生物技术通报
· 研究报告 · BIOTECHNOLOGY BULLETIN 2026, 42(10):216-228

九种金丝桃属植物叶绿体全基因组:基因组结构、比较分析和系统发育关系

范照睿[1#] 王佳敏[2#] 童凯[1*] 康振辉[1*]
(1.四川轻化工大学食品与酿酒工程学院,宜宾 644000; 2.兰州大学生态学院,兰州 730000)

摘 要: 【目的】分析川滇金丝桃(*Hypericum forrestii*)在内的金丝桃属物种叶绿体基因组结构、基因组成及系统发育关系,为物种鉴定和进化研究提供基因组学基础。【方法】采用高通量测序技术对川滇金丝桃叶绿体基因组进行测序与组装,并结合NCBI数据库获取的8种金丝桃属已发表叶绿体基因组数据,共计9个物种开展比较基因组学分析。利用相关生物信息学工具对叶绿体基因组进行基因注释,并进一步开展简单重复序列(SSR)分析、密码子使用偏好分析、IR边界结构比较以及核苷酸多样性分析,同时基于完整叶绿体基因组序列构建系统发育树。【结果】对9种金丝桃属植物的叶绿体基因组进行分析,结果表明,所有物种的基因组结构均呈典型的四分体环状结构,基因数在110-130个之间,符合已报道的金丝桃属叶绿体基因组特征。基因组长度范围为122 570-168 150 bp,GC含量为37.04%-38.14%。SSR分析表明,单核苷酸重复类型占主导,且以A/T碱基为主。密码子使用分析表明,A/U结尾的密码子使用偏好性明显。核苷酸多样性分析显示,LSC和SSC区域为主要的变异热点,尤其是*trnH-psbA*、*rpoC2-rpoC1*、*ndhF-rpl32*等基因间区。系统发育分析表明,金丝桃属为单系群,支持与川苔草科的亲缘关系。【结论】9种金丝桃属植物叶绿体基因组均呈典型四分体结构,IR区的动态变化是导致基因组长度差异的主要因素,LSC和SSC区的高变异热点区域可作为潜在的分子标记。系统发育分析支持金丝桃属为单系群,并与川苔草科具有较近亲缘关系。

关键词: 川滇金丝桃;叶绿体基因组;系统发育分析;比较基因组学;简单重复序列(SSR);核苷酸多样性;IR边界
DOI: 10.13560/j.cnki.biotech.bull.1985.2025-1253

Chloroplast Whole Genomes of Nine *Hypericum* Species: Genome Structure, Comparative Analysis, and Phylogenetic Relationships

FAN Zhao-rui[1#] WANG Jia-min[2#] TONG Kai[1*] KANG Zhen-hui[1*]
(1. *School of Food and Brewing Engineering, Sichuan University of Science & Engineering, Yibin* 644000; 2. *College of Ecology, Lanzhou University, Lanzhou* 730000)

Abstract: 【Objective】This study aimed to analyze the chloroplast genome structure, gene composition, and phylogenetic relationships of *Hypericum forrestii* and related *Hypericum* species, providing a genomic foundation for species identification and evolutionary research. 【Method】The chloroplast genome of *H. forrestii* was sequenced and assembled using high-throughput sequencing technology. Together with sequences of eight published *Hypericum* species retrieved from the NCBI database, a total of nine chloroplast genomes were analyzed. Comparative genomic analyses were performed, including genome annotation, simple sequence repeat (SSR) identification, codon usage bias analysis, inverted repeat (IR) boundary comparison, nucleotide diversity analysis, and phylogenetic reconstruction based on complete chloroplast genome sequences. 【Result】The chloroplast genomes of the nine *Hypericum* species exhibited a typical quadripartite circular structure, with gene numbers ranging from 110 to 130, consistent with previously reported features in the genus. Genome sizes ranged from 122 570 to 168 150 bp, with GC content of 37.04%-38.14%. SSR analysis showed that single nucleotide repeat were dominant, and

收稿日期: 2025-11-19
基金项目: 国家自然科学基金项目(31601291),四川省科技计划重点项目(2020YFN0023)
*通信作者: 康振辉 zhkang85@126.com,童凯 a1149762188@gmail.com
#同等贡献

mononucleotide A/T repeats were the most abundant repeat type. Codon usage analysis showed a clear preference for A/U-ending codons. Nucleotide diversity analysis indicated that the LSC and SSC regions were the primary variation hotspots, particularly the intergenic spacers *trnH-psbA*, *rpoC2-rpoC1*, and *ndhF-rpl32*. Phylogenetic analysis supported the monophyly of *Hypericum* and its close relationship with Podostemaceae.【Conclusion】 The chloroplast genomes of nine *Hypericum* species all exhibit a typical quadripartite structure. Dynamic changes in the IR regions are the main mechanism responsible for genome size variation, while highly variable hotspots in the LSC and SSC regions serve as promising molecular markers. Phylogenetic analysis supports the monophyly of *Hypericum* and its close relationship with Podostemaceae.

Keywords: *Hypericum forrestii*; chloroplast genome; phylogenetic analysis; comparative genomics; simple sequence repeats (SSRs); nucleotide diversity; IR boundary

金丝桃属（*Hypericum*）是金丝桃科中最大的属之一，约包含500个物种，主要分布于欧亚大陆和南美安第斯山区[1]。许多金丝桃属植物具有重要的药用价值，例如贯叶连翘（*H. perforatum*），又名圣约翰草，作为著名的传统草药，被广泛用于治疗轻、中度抑郁症和抗炎症治疗等[2-3]。川滇金丝桃（*H. forrestii*）作为金丝桃属中中国西南高山特有物种，具有潜在抗抑郁和抗氧化药用价值，且此前未有完整叶绿体基因组报道。鉴于该属物种多样性及其重要的药用和生态价值，基于基因组的系统研究对于明确物种界限、揭示亲缘关系和开发药用资源的分子标记具有重要意义。

叶绿体基因组在结构和基因含量上通常较为保守，呈现典型的四元循环结构，包含1个大单拷贝区（large single-copy region, LSC）、1个小单拷贝区（small single-copy region, SSC）和2段倒位重复区（inverted repeat, IR），基因组大小一般在120–160 kb之间[4]。与核基因组和线粒体基因组相比，叶绿体基因组的基因组组成、结构及基因数目相对稳定，因此常被用作植物系统发育研究的分子标记。

在叶绿体基因组比较研究中，简单序列重复（SSR）和其他长重复序列常被用作遗传标记和进化研究的工具。先前对金丝桃属和相关植物的研究表明，叶绿体SSR多集中在LSC区，以单核苷酸重复为主[5]。此外，密码子使用偏好可以反映基因组的演化特征。研究发现*Hypericum*叶绿体编码区偏好使用以A或U（DNA上对应A/T）结尾的密码子。叶绿体基因组的大小变化往往与IR区边界的扩张或收缩密切相关，被认为是基因组大小差异的主要机制。此外，叶绿体基因组长度的物种间差异主要由倒位重复区（IR）边界的扩张或收缩造成，IR的动态变化会导致基因拷贝数和边界邻近基因（如*rps19*、*ycf1*等）位置的重排或丧失，因此比较IR边界为解析基因组结构演化提供了直接证据[6]。近年来的方法学进展（例如IR检测与可视化工具）也使得对这些边界事件的定量比较更为可靠。比较基因组学和滑动窗口分析在多个*Hypericum*和相关研究中反复识别出若干高变热点区域，这些高变区可作为潜在的DNA条码或群体遗传学标记，便于物种鉴定与群体水平研究的开展[7]。叶绿体基因组的系统发育研究或许可为*Hypericum*属域内支系划分提供更高分辨率的依据，正如近期有研究表明通过增加物种样本量更能支持科内的部族关系与属内近缘关系的划分，而仅利用核基因组数据检验叶绿体与核基因组间的核质冲突与基因渐渗时可能存在谱系不一致或杂交信号[8-9]。

本研究以9种金丝桃属植物为对象（其中*H. forrestii*为首次测序物种），结合前人公开的叶绿体基因组数据，对其基因组结构特征进行系统比较分析。利用高通量测序技术对*H. forrestii*的叶绿体基因组进行组装与注释，分析叶绿体基因组中SSR分布和其他长重复序列特征，研究基因组的密码子使用偏好及IR区边界的扩张/收缩情况，并开展多序列比对和系统发育分析，以揭示物种间的亲缘关系。本研究旨在通过揭示金丝桃属叶绿体基因组的结构多样性及物种间演化关系，为金丝桃属物种的鉴定和保护提供新的分子标记和演化信息。

1　材料与方法

1.1　材料

川滇金丝桃（*H. forrestii*）的新鲜叶片采自四川省乐山市（29°32′N，103°21′E）。凭证标本（编号：ghao220810）保存于峨眉山标本馆。植物标本的采

集符合《中华人民共和国野生植物保护条例》的相关规定。总基因组DNA采用改良CTAB法提取[10]。基因组测序由中国甘肃紫沐生物科技有限公司在Illumina HiSeq平台上完成，产生约5.77 Gb原始数据。此外，从NCBI数据库下载其余8种金丝桃属物种（*H. acmosepalum*、*H. addingtonii*、*H. ascyron*、*H. beanii*、*H. monogynum*、*H. perforatum*、*H. petiolulatum*和*H. sampsonii*）的完整叶绿体基因组序列，用于后续比较分析。

1.2 方法

1.2.1 叶绿体基因组组装与注释 原始测序数据经Trimmomatic v0.36进行质量控制[11]，使用默认参数。修剪后的高质量读段采用SPAdes v3.14.1进行组装[12]。组装结果通过CPGAVAS2进行基因注释[13]，并在Notepad++及Geneious软件中进行人工校正[14]，以获得最终注释文件。环状叶绿体基因组图谱通过OGDRAW（http://ogdraw.mpimp-golm.mpg.de）生成，采用默认参数，以确定倒置重复（IR）区与大单拷贝（LSC）、小单拷贝（SSC）区的相对位置[15]。

1.2.2 简单重复序列与分散重复序列分析 使用MISA分析9种金丝桃属植物叶绿体基因组中的简单重复序列（SSR）[16]，确定SSR的类型、数量与分布位置。分析参数设置如下：单核苷酸重复≥10，二核苷酸≥6，三至六核苷酸均≥5，SSR之间的最大间距为100 bp。长重复序列（包括回文、正向、反向及互补重复）通过REPuter检测，参数设定为Hamming距离3，最小重复长度30 bp[17]。

1.2.3 密码子使用情况分析 利用cpstools（https://github.com/Xwb7533/CPStools）从GenBank文件中提取9种金丝桃属植物的CDS序列，去除重复项并筛除长度小于300 bp及非ATG起始的序列[18]。使用RSCUcaller（https://github.com/Mordziarz/RSCUcaller）计算相对同义密码子使用率（relative synonymous codon usage, RSCU）、密码子组统计及氨基酸组成特征，并对结果进行可视化分析[19]。

1.2.4 基因组比较与结构分析 通过IRscope（https://irscope.shinyapps.io/irapp/）生成IR边界比较图[20]，分析不同物种间4个边界区域（LSC-IRa、IRa-SSC、SSC-IRb、IRb-LSC）及其相邻基因的相对位置变化，从而探讨金丝桃属叶绿体基因组结构演化特征。使用mVISTA（https://genome.lbl.gov/vista/mvista/submit.shtml）在Shuffle-LAGAN模式下对叶绿体基因组进行全局比对，其他参数为默认设置，用以检测基因重排与倒位现象[21]。

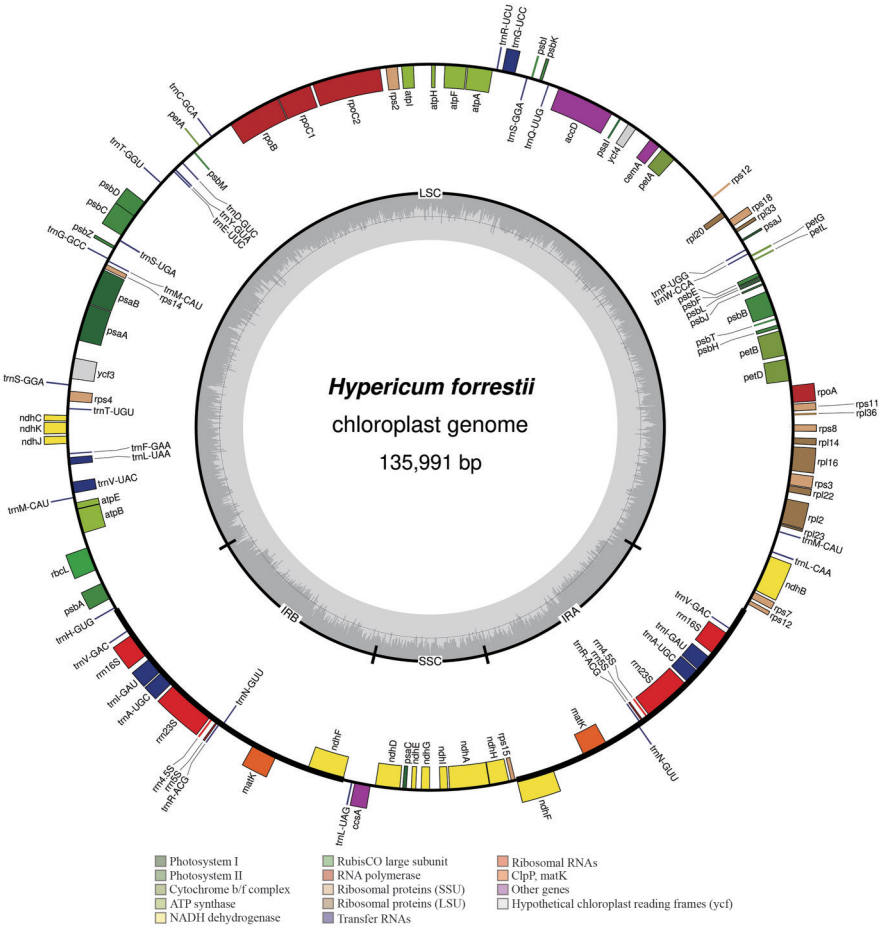
1.2.5 系统发育与核苷酸多态性分析 利用PhyloSuite v1.2.3从9种金丝桃属植物及外群物种的叶绿体基因组中提取48个共有蛋白编码基因（PCGs）[22]。序列经MAFFT v7.4比对并拼接[23]，随后使用Gblocks v0.91b提取保守区段[24]。最大似然（ML）系统发育树通过IQ-TREE v2.1.2构建，模型由ModelFinder自动选择（GTR+F+I+G4），并进行5 000次超快速自举验证[25]。系统发育树在iTOL平台（https://itol.embl.de/）进行可视化与美化[26]。核苷酸多态性分析使用DnaSP v6.12完成，并将结果绘制为折线图以展示变异分布特征[27]。

2 结果

2.1 金丝桃属物种叶绿体基因组的一般特征

9种*Hypericum*属植物的完整叶绿体基因组均呈典型的四分体环状结构（附表1），包含大单拷贝区（LSC）、小单拷贝区（SSC）和1对反向重复区（IR）（图1，附图1）。基因组长度范围为122 570–168 150 bp。其中*H. forrestii*的基因组总长度为135 991 bp，由LSC区（90 710 bp）、SSC区（10 425 bp）和2段IR区（各34 856 bp）组成。主要由于SSC区和IR区的缩短，*H. sampsonii*和*H. ascyron*的基因组较长（168 150 bp和162 286 bp），主要归因于IR区扩增。*H. forrestii*的基因组总长度为135 991 bp，由LSC区（90 710 bp）、SSC区（10 425 bp）和2段IR区（各34 856 bp）组成。各物种整体GC含量变化较小，为37.04%–38.14%，但IR区GC含量相对较高或波动较大，与该区域富含rRNA基因及重复序列有关。整体GC含量在9个物种中变化较小，为37.04%–38.14%。在不同功能区中，GC含量分布存在一定差异，LSC区和SSC区相对较低，而IR区普遍较高或波动较大。

注释结果显示（附表2），*H. forrestii*叶绿体基因组共鉴定115个基因，其中包括73个蛋白编码基因（PCG）、34个tRNA基因和8个rRNA基因。部分基因在IR区中发生重复，且存在顺式剪接和反式剪接基因，表明其基因结构具有一定复杂性（附图2）。



图中展示了4个基因组区域(大单拷贝区LSC、小单拷贝区SSC、反向重复区IRA和反向重复区IRB)。圆内的基因顺时针转录,圆外的基因逆时针转录。不同功能类别的基因用不同颜色标记。内圈显示GC含量(深灰色)和AT含量(浅灰色)

The figure shows four genomic regions: the Large Single-Copy region (LSC), the Small Single-Copy region (SSC), and the two Inverted Repeat regions (IRA and IRB). Genes located inside the circle are transcribed clockwise, while those outside are transcribed counter clockwise. Genes belonging to different functional categories are marked in different colors. The inner circle indicates the GC content (dark gray) and AT content (light gray)

图1　川滇金丝桃叶绿体基因组图谱

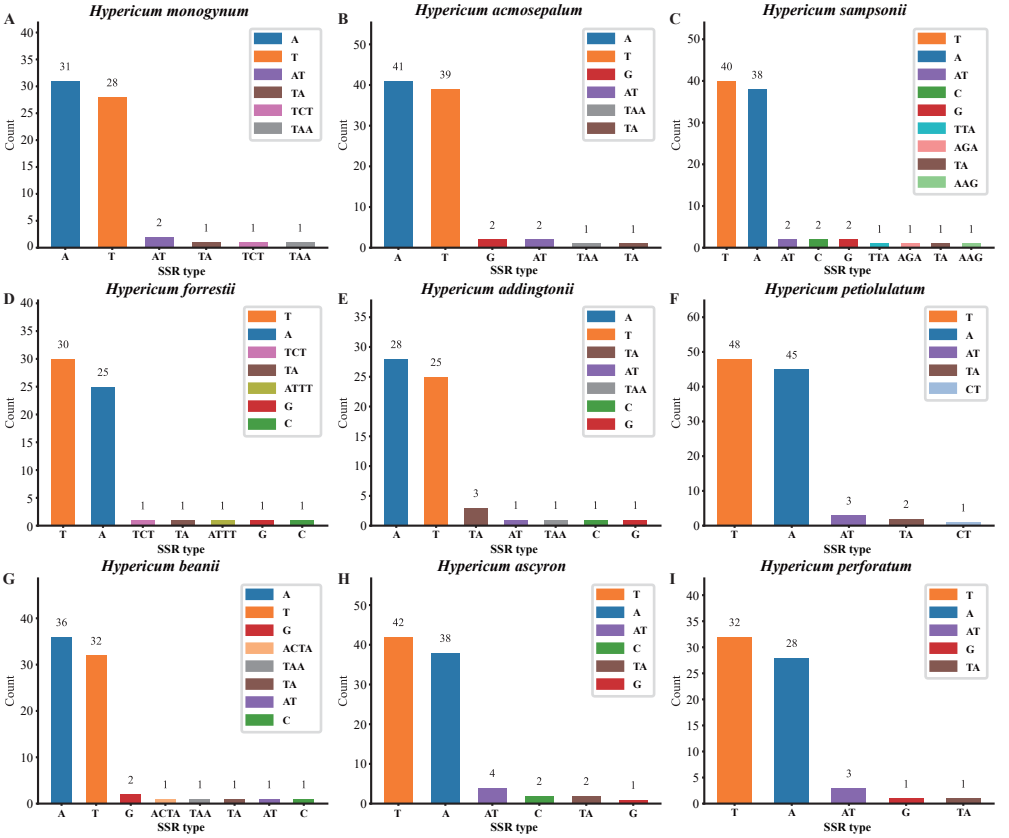
Fig. 1　Chloroplast genome map of *Hypericum forrestii*

从9个物种的整体比较结果来看，基因数在112–127个之间，蛋白编码基因为72–86个，tRNA为29–37个，rRNA为6–8个。这些基因主要参与光合作用、自我复制及蛋白合成等功能，显示出较高的保守性。

2.2　九种金丝桃属物种叶绿体基因组中的重复序列

针对9种基因组的SSR分析显示，在所有样品中均检测到单核苷酸、二核苷酸及少量三核苷酸的简单重复序列，主要分布在LSC区，其中单核苷酸重复为主要类型且数量显著高于其他类型（图2，图3，附表3）。A/T型单核苷酸重复在所有物种中最为丰富，体现出显著的A/T偏好性。SSR总数在物种间存在差异，*H. petiolulatum*的SSR数量最高，A与T单核分别达到48和45个；*H. addingtonii*和*H. monogynum*的SSR数量相对较少（约60个左右）。二核与三核类型（例如AT、TA、TAA、TCT等）普遍较少，占总体SSR的比例小，三核及以上的重复则较为罕见。

SSR长度主要集中在10–15 bp，其中10–12 bp最为常见，占总SSR的大部分比例，显示出较高的种间变异性。高变异基因间区（如*trnH-psbA*、*ndhF-*



A–I: 在不同重复类型中识别出的SSR基序的频率

A–I: Frequency of SSR motifs identified in different repeat types

图2 9种金丝桃属植物叶绿体基因组的SSR位点分析

Fig. 2 Analysis of SSR loci in the chloroplast genomes of nine *Hypericum* species

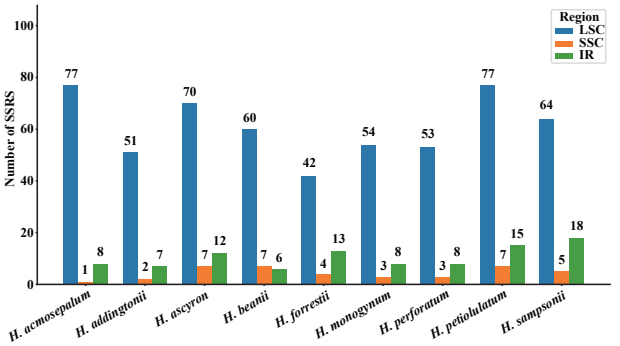
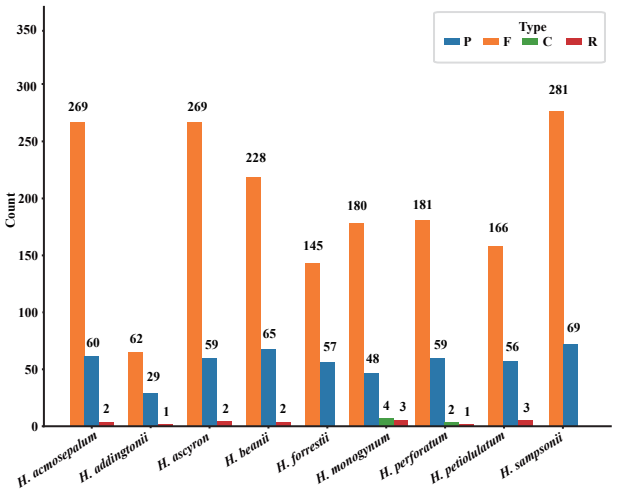


图3 SSR位点在LSC、SSC和IR区域内的重复类型频率

Fig. 3 Frequency of SSR repeat types in the LSC, SSC, and IR regions

rpl32）SSR长度和密度的差异尤为明显，提示其具有较高的遗传分辨力，可作为潜在的分子标记资源（附表4）。

对9种金丝桃属植物叶绿体基因组中散在重复序列进行计数，共鉴定出2 303对重复序列（图4和附表5），具体为：14对反向重复序列（R），502对回文重复序列（P），1 781对正向重复序列（F），6对互补重复序列（C）。重复序列长度主要集中在30–49 bp，短重复（30–39 bp）最为丰富，而较长重复（>49 bp）相对稀少，但在*H. sampsonii*等物种中仍有明显积累。重复序列主要分布于基因间区，少数嵌入蛋白编码基因，如*ycf2*、*rps16*等。不同物种在重复序列的数量、类型及长度分布上存在差异，这可能反映了叶绿体基因组在演化过程中不同物种的结构多样性和物种特异性特征。总体而言，SSR与散在重复序列的分布模式不仅揭示了基因组保守性与变异热点，也为未来*Hypericum*属的系统发育研究和分子标记开发提供了重要信息。



图中蓝色条状P代表回文重复序列,橙色条状F代表正重复序列,绿色条状C代表互补重复序列,红色条状R代表反向重复序列

The blue bars labeled P represent palindromic repeats, the orange bars labeled F represent forward repeats, the green bars labeled C represent complementary repeats, and the red bars labeled R represent reverse repeats

图4　9种金丝桃属植物的长重复序列数和类型

Fig. 4　Number and types of long repeat sequences in nine *Hypericum* species

2.3　密码子使用偏好

对9种金丝桃属物种蛋白编码基因进行相对同义密码子使用率（RSCU）分析，结果显示所有样本均使用64个密码子，编码19种氨基酸。Leu、Arg与Ser的密码子类型最多，均为6个，Met与Trp各由1个密码子编码。各物种的RSCU柱状图整体分布趋势相似。多数密码子的RSCU值在0.5–1.5之间（图5-A），表明金丝桃属植物叶绿体基因组中密码子使用较为稳定。其中，以A或U结尾的密码子偏好性明显高于以G或C结尾的类型，如UUA（编码Leu）、AUA（编码Ile）、GAA（编码Glu）及AAA（编码Lys）等的RSCU值均显著高于1.0，显示出显著的A/U偏好性。相反，G/C结尾的密码子（如CCG、GGC、CGC等）RSCU值均低于1.0，使用频率相对较低。

热图进一步揭示了9种金丝桃属植物间的密码子使用差异（图5-B）。颜色变化反映了RSCU值的高低，其中蓝色表示高频密码子（RSCU>1），红色表示低频密码子（RSCU<1）。值得注意的是，亮氨酸的密码子TTA在所有物种中具有最高的RSCU值（约2.0），而同样编码亮氨酸的CTG的RSCU值最低（约0.3），表明金丝桃属植物对以A/U结尾的亮氨酸密码子具有显著偏好。少数物种间差异如*H. petiolulatum*和*H. ascyron*中部分A/U结尾的密码子偏好略强，而*H. monogynum*与*H. addingtonii*在少数G/C结尾的密码子上呈现相对均衡的使用倾向。

总体而言，金丝桃属叶绿体基因组的密码子使用模式高度保守，存在显著的A/U偏好，与大多数被子植物叶绿体基因组一致，同时显示出物种间有限的种内变异。

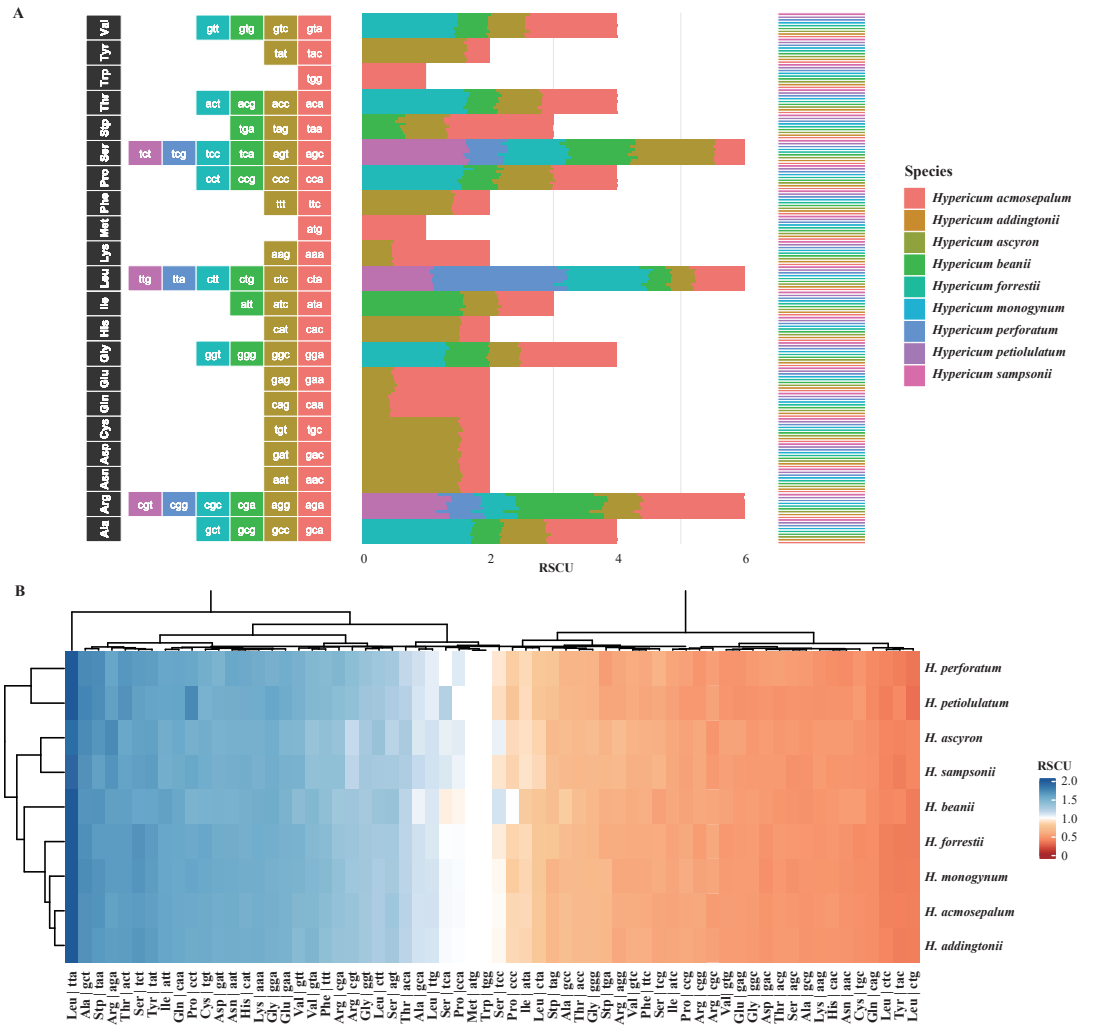
2.4　IR区的扩展与收缩

在对9种金丝桃属物种的叶绿体基因组比较中，比较了4个叶绿体边界区域，即LSC-IRb交界（JLB）、SSC-IRb交界（JSB）、SSC-IRa交界（JSA）和LSC-IRa交界（JLA），发现LSC-IRb边界普遍不与基因重叠（图6），仅*H. addingtonii*的LSC-IRb边界与*rrn5*基因相距25 bp。*H. ascyron*、*H. forrestii*、*H. perforatum*和*H. petiolulatum*的IRb-SSC边界均距*ndhF*基因26 bp；而在*H. addingtonii*中，IRb-SSC边界与*rrn16*基因相距17 bp。相比之下，其余金丝桃属物种在该边界处缺失*ndhF*。在SSC-IRa边界上，*H. ascyron*、*H. forrestii*与*H. monogynum*的*ndhF*基因均跨越该边界11 bp；*H. addingtonii*的*rrn16*基因跨越该边界16 bp；而*H. perforatum*与*H. petiolulatum*的*ndhF*基因则较近，均距该边界25 bp。

2.5　叶绿体基因组序列比较与核苷酸多样性

基于mVISTA的全局比对（Shuffle-LAGAN）显示，9种叶绿体基因组在结构与基因顺序上高度保守（图7），蛋白编码区（exon）和rRNA/tRNA区域的序列同源性较高，非编码区（CNS）出现较多波动与显著的低同源片段，是叶绿体基因组中主要的变异热点。且IR区比单拷贝区更为保守。比对结果揭示若干高变异热点，在*rpoC2-rpoC1*、*psaA-ycf3*、*ndhF-ccsA*等区域曲线下降明显，提示这些基因间区为金丝桃属叶绿体中较高变异区域。

核苷酸多样性（π）滑动窗口分析结果表明，9种金丝桃属植物叶绿体基因组整体变异水平较低（π≈0.000 5–0.002 5）（图8），表明该属植物的叶绿体基因组在种间层面上高度保守，变异主要集中在一些特定的非编码区或特定基因区段，例如*trnH-*



A:9种金丝桃属植物不同密码子RSCU柱状图;B:9种金丝桃属植物RSCU值的热图

A: Bar chart of RSCU for different codons across nine *Hypericum* species. B: Heatmap of RSCU values for nine *Hypericum* species

图5 金丝桃属植物的相对同义密码子使用度(RSCU)

Fig. 5 Relative synonymous codon usage (RSCU) in *Hypericum* species

psbA、*rpoC2-rpoC1*、*ndhF-rpl32*和*ycf1*等区域，这些区域的π值峰值明显，进一步确认了它们是潜在的变异热点。这些高变异区域通常位于LSC和SSC区域，而IR区则呈现出较低的变异性，符合被子植物叶绿体基因组的普遍规律，即IR区的序列通常高度保守。

在9种金丝桃属叶绿体基因组中，存在高变异的基因，如*ndhF*（π>0.005，间隔区变异高）、*rpoC2*（转录相关）和*ycf1*（假设蛋白，边界变异）。此外，存在PAV（presence-absence variation）变异显著的基因：tRNA基因数量在29–37个之间波动，*ndhF*在部分种（如*H. addingtonii*）中缺失或伪基因化，*ycf1*拷贝数变异源于IR扩张。这些快速进化的基因与特定功能相关：*ndhF*参与光合电子传递和抗逆应激，*rpoC2*调控叶绿体转录表达，*ycf1*涉及膜蛋白稳定性和蛋白导入，可能帮助植物适应高山或干旱环境。金丝桃属中的这些高变基因与其他植物已报道的叶绿体快速进化基因高度一致，例如在*Prunus*属中*ycf1-ndhF*为常见热点，在*Panax*属中*rpoC2*显示高变异率，在*ndh*家族中PAV变异普遍存在，进一步支持这些区域作为跨属分子标记的潜力。鉴于这些变异热点区域的高核苷酸多样性，它们具有作为分子

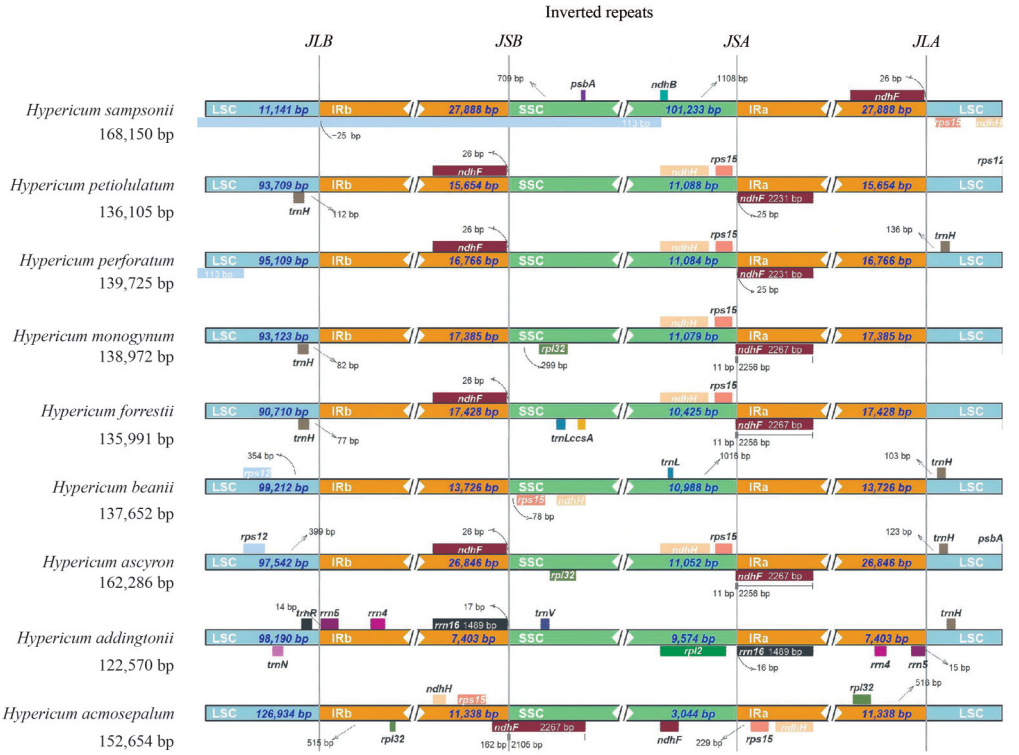


图6　9种金丝桃属叶绿体基因组中LSC、SSC和IR区域边界的比较

Fig. 6　Comparison of the boundaries of LSC, SSC, and IR regions in the chloroplast genomes of nine *Hypericum* species

标记的潜力，可以用于开发叶绿体DNA条形码或高分辨率的系统发育标记，进而在金丝桃属物种间进行细致的种间区分或类群辨识。

2.6　系统发育分析

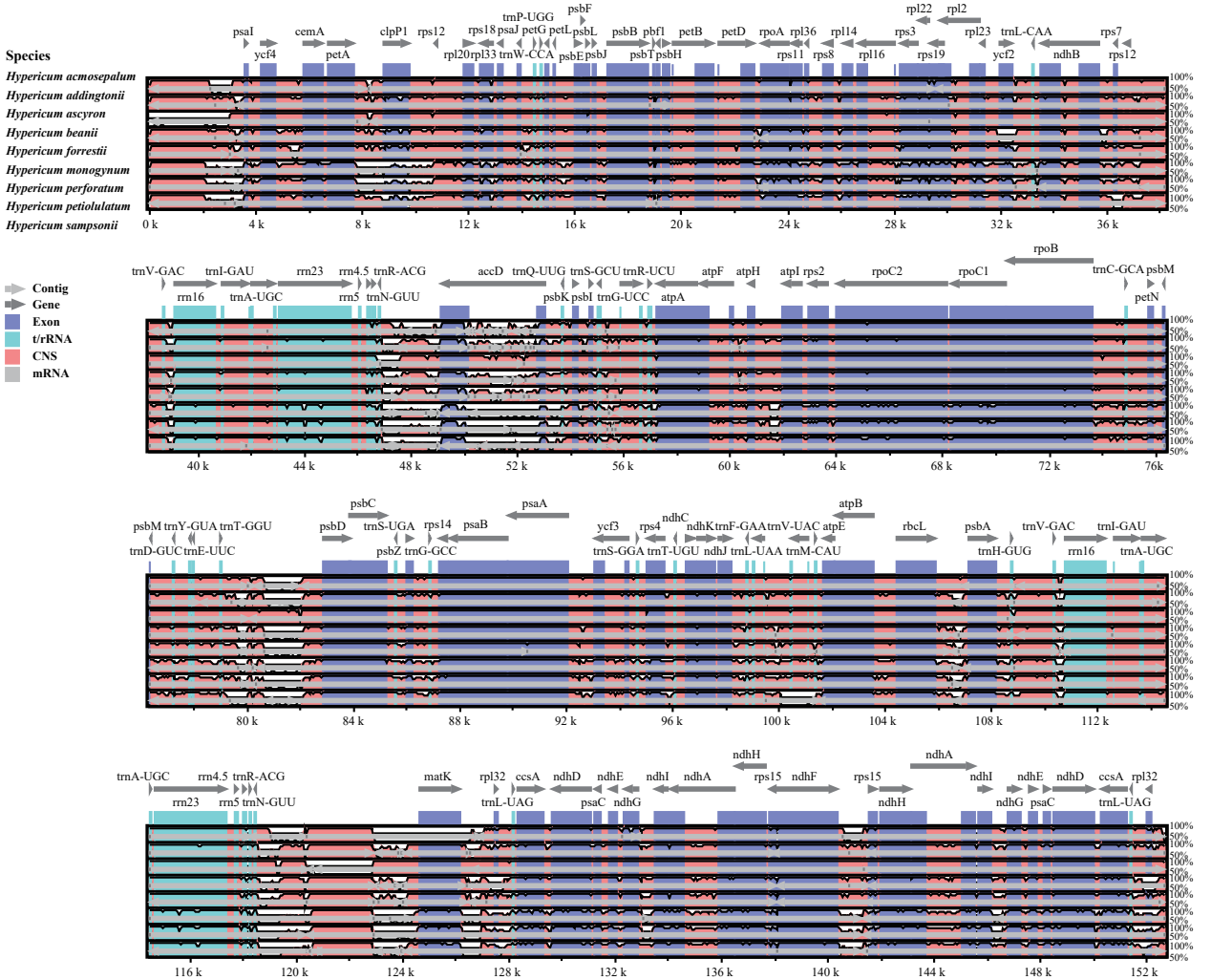
为了推断金丝桃科植物各物种与金虎尾目其他科之间的系统发育关系，基于48个共有蛋白编码基因（PCGs）构建最大似然树（图9）。生成的拓扑结构具有一致的系统发育关系且分辨率较高，其中金丝桃科（Hypericaceae）和Podostemaceae科形成一个紧密的进化枝，支持度极高（BS=100%和PP=1.00），而这一分支进一步与Calophyllaceae科聚类，具有强烈的统计支持（BS=100%和PP=1.00）。相比之下，这一簇与Clusiaceae科的亲缘关系较远，后者独立形成一个分支，支持度为100%。

在金丝桃科内部，该科分为2个主要进化枝——Cratoxyleae族和Hypericeae族，支持度很高（BS=100%）。Cratoxyleae族包含多个*Cratoxylum*种，包括*C. arborescens*、*C. cochinchinense*、*C. formosum*、*C. maingayi*和*C. sumatranum*，其中*C. arborescens*作为该族最早分化出的物种，具有高支持度（BS=100%），其余物种依次分化，形成一个单系群，分支长度较短表明遗传变异较低。

在*Hypericum*属内部，9种植物形成一个进化枝，其中*H. petiolatum*作为第一个分化出的物种，支持度为100%，随后是*H. perforatum*和*H. ascyron*的分化（支持度100%），*H. addingtonii*、*H. acmosepalum*、*H. beanii*形成子簇，其中*H. acmosepalum*和*H. beanii*亲缘最近（支持度100%），它们与*H. addingtonii*共享祖先（支持度100%），*H. forrestii*则与*H. monogynum*亲缘关系最为接近。

3　讨论

九种金丝桃属植物的叶绿体基因组均呈典型的四分体结构，包含大单拷贝区（LSC）、小单拷贝区（SSC）以及2对反向重复区（IR）。这一结构与大多数被子植物的叶绿体基因组一致，表明金丝桃属植物在叶绿体基因组的进化上保持了较高的保守性[28]。整体GC含量较为一致，表明其核苷酸组成



x轴代表核苷酸序列，y轴代表序列同源性百分比，范围50%–100%。曲线下方突出显示了保守区域，红色区域为“保守非编码序列”（CNS），深蓝色区域为外显子，浅蓝色区域为tRNA或rRNA

The x-axis represents the nucleotide sequence, and the y-axis represents the sequence homology percentage, ranging from 50% to 100%. Conserved regions are highlighted below the curve, with red indicating “Conserved Noncoding Sequences” (CNS), dark blue indicating exons, and light blue indicating tRNA or rRNA

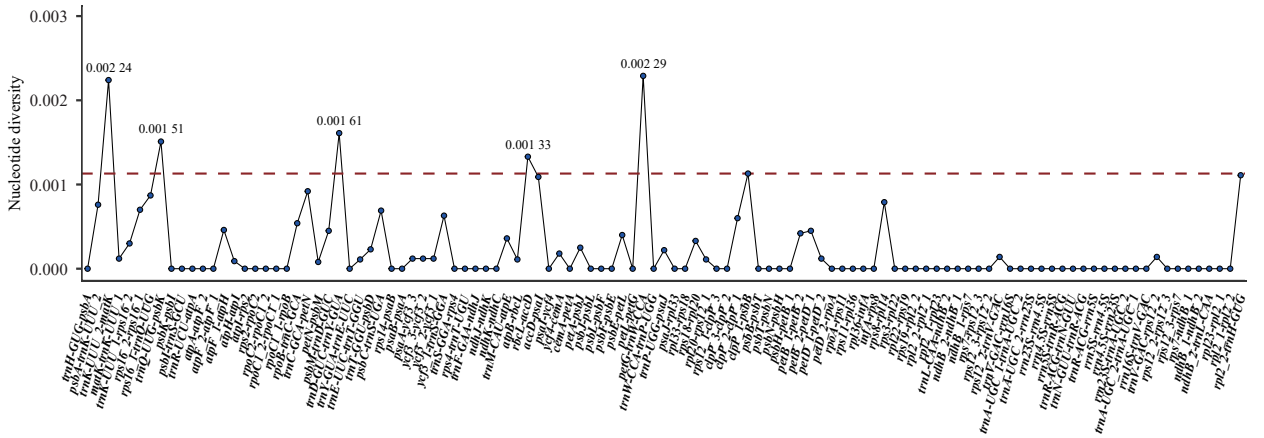
图7 9种金丝桃属叶绿体基因组序列特征比较图谱

Fig. 7 Comparative map of sequence features of the chloroplast genomes of nine *Hypericum* species

在进化过程中保持相对稳定。尽管如此，IR区GC含量较高或波动较大的现象值得注意，这可能与富含rRNA基因及重复序列的区域特点有关[29]。与若干近缘研究比较，本研究样本覆盖更多种间差异，能够更全面地反映金丝桃属叶绿体基因组的整体特征及种间差异。

简单序列重复分析结果表明，单核苷酸重复（尤其是A/T类型）在9种金丝桃属物种中普遍存在，且在LSC区的分布最为集中。A/T的偏好性与大多数被子植物相似，这表明金丝桃属植物叶绿体基因组在重复序列的类型和分布上与其他植物相似[30]。通过分析重复序列的种类和数量，发现正向重复和回文重复占主导地位，以30–39 bp的短重复最为丰富，且大多数位于基因间区，理论上，这些特征可能促进局部重组或微小结构变动，影响基因组稳定性与IR边界演化[31]。

在密码子使用上，本研究观察到金丝桃属植物叶绿体基因组存在显著的A/U偏好性，这一现象在



x轴：窗口中点的位置；y轴：每个窗口的核苷酸多样性；红色虚线代表9种植物叶绿体基因组的平均核苷酸多样性水平

x-axis: midpoint position of each window; y-axis: nucleotide diversity of each window. The red dashed line represents the average nucleotide diversity across the complete chloroplast genomes of the nine *Hypericum* species

图8　9种金丝桃属植物的全叶绿体基因组滑动窗口分析

Fig. 8　Sliding window analysis of the complete chloroplast genomes of nine *Hypericum* species

所有物种中普遍存在。Leu、Arg与Ser的密码子类型最多，且A/U结尾的密码子使用频率显著高于G/C结尾的密码子，这种偏好性与其他被子植物的叶绿体基因组一致。这表明金丝桃属植物在叶绿体基因组的进化过程中保持了较为一致的密码子使用模式，但在不同物种之间仍存在微小的差异[32]，反映了轻微的选择性或突变偏向差异。

通过对叶绿体基因组的核苷酸多样性（π）滑动窗口分析，发现9种金丝桃属植物的叶绿体基因组在整体上显示出较低的多样性（π≈0.000 5–0.002 5），表明这些基因组在种间层面上保持了较高的保守性。然而，分析也揭示了一些高变异区域，特别是在*trnH-psbA*、*rpoC2-rpoC1*、*ndhF-rpl32*和*ycf1*等基因区段。这些区域使其成为潜在的分子标记，可用于物种鉴定和系统发育研究。与IR区域相比，LSC和SSC区域的变异性较高，符合被子植物叶绿体基因组普遍规律，即IR区域通常更加保守[33]。

通过基于106个叶绿体基因构建的最大似然（ML）系统发育树显示，*Hypericaceae*与*Podostemaceae*亲缘关系密切，在*Hypericum*属内形成单系进化枝，其中*H. petiolulatum*为最早分化物种，*H. forrestii*与*H. monogynum*亲缘关系较近，为该属系统发育提供了重要分子证据。

尽管叶绿体基因组在解析属内系统关系和识别高变区方面具有重要价值，但其单一遗传系统的局限性（如母系遗传和较低的有效群体大小）可能削弱对物种边界及复杂杂交历史的解析能力。因此，未来研究应结合核基因组SNP、全基因组测序或目标捕获数据等多组学信息进行综合分析，以更全面地揭示金丝桃属植物的系统发育格局与演化历史[34]。近年来，随着高通量测序和比较基因组学技术的快速发展，叶绿体基因组在植物系统发育、群体遗传和物种鉴定等领域的应用愈加广泛。本研究揭示了叶绿体基因组结构特征、重复序列分布及高变区段，为进一步探讨该属的系统演化与遗传多样性奠定了基础。未来可进一步整合叶绿体与核基因组信息（如genome skimming或target-capture技术），以检测属内潜在的杂交、基因渗流及叶绿体捕获事件，从而提升物种界限与演化历史的解析精度[35–36]。同时，应针对本研究中识别的高变区段设计特异性引物，并在更大样本与更广地理分布范围内进行验证，以评估其作为DNA条码或群体标记的实用性与稳定性[37]。此外，在关键样本中采用长读或混合组装策略，将有助于解析IR边界、重复序列及复杂结构区域；结合*dN/dS*或位点选择性分析方法，对高变基因的功能演化与适应性分化进行探讨[38–39]。通过这些多层次的研究策略，不仅可进一步完善金丝桃属系统发育与演化研究框架，还

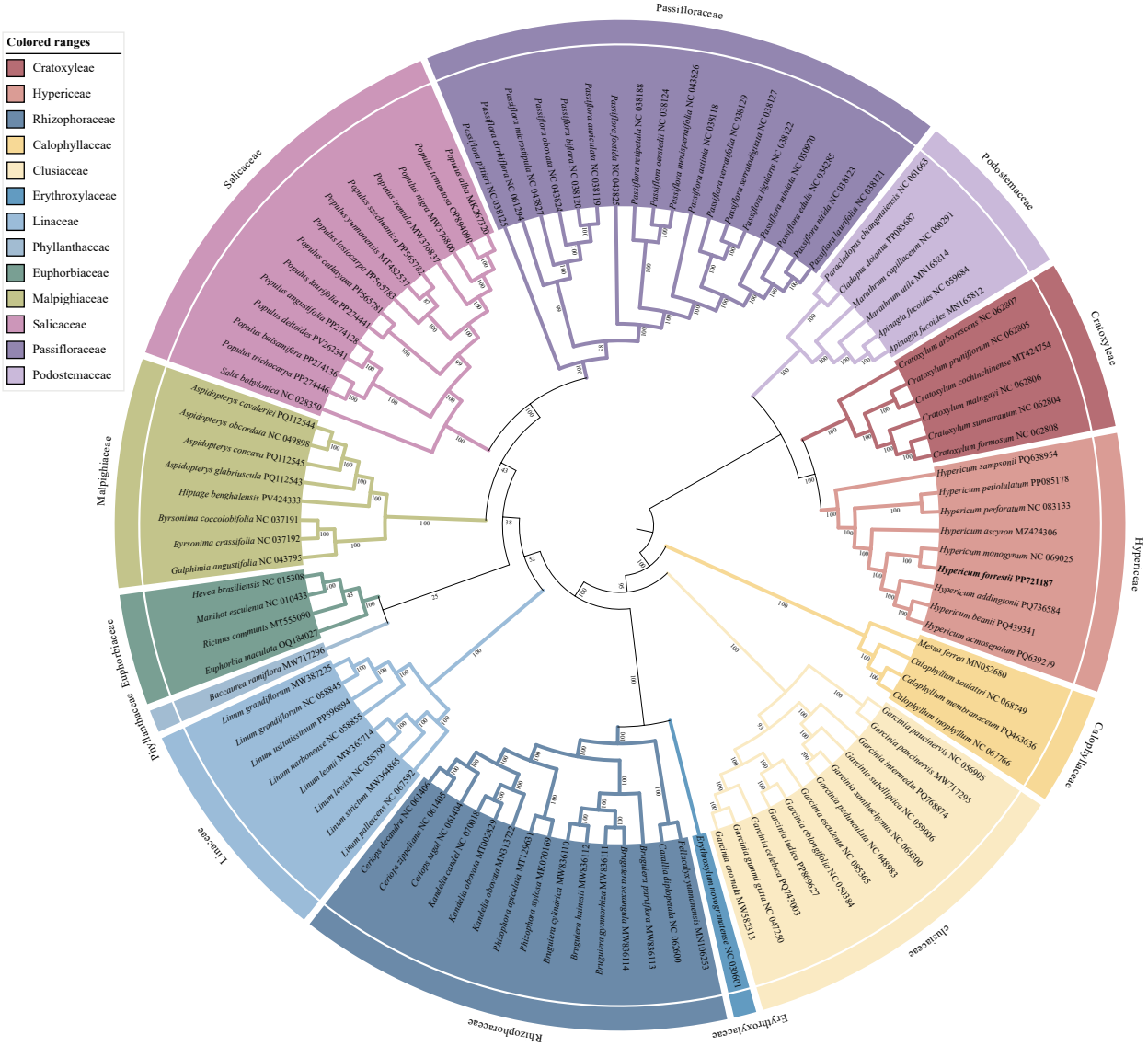


图9 基于最大似然法(ML)构建的金虎尾目植物叶绿体基因组系统发育树

Fig. 9 Maximum likelihood phylogenetic tree of Malpighiales based on chloroplast genomes

能为该属药用资源的开发与遗传多样性保护提供坚实的分子基础。

4 结论

九种金丝桃属植物叶绿体基因组均为四分体结构，基因组长度122 570–168 150 bp，其差异主要与IR区扩张或收缩及SSC区长度变化有关。各物种整体GC含量和基因组成较稳定。SSR以A/T单核苷酸重复为主，密码子使用呈明显A/U偏好。核苷酸多样性分析表明LSC和SSC为主要变异区域，*trnH-psbA*、*rpoC2-rpoC1*等为高变区。系统发育分析支持金丝桃属为单系群，并与Podostemaceae关系较近。

文章所有附图附表请点击链接获取（https://biotech.aiijournal.com/fileup/1002-5464/NEWS/20260509173108_NewsFile_25.pdf）。

参考文献

[1] Kladar N, Božin B, Bijelić K, et al. Biological activity of genus *Hypericum* Sect. *Hypericum* species——*H. Tetrapterum*, *H.*

maculatum subsp. *immaculatum*, *H. triquetrifolium* [J]. Molecules, 2023, 28(17): 6218.

[2] Apaydin EA, Maher AR, Shanman R, et al. A systematic review of St. John's wort for major depressive disorder [J]. Syst Rev, 2016, 5: 148.

[3] Barnes J, Anderson LA, Phillipson JD. St John's wort (*Hypericum perforatum* L.): a review of its chemistry, pharmacology and clinical properties [J]. J Pharm Pharmacol, 2001, 53(5): 583-600.

[4] Yan K, Lu X, Li WD, et al. Chloroplast genome diversity and molecular evolution in Hypericaceae: new insights from three *Hypericum* species [J]. Int J Mol Sci, 2025, 26(1): 323.

[5] Khade YP, Mainkar P, Chandanshive A, et al. Harnessing chloroplast SSRs to decipher genetic diversity in underutilized *Allium* species [J]. Front Plant Sci, 2025, 16: 1645145.

[6] Plunkett GM, Downie SR. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily apioideae [J]. Syst Bot, 2000, 25(4): 648.

[7] Dong WP, Liu J, Yu J, et al. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding [J]. PLoS One, 2012, 7(4): e35071.

[8] Cai LM, Xi ZX, Lemmon EM, et al. The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, malpighiales [J]. Syst Biol, 2021, 70(3): 491-507.

[9] Tokuoka T, Tobe H. Phylogenetic analyses of Malpighiales using plastid and nuclear DNA sequences, with particular reference to the embryology of Euphorbiaceae Sens. str [J]. J Plant Res, 2006, 119(6): 599-616.

[10] Pahlich E, Gerlitz C. A rapid DNA isolation procedure for small quantities of fresh leaf tissue [J]. Phytochemistry, 1980, 19: 11-13.

[11] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data [J]. Bioinformatics, 2014, 30(15): 2114-2120.

[12] Bankevich A, Gurevich AA, Dvorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing [J]. J Comput Biol, 2012, 19(5): 455-477.

[13] Shi LC, Chen HM, Jiang M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. Nucleic Acids Res, 2019, 47(W1): W65-W73.

[14] Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. Bioinformatics, 2012, 28(12): 1647-1649.

[15] Lohse M, Drechsel O, Kahlau S, et al. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. Nucleic Acids Res, 2013, 41(W1): W575-W581.

[16] Beier S, Thiel T, Münch T, et al. MISA-web: a web server for microsatellite prediction [J]. Bioinformatics, 2017, 33(16): 2583-2585.

[17] Kurtz S. REPuter: the manifold applications of repeat analysis on a genomic scale [J]. Nucleic Acids Res, 2001, 29(22): 4633-4642.

[18] Huang LJ, Yu HX, Wang Z, et al. CPStools: a package for analyzing chloroplast genome sequences [J]. iMetaOmics, 2024, 1(2): e25.

[19] Maździarz M, Zając S, Paukszto Ł, et al. RSCUcaller: an R package for analyzing differences in relative synonymous codon usage (RSCU) [J]. BMC Bioinform, 2025, 26: 141.

[20] Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes [J]. Bioinformatics, 2018, 34(17): 3030-3031.

[21] Frazer KA, Pachter L, Poliakov A, et al. VISTA: computational tools for comparative genomics [J]. Nucleic Acids Res, 2004, 32: W273-W279.

[22] Zhang D, Gao FL, 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.

[23] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability [J]. Mol Biol Evol, 2013, 30(4): 772-780.

[24] Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis [J]. Mol Biol Evol, 2000, 17(4): 540-552.

[25] Nguyen LT, Schmidt HA, von Haeseler A, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies [J]. Mol Biol Evol, 2015, 32(1): 268-274.

[26] Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool [J]. Nucleic Acids Res, 2024, 52(W1): W78-W82.

[27] Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets [J]. Mol Biol Evol, 2017, 34(12): 3299-3302.

[28] Cauz-Santos LA. Beyond conservation: the landscape of chloroplast genome rearrangements in angiosperms [J]. New Phytol, 2025, 247(6): 2571-2580.

[29] Fu N, Ji MY, Rouard M, et al. Comparative plastome analysis of Musaceae and new insights into phylogenetic relationships [J]. BMC Genom, 2022, 23: 223.

[30] Shen LW, Chen SQ, Liang M, et al. Comparative analysis of codon usage bias in chloroplast genomes of ten medicinal species of Rutaceae [J]. BMC Plant Biol, 2024, 24(1): 424.

[31] Dekhkonov D, Asatulloev T, Yusupov Z, et al. Comparative chloroplast genome and phylogenetic analysis of Central Asian

tulips [J]. Nord J Bot, 2025, 2025(7): e04460.

[32] Li JX, Du Y, Xie L, et al. Comparative plastome genomics and phylogenetic relationships of the genus *Trollius* [J]. Front Plant Sci, 2023, 14: 1293091.

[33] Meena RK, Negi N, Uniyal N, et al. Chloroplast-based DNA barcode analysis indicates high discriminatory potential of matK locus in Himalayan temperate bamboos [J]. 3 Biotech, 2020, 10 (12): 534.

[34] Fan YJ, Li CG, Ma C, et al. Holistic approach for plant species circumscription integrating standard barcodes, chloroplast genomes, single-copy nuclear genes and micro-morphological data: a case study in *Epimedium* (Berberidaceae) [J]. BMC Genom, 2025, 26: 931.

[35] Andermann T, Torres Jiménez MF, Matos-Maraví P, et al. A guide to carrying out a phylogenomic target sequence capture project [J]. Front Genet, 2020, 10: 1407.

[36] Trevisan B, Alcantara DMC, Machado DJ, et al. Genome skimming is a low-cost and robust strategy to assemble complete mitochondrial genomes from ethanol preserved specimens in biodiversity studies [J]. PeerJ, 2019, 7: e7543.

[37] Antil S, Abraham JS, Sripoorna S, et al. DNA barcoding, an effective tool for species identification: a review [J]. Mol Biol Rep, 2023, 50(1): 761-775.

[38] Wang WW, Schalamun M, Morales-Suarez A, et al. Assembly of chloroplast genomes with long- and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case [J]. BMC Genom, 2018, 19: 977.

[39] Wu H, Li DZ, Ma PF. Unprecedented variation pattern of plastid genomes and the potential role in adaptive evolution in Poales [J]. BMC Biol, 2024, 22: 97.

（责任编辑　高洁）