安徽宣城木瓜叶绿体基因组特征与系统发育关系研究

汪思嘉[1], 戚康茹[1], 赵志强[1], 方清影[1], 余函纹[1], 梁　娟[1*], 查良平[1,2*]

(1. 安徽中医药大学药学院, 安徽 合肥 230012;　2. 省部共建安徽道地中药材品质提升协同创新中心, 安徽 合肥 230012)

摘要: 中药木瓜来源于蔷薇科植物贴梗海棠 *Chaenomeles speciosa* (Sweet) Naka 的干燥果实, 安徽宣城产的“宣木瓜”为道地药材, 其药用历史悠久, 也是著名的药食同源中药。本研究采用 Illumina 平台测序, 组装注释获得宣城木瓜的叶绿体全基因组, 利用 REPuter、MISA 分析重复序列, 基于相对同义密码子使用度 (relative synonymous codon usage, RSCU) 分析密码子使用偏好性, 并结合木瓜属近缘种进行反向重复区 (inverted repeat, IR) 边界比较与系统发育分析。结果显示, 宣城木瓜叶绿体基因组全长 159 991 bp, 呈典型四分体结构, GC 含量为 36.5%。共注释获得 126 个基因, 包括 83 个蛋白编码基因、35 个 tRNA 和 8 个 rRNA。共鉴定到 76 个简单重复序列 (SSR) 和 55 个长重复序列, SSR 位点以 A/T 单核苷酸为主, 多分布于大单拷贝区 (large single copy, LSC) 及间隔区。密码子使用偏好性分析表明, 30 种密码子具有偏好性 (RSCU > 1), 多偏好 A/U 结尾。IR 边界比较显示木瓜属内高度保守。基因组变异分析, 鉴定出位于 LSC 区域的 4 个高度可变区 *trnR-atpA*、*petN-psbM*、*trnT-GGU-psbD*、*trnV-M*。系统发育分析表明, 木瓜属形成一个单系分支, 其中宣木瓜与毛叶木瓜、西藏木瓜亲缘关系最近。本研究首次对宣城木瓜叶绿体全基因组和系统发育进行了全面解析, 筛选的 SSR 位点和 4 个基因高变异区可为其种质鉴定与系统发育研究提供基础。

关键词: 宣木瓜; 叶绿体基因组; 密码子偏好性; 重复序列; 系统发育

中图分类号: R914　　**文献标识码**: A　　**文章编号**: 0513-4870(2026)08-3207-10

Characterization and phylogenetic analysis of the complete chloroplast genome of *Chaenomeles speciosa* from Xuancheng, Anhui

WANG Si-jia[1], QI Kang-ru[1], ZHAO Zhi-qiang[1], FANG Qing-ying[1], YU Han-wen[1], LIANG Juan[1*], ZHA Liang-ping[1,2*]

(*1. School of Pharmacy, Anhui University of Chinese Medicine, Hefei 230012, China;　2. Anhui Collaborative Innovation Center for Quality Improvement of Genuine Chinese Medicinal Materials, Hefei 230012, China*)

Abstract: The traditional Chinese medicinal material Mugua is derived from the dried fruit of *Chaenomeles speciosa* (Sweet) Nakai (Rosaceae). The material produced in Xuancheng, Anhui Province, is regarded as a geo-authentic medicinal material and is commonly referred to as "Xuan Mugua". *C. speciosa* is a medicinal and edible plant with a long history of use in traditional Chinese medicine. In this study, we sequenced and assembled the complete chloroplast genome of *C. speciosa* using the Illumina platform and performed comprehensive genome annotation. Repetitive sequences were analyzed with REPuter and MISA, codon usage bias was evaluated *via* relative synonymous codon usage (RSCU), and comparative analyses of inverted repeat (IR) boundaries and phylogenetic relationships were conducted among closely related *Chaenomeles* species. The chloroplast genome is

收稿日期: 2026-03-31;　修回日期: 2026-05-14.

基金项目: 安徽高校协同创新项目 (GXXT-2023-072); 安徽中医药大学“常青藤工程”专项基金 (CQT20250204); 国家中医药管理局高水平中医药重点学科中药资源学 (药用植物学) 建设项目 (zyyzdxk-2023095); 安徽省高校杰出青年科研项目 (2023AH020036); 安徽省优秀青年教师重点培育项目 (YQZD2023049).

*通讯作者 E-mail: Ljuan1875@ahtcm.edu.cn; zlp_ahtcm@126.com

DOI: 10.16438/j.0513-4870.2026-0353

159 991 bp in length, exhibiting a typical quadripartite structure with a GC content of 36.5%, and encodes 126 genes, including 83 protein-coding genes, 35 tRNAs, and 8 rRNAs. Seventy-six simple sequence repeats (SSRs) and 55 long repeat sequences were identified, with SSRs primarily consisting of A/T mononucleotides and predominantly located in the large single copy (LSC) and intergenic regions. Codon usage analysis showed that 30 codons have a preference (RSCU > 1), mostly ending with A/U. Comparative analysis of IR boundaries indicated high conservation within the genus. Genome variation analysis revealed four highly variable regions in the LSC (*trnR-atpA*, *petN-psbM*, *trnT-GGU-psbD*, *trnV-M*). Phylogenetic analysis confirmed that *Chaenomeles* is monophyletic, with *C. speciosa* closely related to *C. japonica* and *C. tibetica*. This work provides the first detailed characterization of the *C. speciosa* chloroplast genome and phylogeny, and the identified SSRs and highly variable regions represent valuable resources for germplasm identification and evolutionary studies in *Chaenomeles*.

Key words: *Chaenomeles speciosa*; chloroplast genome; codon usage bias; repetitive sequence; phylogeny

中药木瓜来源于蔷薇科植物贴梗海棠*Chaenomeles speciosa* (Sweet) Nakai的干燥近成熟果实, 该植物又名皱皮木瓜、贴梗海棠、铁脚梨等, 主产于安徽宣城、湖北长阳、重庆綦江等地, 以安徽宣城产者为道地药材, 故名“宣木瓜”[1]。宣木瓜入药始载于《名医别录》, 其果实味酸、性温, 归肝、脾经, 具有舒筋活络、和胃化湿的功效, 用于治疗湿痹拘挛、腰膝关节酸重疼痛等症[2]。现代药理研究表明, 宣木瓜具有抗炎镇痛、抗氧化、免疫调节、抗肿瘤及降血糖等药理活性[3]。作为首批入选药食同源目录的中药, 宣木瓜被广泛应用于中医临床、食品与饮品开发、现代保健品与药品研发领域[4,5]。

《中国植物志》收载了木瓜属5种植物, 包括皱皮木瓜*C. speciosa*、光皮木瓜*Chaenomeles sinensis* (Thouin) Koehne、日本木瓜*Chaenomeles japonica* (Thunb.) Lindl. ex Spach、西藏木瓜*Chaenomeles thibetica* T. T. Yu、毛叶木瓜*Chaenomeles cathayensis* (Hemsl.) C. K. Schneid[6]。木瓜属种类较少, 但近缘种间形态特征相似, 且部分分类特征易受环境和栽培条件影响, 不同品种及居群间又存在较复杂的遗传分化[7]。木瓜属不同品种间表现出较高遗传多样性, 在品种鉴定过程中, 除了形态学鉴定等传统方法, 还需借助随机扩增多态性DNA (random amplified polymorphic DNA, RAPD) 分子标记[8]和DNA条形码[9]等技术加以辅助。近年来, 皱皮木瓜核基因组已被解析, 其大小约为650 Mb, 具有较高的杂合度, 显示出较为复杂的基因组结构[10]。然而, 传统DNA条形码在区分近缘物种时分辨率有限[11], 核基因组的高杂合度及复杂重复序列可能影响序列解析, 同时其双亲遗传模式也增加了谱系分析的复杂性[12], 这些方法在实际应用中仍存在一定局限性。因此, 有必要结合更高分辨率的分子数据对木瓜属植物开展系统研究。

叶绿体基因组具有母系遗传、结构保守及无重组的特性, 能有效规避核基因组的复杂干扰[13]。叶绿体全基因组分析已被广泛应用于植物系统发育重建、物种界定、DNA条形码开发及遗传多样性评估等研究领域, 通过比较基因组学可筛选高变异区域作为物种特异性标记, 用于中药材的快速鉴定与种质资源评价[14]。例如针对分类鉴定困难的黄精属药用植物, 筛选出14个种间高变异区, 通过序列联合分析将物种鉴定分辨率提升至73.33%, 成功筛选出*psaJ-rpl33*、*rps16-trnQ*等高分辨率分子标记组合[15]。Li等[16]以大黄药材3个基原物种为研究对象, 鉴定出*rps16-trnQ*、*psaA-ycf3*等5个高变异区, 据此构建的特异性DNA条形码可准确区分大黄的来源, 为大黄药材市场的规范化管理提供了可靠的分子鉴定依据。目前, 木瓜属部分植物的叶绿体基因组已有报道[17,18], 然而, 作为“木瓜”药材的重要基原之一, 宣城木瓜叶绿体基因组的系统研究尚不充分, 其基因组结构特征、重复序列组成、潜在分子标记及系统发育位置仍有待进一步明确。

本研究基于高通量测序技术对宣城木瓜叶绿体全基因组进行组装与注释; 结合已发表的木瓜属4种植物的叶绿体基因组信息, 利用生物信息学手段对5个木瓜属植物叶绿体的密码子使用偏好性、叶绿体基因组边界特征及系统发育关系进行分析, 研究结果将为木瓜属植物的叶绿体遗传多样性、进化及系统发育研究提供参考。

材料与方法

数据来源与植物材料 本研究所用样品于2025年7月采集于安徽省宣城市宣木瓜产业振兴发展有限公司基地种植的宣木瓜植株。样品生长于人工栽培环境。采集其生长状态良好且无病虫害的幼嫩叶片, 液氮速冻, 于−80 ℃保存。凭证标本保存于安徽中医药大学中药资源中心。凭证标本编号为: AHCM-20250715-001。

基因组DNA提取和测序 使用植物DNA提取试剂盒 (TaKaRa Mini BEST Plant Genomic DNA Extrac-

tion Kit) 提取DNA。通过1%琼脂糖凝胶电泳评估提取DNA的质量, 并使用NanoDrop2000分光光度计 (美国Thermo Fisher Scientific公司) 测定其浓度。随后, 制备测序文库, 使用Agilent2100生物分析仪 (美国Agilent公司) 进一步验证片段大小分布。使用Qubit3.0荧光计 (美国Thermo Fisher Scientific公司) 对文库进行精确定量, 构建插入片段为350 bp的paired-end (PE) 文库, 利用Illumina NovaSeq6000平台进行高通量测序, 为保障测序数据的可靠性, 采用fastp软件对原始数据进行质控, 去除低质量reads及接头序列, 获得高质量cleandata用于后续组装分析。

叶绿体基因组组装和注释 为确保叶绿体基因组的准确性, 选择皱皮木瓜 (OL450372.1) 作为参照基因组, 采用metaSPAdes软件对宣木瓜叶绿体基因组的测序序列片段进行初步组装。使用在线工具CPGA-VAS2 (http://47.96.249.172: 16019/analyzer/annotate) [19]注释叶绿体基因软件进行基因注释, 并进行了手动人工校正。为了进一步验证编码区域的准确性, 利用BLAST进行序列的比对检索。利用OGDRAW软件 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html)[20]。将GenBank文件绘制叶绿体基因组图谱并将序列数据和基因注释信息上传至NCBI (National Center for Biotechnology Information) 网站数据库。

重复序列分析 采用REPuter在线软件 (https://bibiserv.cebitec.uni-bielefeld.de/reputer/)[21]精准识别基因组中可能存在的正向、反向、回文及互补等长重复序列 (long repeat sequence, LSR)。参数设置如下: 最小重复长度设定为30 bp, Hamming距离阈值为3, 最大重复距离限制在5 000 bp以内。此外, 利用TandemRepeats-Finder在线网站识别串联重复序列, 参数均设置为默认值。使用在线工具MISA分析宣木瓜叶绿体基因组中的简单重复序列 (simple sequence repeat, SSR)。设定各重复单元的最小重复次数如下: 单核苷酸为10次, 二核苷酸为5次, 三核苷酸为4次, 四、五及六核苷酸均为3次。

密码子使用偏好性分析 使用CPStools软件 (https://doi.org/10.1002/imo2.25)[22]中的相对同义密码子使用度 (relative synonymous codon usage, RSCU) 模块进行计算, 预测密码子的偏好性, 其中RSCU > 1代表具有密码子使用偏好性, RSCU < 1表示密码子使用率较低, 而RSCU = 1则表示密码子没有偏好性。

叶绿体基因组序列结构变异分析以及IR边界收缩和扩张分析 选取与宣木瓜同属的木瓜属物种的叶绿体全基因组序列, 从NCBI数据库下载GenBank格式文件。使用在线工具IRscope (http://genocat.tools/tools/irscope.html)[23]识别各物种的大单拷贝区 (large single copy, LSC)、小单拷贝区 (small single copy, SSC)、两个反向重复区 (inverted repeat a/b, IRa/b) 区域边界并进行可视化分析。使用在线比对工具mVISTA (https://genome.lbl.gov/vista/mvista/submit.shtml)[24], 选择Shuffle-LAGAN模式进行全基因组多重比对, 识别保守区域和高变异区域。基于DnaSP v5.10软件, 本研究对木瓜属物种叶绿体基因组进行了滑动窗口分析, 以解析其核苷酸多样性 (nucleotide diversity, Pi) 的分布特征。分析设置窗口长度为800 bp, 步长为200 bp, 旨在筛选高变异区段 (Pi > 0.015) 作为潜在的分子鉴定标记。

系统发育分析 从GeneBank中下载已公布的12种物种完整的叶绿体基因组序列, 包括11种蔷薇科植物日本木瓜 (*C. japonica*)、西藏木瓜 (*C. thibetica*)、毛叶木瓜 (*C. cathayensis*)、光皮木瓜 (*C. sinensis*)、云南多依 [*Docynia delavayi* (Franch.) C.K.Schneid]、山芥子 [*Malus baccata* (L.) Borkh.]、佛罗伦萨海棠 [*Eriolobus florentinus* (Zuccagni) Stapf]、沙梨 [*Pyrus pyrifolia* (Burm. f.) Nakai]、桃 [*Prunus persica* (L.) Batsch]、野草莓 (*Fragaria vesca* L. subsp. *vesca*)、空心泡 (*Rubus rosifolius* Sm)。并以鼠李科植物枣 (*Ziziphus jujuba* Mill) 作为外类群, 采用CPStools软件[22]提取上述12个物种叶绿体基因组中所有共有的CDS基因, 确保比对序列的同源性。用MAFFT软件对共同序列进行多重比对, 使用MEGA7软件[25]中的最大似然法 (maximum likelihood, ML) 构建系统发育树, 重复次数设定为1 000次。通过iTOL在线工具 (https://itol.embl.de/tree/)[26]对生成的系统发育树进行可视化与注释, 以便清晰地展示物种间的亲缘关系和分支结构。

结果

1 宣城木瓜叶绿体基因组的结构与特征分析

对宣城木瓜叶绿体基因组进行测序、组装及注释, 获得其完整序列 (GenBank登录号: PZ183445)。该基因组全长为159 991 bp, 呈现典型的环状四分体结构, LSC区长度为88 001 bp、SSC区长度为19 230 bp, 以及一对IRa和IRb区长度均为26 380 bp。基因组总GC含量为36.5%, 其中IR区GC含量高达42.6%, 显著高于LSC和SSC区, 这与高等植物叶绿体基因组中IR区含有高GC含量的rRNA基因 (*rrn16*、*rrn23*、*rrn4.5*、*rrn5*) 的保守特征一致 (图1)。基因注释结果显示, 该基因组共编码126个功能基因, 包括83个CDS基因、35个tRNA基因及8个rRNA基因, 基因组成与数量与已报道的木瓜属及其近缘种高度相似 (表1)。14个不

同的基因有一个内含子, 3 个基因 (*rps12*、*clpP* 和 *ycf3*) 有两个内含子。上述结果表明, 宣城木瓜叶绿体基因组在结构、GC 含量分布及基因组成上均表现出高度的保守性, 进一步证实了木瓜属叶绿体基因组的整体进化稳定性。

2 重复序列和 SSR 位点分析

应用 MISA 分析宣木瓜 SSR 位点, 共检测到 76 个 SSR 位点, 其中以单核苷酸最多, 共有 72 个 (图 2)。其中 A/T 单核苷酸占比 97.2%, 3 个 SSR 为 AT/TA 二核苷酸; 1 个四核苷酸, 64 个 SSR 分布在 LSC 区, 10 个 SSR 分布在 SSC 区, 1 个 SSR 分布在 IRa 区, 1 个 SSR 分布在 IRb 区; 8 个 SSR 分布在 CDS 基因, 56 个 SSR 分布在基因间隔区 (intergenic spacer, IGS), 12 个 SSR 分布在 intron。此外在宣木瓜叶绿体基因组中共检测到 55 个长度在 30～58 bp 的 LSR, 其中 3 个互补 (complement, C), 29 个 LSR 为正向 (forward, F), 17 个 LSR 为回文 (palindrome, P), 6 个 LSR 为反向 (reverse, R), 主要分布在 IGS 区。

3 密码子使用偏好性分析

对 5 个木瓜属物种的密码子使用偏好性分析, 结果显示 RSCU 值在各物种间高度一致, 无明显差异 (图 3)。共统计了 64 个密码子 (包括 3 个终止密码子), 其中 30 个密码子表现出显著偏好性 (RSCU > 1), 32 个密码子使用频率较低 (RSCU < 1), 而色氨酸和甲硫氨酸对应的密码子无明显偏好 (RSCU = 1)。偏好密码子多以 A 或 U 结尾, 如亮氨酸的 TTA (RSCU = 2.03) 和丙氨酸的 GCT (RSCU = 1.85), 而使用频率低的密码子多以 G 或 C 结尾, 如亮氨酸的 CTC (RSCU = 0.35) 和精氨酸的 CGC (RSCU = 0.43)。RSCU 值最高的密码子为亮氨酸的 TTA (2.03), 最低的为亮氨酸的 CTC (0.35)。终止密码子中, TAA 的 RSCU 最高 (1.70)。各氨基酸的密码子使用模式显示, 亮氨酸的 6 个密码子中, TTA 和 CTT 为偏好密码子, 而 CTC 和 CTG 使用频率极低; 精氨酸的 6 个密码子中, AGA 和 CGA 为偏好密码子, 而 CGC 和 CGG 使用频率较低。上述结果表明木瓜属叶绿体基因组密码子强烈偏好以 A/U 结尾, 这一特征与大多数被子植物叶绿体基因组的演化规律一致。

4 叶绿体基因组的结构变异分析

以宣城木瓜叶绿体基因组为参考, 利用 mVISTA 软件对木瓜属 5 个物种的全基因组进行全局比对 (图 4)。结果显示, 木瓜属叶绿体基因组在整体结构上高度保守, 未检测到大片段重排或倒位事件。蛋白编码区和外显子区域序列表现出较高的保守性, 而序列变异主要发生在非编码区。除此之外, LSC 区和 SSC 区的变异程度明显高于 IR 区, IR 区在全长范围内均表现为连续的保守区域。在 5 种木瓜属绿体基因组的核苷

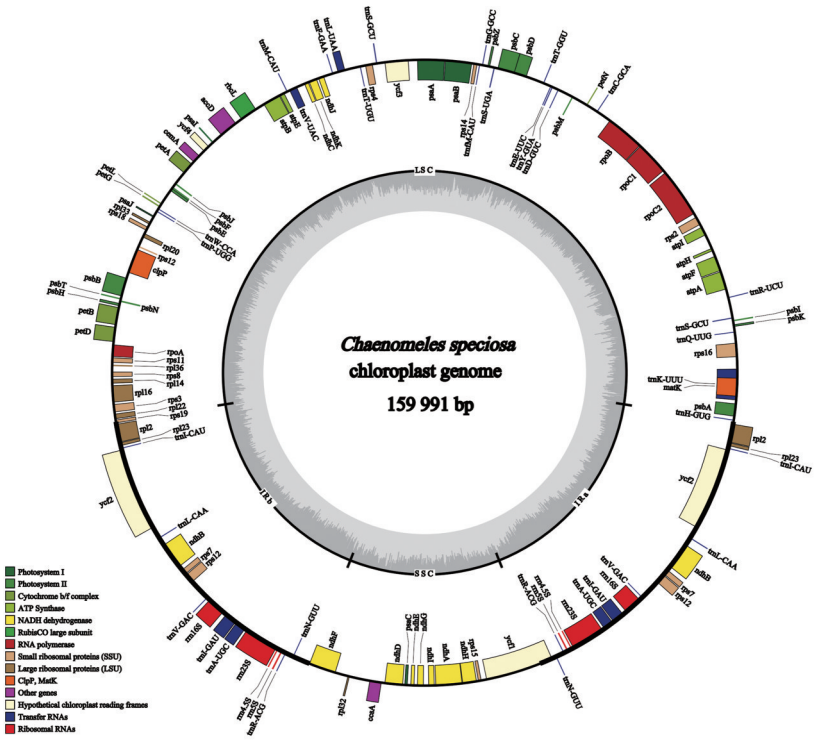


Figure 1 Chloroplast genome map of *Chaenomeles speciosa*. Genes are color-coded by function as indicated in the legend. Genes inside and outside the circle are transcribed in clockwise and counterclockwise directions, respectively. The gray region of the inner circle represents the GC content of the chloroplast genome. LSC: Large single-copy; SSC: Small single-copy; IRa: Inverted repeat A; IRb: Inverted repeat B

Table 1 Annotation information of the chloroplast genome of *Chaenomeles speciosa*. (2): Copy Number; #: One intron; ##: Two introns

Category	Gene group	Gene name	Number
Photosynthesis	Subunits of photosystem I	*psaA*, *psaB*, *psaC*, *psaJ*, *psaI*	5
	Subunits of photosystem II	*psbA*, *psbK*, *psbI*, *psbM*, *psbD*, *psbC*, *psbZ*, *psbJ*, *psbF*, *psbE*, *psbB*, *psbT*, *psbN*, *psbH*	14
	Subunits of NADH dehydrogenase	*ndhJ*, *ndhK*, *ndhC*, *ndhB*$^{(2)\#}$, *ndhF*, *ndhD*, *ndhE*, *ndhG*, *ndhI*, *ndhA*$^{\#}$, *ndhH*	12
	Subunits of cytochrome b/f complex	*petN*, *petA*, *petL*, *petG*, *petB*$^{\#}$, *petD*$^{\#}$	6
	Large subunit of rubisco	*rbcL*	1
	Subunits of ATP synthase	*atpA*, *atpF*$^{\#}$, *atpH*, *atpI*, *atpE*, *atpB*	6
Self-replication	Proteins of large ribosomal subunit	*rpl33*, *rpl20*, *rpl36*, *rpl14*, *rpl16*$^{\#}$, *rpl22*, *rpl2*$^{(2)\#}$, *rpl23*$^{(2)}$, *rpl32*	11
	Proteins of small ribosomal subunit	*rps12*$^{(2)\#\#}$, *rps16*$^{\#}$, *rps2*, *rps14*, *rps4*, *rps18*, *rps11*, *rps8*, *rps3*, *rps19*, *rps7*$^{(2)}$, *rps15*	14
	Subunits of RNA polymerase	*rpoC2*, *rpoC1*$^{\#}$, *rpoB*, *rpoA*	4
	Ribosomal RNAs	*rrn16S*$^{(2)}$, *rrn23S*$^{(2)}$, *rrn4.5S*$^{(2)}$, *rrn5S*$^{(2)}$	8
	Transfer RNAs	*trnH-GUG*, *trnK-UUU*$^{\#}$, *trnQ-UUG*, *trnS-GCU*$^{(2)}$, *trnR-UCU*, *trnC-GCA*, *trnD-GUC*, *trnY-GUA*, *trnE-UUC*, *trnT-GGU*, *trnS-UGA*, *trnG-GCC*, *trnfM-CAU*, *trnT-UGU*, *trnL-UAA*$^{\#}$, *trnF-GAA*, *trnV-UAC*$^{\#}$, *trnM-CAU*, *trnW-CCA*, *trnP-UGG*, *trnI-CAU*$^{(2)}$, *trnL-CAA*$^{(2)}$, *trnV-GAC*$^{(2)}$, *trnI-GAU*$^{(2)\#}$, *trnA-UGC*$^{(2)\#}$, *trnR-ACG*$^{(2)}$, *trnN-GUU*$^{(2)}$	35
Other genes	Maturase	*matK*	1
	Protease	*clpP*$^{\#\#}$	1
	Envelope membrane protein	*cemA*	1
	Acetyl-CoA carboxylase	*accD*	1
	c-Type cytochrome synthesis gene	*ccsA*	1
	Conserved open reading frames	*ycf3*$^{\#\#}$, *ycf4*, *ycf2*$^{(2)}$, *ycf1*	5
Total			126

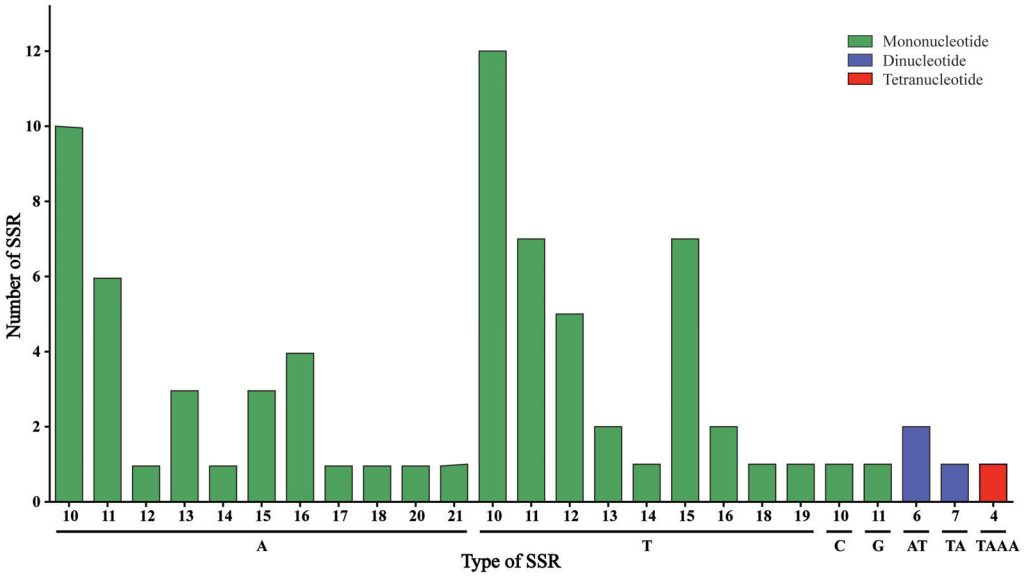


Figure 2 Types and numbers of SSRs in the chloroplast genome of *Chaenomeles speciosa*

酸多态性分析结果中。窗口中的Pi值从0到0.021 00不等, 平均值为0.008 13。根据 (Pi > 0.015) 在分析的序列中鉴定出四个高度可变区 (图5), 位于LSC区的*trnR-atpA*、*petN-psbM*、*trnT-GGU-psbD*、*trnV-M*。为进一步评估这些区域作为分子标记的潜力, 计算了各区域的长度和Pi: *trnR-atpA* 长度为1 302 bp, Pi为0.020 00; *petN-psbM* 长度为1 037 bp, Pi为0.021 00; *trnT-GGU-psbD* 长度为771 bp, Pi为0.016 00; *trnV-M* 长度为1 167 bp, Pi为0.018 30。

5 IR边界收缩与扩张分析

为探究木瓜属及其近缘物种叶绿体基因组的IR边界扩张与收缩特征, 对宣木瓜、日本木瓜、西藏木瓜、

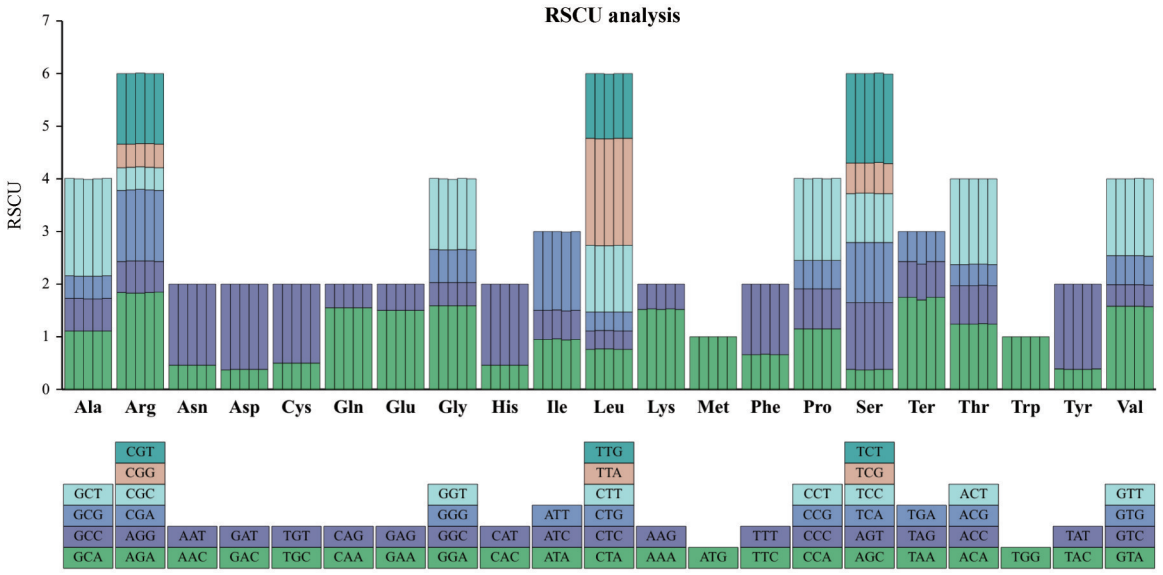


Figure 3 Codon content of the 20 amino acids in all protein-coding genes of the chloroplast genome of *Chaenomeles* species. Different colors in the bars correspond to the codons displayed in the same colors below the figure. RSCU: Relative synonymous codon usage; Ala: Alanine; Arg: Arginine; Asn: Asparagine; Asp: Aspartic acid; Cys: Cysteine; Gln: Glutamine; Glu: Glutamic acid; Gly: Glycine; His: Histidine; Ile: Isoleucine; Leu: Leucine; Lys: Lysine; Met: Methionine; Phe: Phenylalanine; Pro: Proline; Ser: Serine; Thr: Threonine; Trp: Tryptophan; Tyr: Tyrosine; Val: Valine

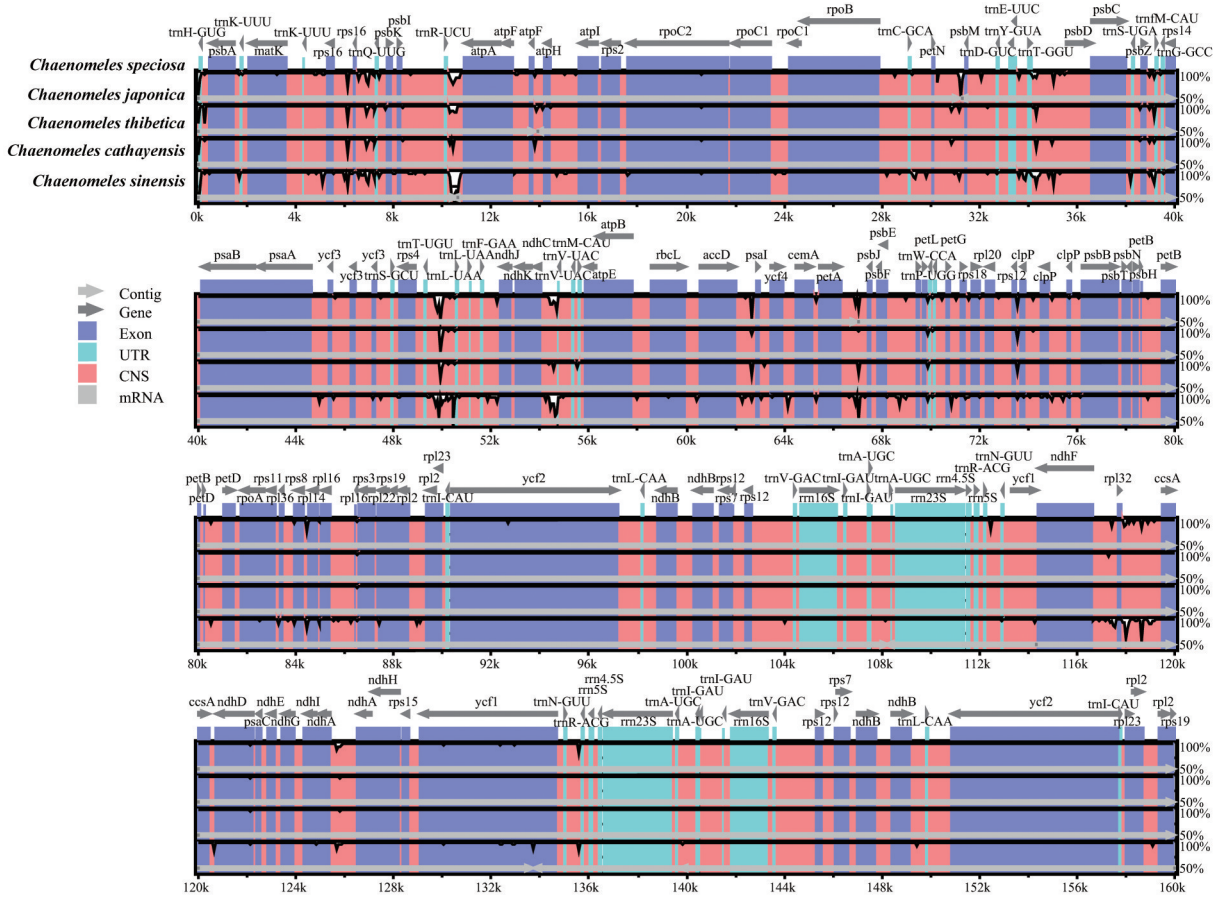


Figure 4 Comparison of chloroplast genome sequence identity between *Chaenomeles speciosa* and four other *Chaenomeles* species. The horizontal axis represents genome size, while the vertical axis denotes sequence similarity ranging from 50% to 100%. Coding regions, non-coding regions, and tRNA or rRNA genes are distinguished in the graph using different colors

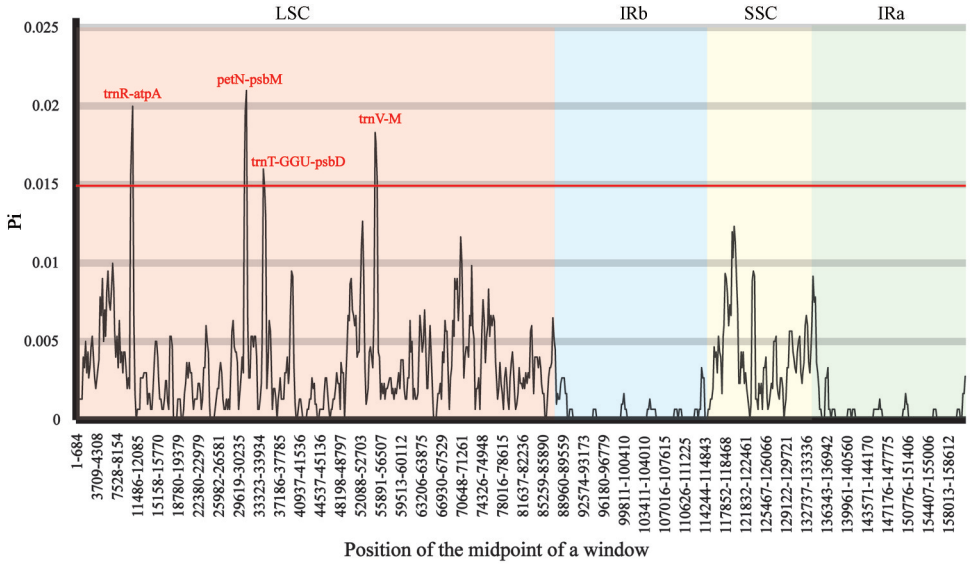


Figure 5 Nucleotide diversity of the five *Chaenomeles* species chloroplast genomes. Window length: 600 bp; Step size: 200 bp. Pi: Nucleotide diversity

毛叶木瓜、光皮木瓜五个木瓜属植物叶绿体基因组的IR边界及长度特征进行比较分析 (图6), 结果表明木瓜属IR区整体高度保守, 但种间存在细微差异。五种木瓜的IR长度范围为26 357～26 393 bp, 其中日本木瓜IR最长26 393 bp, 光皮木瓜最短26 357 bp, 宣城木瓜、西藏木瓜、毛叶木瓜IR长度完全一致为26 379 bp。从基因组总长来看, 毛叶木瓜最大长度为160 040 bp, 光皮木瓜最小长度为159 436 bp, 差异主要来源于LSC区和SSC区的长度变异。

6 叶绿体基因组系统发育分析

本研究以鼠李科植物枣 (*Z. jujuba*) 为外类群, 采用最大似然法构建了蔷薇科多个属的系统发育关系。结果显示, 木瓜属的5个物种构成一个高支持率的单系分支 (图7), 其中毛叶木瓜 (*C. cathayensis*) 与西藏木瓜 (*C. thibetica*) 亲缘关系最近, 支持率为99%, 光皮木瓜 (*C. sinensis*) 位于整个属的基部。在整个蔷薇科植

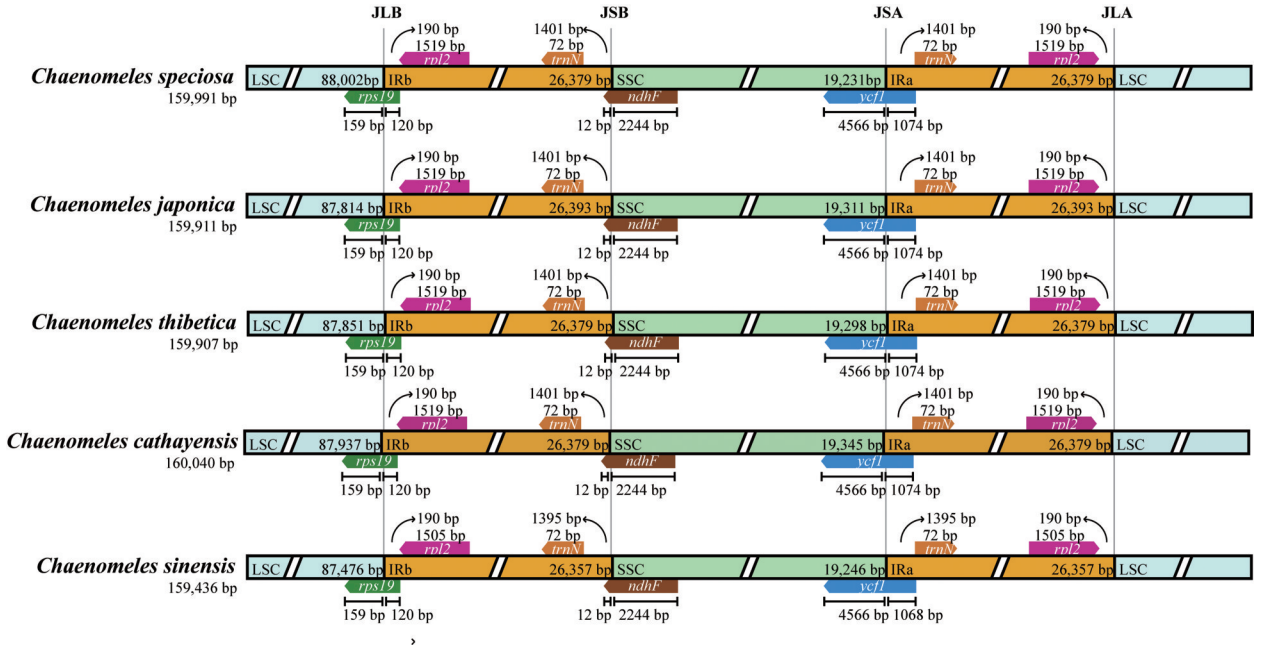


Figure 6 Boundary comparison of IR and SC regions of chloroplast genomes in five *Chaenomeles* species. Gene names and their sequence lengths are displayed in boxes and above the boxes, respectively. JLB: LSC/IRb boundary; JSB: IRb/SSC boundary; JSA: SSC/IRa boundary; JLA: IRa/LSC boundary

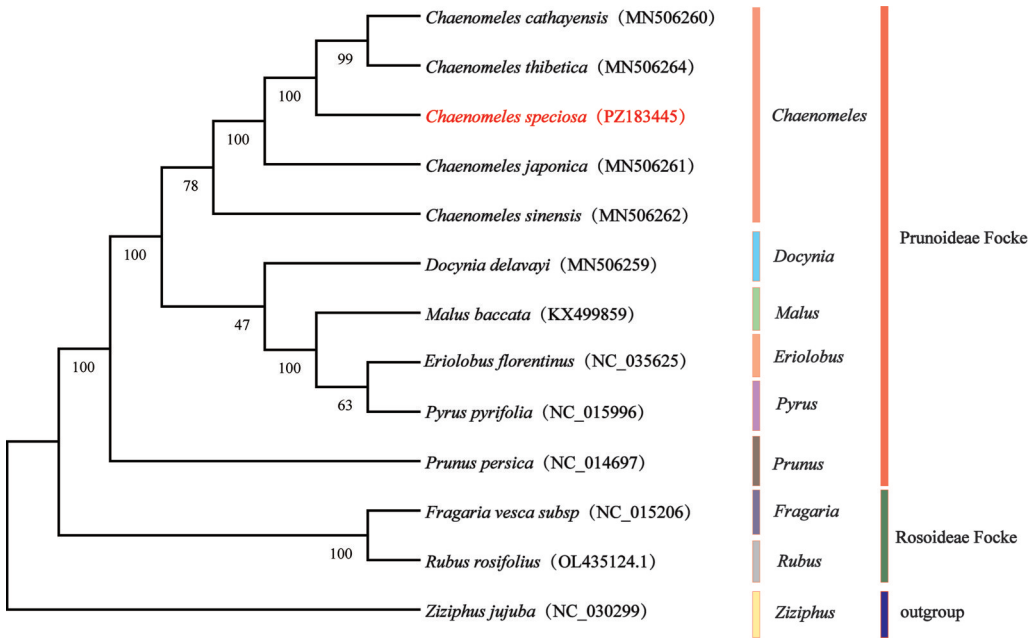


Figure 7 Phylogenetic tree based on common genes of the 13 complete chloroplast genomes using maximum likelihood (ML)

物系统关系上, 与木瓜属植物亲缘关系较近的依次是云南移依 (*D. delavayi*)、山芥子 (*M. baccata*)、佛罗伦萨海棠 (*E. florentinus*)、沙梨 (*P. pyrifolia*)。

讨论

相比核基因组, 叶绿体基因组具有结构保守、无基因沉默且进化速率慢等优势[13]。目前木瓜属多个物种的叶绿体基因组已相继公布[17,18], 但作为中药木瓜正品之一的“宣木瓜”尚缺乏系统性研究。本研究完整解析的宣城木瓜叶绿体基因组全长 159 991 bp, 呈典型环状四分体结构, 总 GC 含量 36.5% (其中 IR 区高达 42.6%), 这与蔷薇科苹果族植物的普遍特征高度一致[27], 证实了木瓜属叶绿体基因组在结构与碱基组成上的高度保守性。

SSR 标记是物种鉴定的理想分子标记[28]。研究表明, 不同植物类群的叶绿体 SSR 普遍以单核苷酸重复为主导且具有 A/T 偏好性, 且三核苷酸重复通常较少, 一般仅为少量存在[29,30], 甚至在部分植物[31]中未被检测到。如太白贝母[32]和乌塌菜[33]中均以单核苷酸重复占比最高, 且碱基组成以 A 和 T 为主。本研究在宣城木瓜中共检测到 76 个 SSR 位点, 单核苷酸重复占绝对优势且 A/T 碱基占比高达 97.2%, 进一步证实了该特征在单子叶与双子叶植物中的普遍性。值得注意的是, 本研究未发现三核苷酸重复, 可能与物种间基因组差异、样本数量或检测参数有关, 但其是否具有物种或属水平的特异性仍有待更多数据进一步验证。此外, 密码子使用偏好性分析同样表明, 偏好密码子多以 A/U 结尾, 与现有研究结果[34,35]相符, 反映出木瓜属叶绿体基因组在碱基组成和遗传密码使用上的保守性。

针对中药材市场宣木瓜常与同属伪品混淆且传统显微鉴别困难的现状[36], 叶绿体全基因组的滑动窗口分析和多重比对是筛选物种特异性 DNA 条形码的高效策略[37]。结果显示变异主要富集于 LSC 区。经严格筛选 (Pi > 0.015), 成功定位了 *trnR-atpA*、*petN-psbM*、*trnT-GGU-psbD* 及 *trnV-M* 四个高变异区。其中, *petN-psbM* 区域显示出最高的核苷酸多样性, 而 *trnT-GGU-psbD* 区域虽然长度较短, 但同样具有较高的变异水平。上述 4 个高变区具备开发为专属分子条形码的潜力, 为宣城木瓜及其混伪品的精准鉴定提供了可靠工具。

IR 区域是叶绿体基因组中相对保守的区域, IR 区的扩张与收缩常驱动基因组大小变异[38], Li 等[39]在对河北梨及苹果属等近缘属的比较中, 发现苹果族展现出极强的演化保守性。本研究证实 5 种木瓜属植物 IR 边界呈极高的属内保守性, 长度波动仅 36 bp, 宣城木瓜与西藏木瓜、毛叶木瓜的 IR 长度完全一致。这表明木瓜属基因组大小差异主要源于单拷贝区序列变异。

本研究基于叶绿体基因组构建的系统发育树, 揭示了木瓜属内部较为清晰的亲缘关系。已有研究[40,41]基于光皮木瓜的叶绿体 DNA 和核糖体 ITS 构建系统发育树, 得到不同聚类结果, 提出光皮木瓜可能具有木瓜属与榅桲属的杂交起源, 并支持将其划分为独立属 *Pseudocydonia*。本研究结果显示木瓜属 5 个物种整体聚为一支且支持率较高, 表明该属在叶绿体基因组水平上具有一定的亲缘一致性。其中, 毛叶木瓜与西藏

木瓜亲缘关系最近, 光皮木瓜位于该分支的基部。该结果与杂交假说中木瓜属作为母本的推测相一致。本研究的结果为进一步探讨木瓜属及其近缘类群的演化关系和光皮木瓜的分类提供了新的分子依据。

作者贡献: 汪思嘉参与数据收集、数据分析、文章初稿撰写和修订; 余函纹、戚康茹、赵志强、方清影参与数据收集、讨论和文章修订; 查良平、梁娟参与研究方案设计、数据分析指导和文章修订。

利益冲突: 所有作者声明本论文不存在利益冲突。

References

[1] Ren AY, Peng HS. Textual research on ancient geo-authentic producing areas of *Chaenomeles speciosa* based on literature and historical records from materia medica and local chronicles [J]. China J Chin Mater Med (中国中药杂志), 2024, 49: 5388-5396.

[2] Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China (中华人民共和国药典) [S]. 2020Ed. Part 1. Beijing: China Medical Science Press, 2020: 62.

[3] Wang ST, Xie XM. Research progress on the active components and pharmacological effects of *Chaenomeles speciosa* [J]. Clin J Tradit Chin Med (中医药临床杂志), 2014, 26: 320-321.

[4] Yue HF, Wang YZ, Zhang N, et al. Recent progress in cultivation and breeding techniques and comprehensive utilization of *Chaenomeles* spp. in China [J]. World For Res (世界林业研究), 2020, 33: 88-93.

[5] Gao X, Zheng RB, Yan YX, et al. Process optimization of *Chaenomeles speciosa* leaf beverage and investigation of its shelf life and antioxidant activity [J]. J Food Saf Qual (食品安全质量检测学报), 2026, 17: 86-99.

[6] Editorial Committee of Flora of China, Chinese Academy of Sciences. Flora of China (中国植物志) [M]. Beijing: Science Press, 1974, 36: 348-354.

[7] Zheng X, Wang H, Zhang P, et al. Chemical composition, antioxidant activity and *α* -glucosidase inhibitory activity of *Chaenomeles speciosa* from four production areas in China [J]. Molecules, 2018, 23: 2518.

[8] Bartish I, Rumpunen K, Nybom H. Genetic diversity in *Chaenomeles* (Rosaceae) revealed by RAPD analysis [J]. Plant Syst Evol, 1999, 214: 131-145.

[9] Lv ZF, Wu HM, Xi WP, et al. Molecular identification of *Chaenomeles* and its closely related genera using ITS2 sequence fragments [J]. Acta Hortic Sin (园艺学报), 2015, 42: 989-996.

[10] He S, Weng D, Zhang Y, et al. A telomere-to-telomere reference genome provides genetic insight into the pentacyclic triterpenoid biosynthesis in *Chaenomeles speciosa* [J]. Hortic Res, 2023, 10: uhad183.

[11] Rani V, Chauhan C, Sengar RS. DNA barcoding markers: A comprehensive review and taxonomic classification across species [J]. Comput Biol Chem, 2026, 122: 108872.

[12] Kong W, Wang Y, Zhang S, et al. Recent advances in assembly of complex plant genomes [J]. Genomics Proteomics Bioinf, 2023, 21: 427-439.

[13] Daniell H, Lin CS, Yu M, et al. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. Genome Biol, 2016, 17: 134.

[14] Dou XM, Zhang JH, Wang X, et al. DNA barcode screening, identification of germplasm resources, and analysis of genetic diversity of *Anemarrhena asphodeloides* [J]. China J Chin Mater Med (中国中药杂志), 2024, 49: 88-99.

[15] Chen JH, Cheng WP, Gao J, et al. Molecular identification of medicinal *Polygonatum* species based on plastid divergence hotspot regions [J]. Acta Pharm Sin (药学学报), 2025, 60: 1543-1554.

[16] Li RJ, Wu LW, Xin TY, et al. Analysis of chloroplast genomes and development of specific DNA barcodes for identifying the original species of Rhei Radix et Rhizoma [J]. Acta Pharm Sin (药学学报), 2022, 57: 1495-1505.

[17] Zhang SD, Jin JJ, Chen SY, et al. Diversification of Rosaceae since the late cretaceous based on plastid phylogenomics [J]. New Phytol, 2017, 214: 1355-1367.

[18] Sun J, Wang Y, Liu Y, et al. Evolutionary and phylogenetic aspects of the chloroplast genome of *Chaenomeles* species [J]. Sci Rep, 2020, 10: 11466.

[19] Shi L, Chen H, Jiang M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. Nucleic Acids Res, 2019, 47: 65-73.

[20] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes [J]. Nucleic Acids Res, 2019, 47: 59-64.

[21] Kurtz S, Choudhuri JV, Ohlebusch E, et al. REPuter: The manifold applications of repeat analysis on a genomic scale [J]. Nucleic Acids Res, 2001, 29: 4633-4642.

[22] Huang L, Yu H, Wang Z, et al. CPStools: A package for analyzing chloroplast genome sequences [J]. iMetaOmics, 2024, 1: e25.

[23] Amiryousefi A, Hyvönen J, Poczai P. IRscope: An online program to visualize the junction sites of chloroplast genomes [J]. Bioinformatics, 2018, 34: 3030-3031.

[24] Frazer KA, Pachter L, Poliakov A, et al. VISTA: Computational tools for comparative genomics [J]. Nucleic Acids Res, 2004, 32: 273-279.

[25] Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets [J]. Mol Biol Evol, 2016, 33: 1870-1874.

[26] Ivica L, Peer B. Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool [J]. Nucleic Acids Res, 2024, 52: 78-82.

[27] Sun J, Zhao D, Qiao P, et al. Phylogeny of genera in Maleae

(Rosaceae) based on chloroplast genome analysis [J]. Front Plant Sci, 2024, 15: 1367645.

[28] Li X, Hu Z, Yu W, et al. Advances and challenges in plant molecular marker technologies and their applications in the artificial intelligence empowered era [J]. Front Plant Sci, 2026, 16: 1757949.

[29] George B, Bhatt BS, Awasthi M, et al. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants [J]. Curr Genet, 2015, 61: 665-677.

[30] Qin Q, Dong Y, Chen J, et al. Comparative analysis of chloroplast genomes reveals molecular evolution and phylogenetic relationships within the Papilionoideae of Fabaceae [J]. BMC Plant Biol, 2025, 25: 157.

[31] Du Q, Wang LQ, Chen ZE, et al. Characterization and phylogenetic analysis of the complete chloroplast genome of *Lycopus europaeus* [J]. Acta Pharm Sin (药学学报), 2022, 57: 2206-2215.

[32] Zhu L, Tong XY, Wang Q, et al. Complete chloroplast genome sequence characteristics and phylogenetic analysis of *Fritillaria taipaiensis* [J]. Chin Tradit Herb Drugs (中草药), 2025, 56: 7157-7167.

[33] Wang YL, Zhu ZX, Zhang HY, et al. Characterisation and phylogeny of the chloroplast genome of Wucai (*Brassica campestir*) [J]. Acta Hortic Sin (园艺学报), 2025, 52: 1326-1340.

[34] Fang ZX, Ji Q, Hu JD, et al. Characterization and phylogenetic analysis of the complete chloroplast genome of *Salvia apiana* Jepson [J]. Acta Pharm Sin (药学学报), 2024, 59: 1484-1493.

[35] Ji KK, Song XQ, Chen CG, et al. Codon usage profiling of chloroplast genome in Magnoliaceae [J]. J Agric Sci Technol (中国农业科技导报), 2020, 22: 52-62.

[36] Zhang YL, Li ZJ, Zhu LL, et al. Identification of prepared slices of Chaenomelis Fructus and adulterants based on HPLC fingerprint [J]. Res Pract Chin Med (现代中药研究与实践), 2020, 34: 41-45.

[37] Zhang M, Zhai X, He L, et al. Morphological description and DNA barcoding research of nine *Syringa* species [J]. Front Genet, 2025, 16: 1544062.

[38] López-Caamal A, Gandee T, Galloway LF, et al. Substantial structural variation and repetitive DNA content contribute to intraspecific plastid genome evolution [J]. BMC Genomics, 2025, 26: 340.

[39] Li Y, Zhang J, Li L, et al. Structural and comparative analysis of the complete chloroplast genome of *Pyrus hopeiensis* and three other *Pyrus* species [J]. Int J Mol Sci, 2018, 19: 3262.

[40] Campbell C, Evans R, Morgan D, et al. Phylogeny of subtribe Pyrinae (formerly the Maloideae, Rosaceae): Limited resolution of a complex evolutionary history [J]. Plant Syst Evol, 2007, 266: 119-145.

[41] Lo EY, Donoghue MJ. Expanded phylogenetic and dating analyses of the apples and their relatives (Pyreae, Rosaceae) [J]. Mol Biol Evol, 2012, 63: 230-243.