional barcodes equally effective via parsimony-informative sites. Plant Diversity, 47, 920–930.

Xu J, Song YG, Deng M, Jiang XL, Zheng SS, Li Y (2020) Seed germination schedule and environmental context shaped the population genetic structure of subtropical evergreen oaks on the Yun-Gui Plateau, Southwest China. Heredity, 124, 499–513.

Yan H, Zhang JS, Ye JF, Guo CC, Xiang XG (2021) Characterization of the complete chloroplast genome of *Pterocarya macroptera* var. *delavayi* (Juglandaceae). Mitochondrial DNA Part B, 6, 1344–1345.

Yan H, Zhou P, Wang W, Ye JF, Tan SL, Guo CC, Zhang WG, Zhu ZW, Liu YZ, Xiang XG (2024) Biogeographic history of *Pterocarya* (Juglandaceae) inferred from phylogenomic and fossil data. Journal of Systematics and Evolution, 62, 1165–1176.

Yan J, Han K, Zeng SY, Zhao P, Woeste K, Li JF, Liu ZL (2017) Characterization of the complete chloroplast genome of *Platycarya strobilacea* (Juglandaceae). Conservation Genetics Resources, 9, 79–81.

Yan M, Zhao XQ, Zhou JQ, Huo Y, Ding Y, Yuan ZH (2019) The complete chloroplast genomes of *Punica granatum* and a comparison with other species in Lythraceae. International Journal of Molecular Sciences, 20, 2886.

Yan MX, Dong SJ, Gong QY, Xu Q, Ge YQ (2023) Comparative chloroplast genome analysis of four *Polygonatum* species insights into DNA barcoding, evolution, and phylogeny. Scientific Reports, 13, 16495.

Yang Y, Forsythe ES, Ding YM, Zhang DY, Bai WN (2023) Genomic analysis of plastid-nuclear interactions and differential evolution rates in coevolved genes across Juglandaceae species. Genome Biology and Evolution, 15, evad145.

Yang YY, Qu XJ, Zhang R, Stull GW, Yi TS (2021) Plastid phylogenomic analyses of Fagales reveal signatures of conflict and ancient chloroplast capture. Molecular Phylogenetics and Evolution, 163, 107232.

Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ (2012) Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). Molecular Ecology Resources, 12, 98–108.

Zhang J, Zhang HX (2024) Characteristics and evolutionary history of the chloroplast genome of *Malus sieversii* in Xinjiang. Arid Zone Research, 41, 2110–2119. (in Chinese with English abstract) [张健, 张宏祥 (2024) 新疆野苹果叶绿体基因组特征及其进化历史. 干旱区研究, 41, 2110–2119.]

Zhang R, Stull GW, Jin JJ, Wang YH, Guo Y, Yang ZY, Li HT, An KL, Charboneau JLM, Folk RA, Cardoso D, de Queiroz LP, Bruneau A, Soltis PS, Soltis DE, Smith SA, Li DZ, Yi TS (2026) Phylogenetic resolution and conflict in the species-rich flowering plant family Leguminosae. Systematic Biology, 75, 236–258.

Zhang ZR, Yang X, Li WY, Peng YQ, Gao J (2022) Comparative chloroplast genome analysis of *Ficus* (Moraceae): Insight into adaptive evolution and mutational hotspot regions. Frontiers in Plant Science, 13, 965335.

Zhang ZR, Shi WB, Hu SQ, Mou LR, Shi C, Zhu BY, Yang J (2025) Comprehensive chloroplast genome analysis of four *Callitriche* (Plantaginaceae) species for phylogenetic and conservation insights. Horticulturae, 11, 66.

Zhao XQ, Yan M, Ding Y, Huo Y, Yuan ZH (2019) Characterization and comparative analysis of the complete chloroplast genome sequence from *Prunus avium* 'Summit'. PeerJ, 7, e8210.

Zhou BF, Yuan S, Crowl AA, Liang YY, Shi Y, Chen XY, An QQ, Kang M, Manos PS, Wang BS (2022) Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. Nature Communications, 13, 1320.

Zhou HJ, Hu YH, Ebrahimi A, Liu PL, Woeste K, Zhao P, Zhang SX (2021) Whole genome based insights into the phylogeny and evolution of the Juglandaceae. BMC Ecology and Evolution, 21, 191.

(责任编委: 邱英雄　责任编辑: 黄祥忠)

附录　Supplementary Material

https://www.biodiversity-science.net/CN/10.17520/biods.2025475