• 药材与资源 •

喜马拉雅大黄叶绿体基因组特征及进化分析

米玛潘多[1]，达瓦潘多[1]，格桑文灯[2]，仁　增[1]，洛桑塔杰[1]，罗桑江才[1]，徐立军[1*]，次　仁[1*]

1. 西藏藏医药大学，西藏 拉萨　850000
2. 上海交通大学，上海　200240

摘　要：目的　喜马拉雅大黄 *Rheum webbianum* 作为青藏高原特有的民族药用资源，药用价值与开发潜力突出，目前其叶绿体（chloroplast，CP）基因组仍未见报道。**方法**　以喜马拉雅大黄为材料，采用高通量测序，首次对喜马拉雅大黄 *CP* 基因组进行测序，并分析重复序列、密码子偏好性等；结合已发表的近缘 *CP* 基因组开展基因组结构比较和系统进化分析。**结果**　喜马拉雅大黄 *CP* 基因组为典型环状结构，长度 162 150 bp，GC 含量 37.40%。基因组共注释 125 个基因，包括 34 个转运 RNA（transfer RNA，tRNA），8 个核糖体 RNA（ribosomal RNA，rRNA），83 个蛋白编码基因（protein coding genes，PCGs）。重复序列分析鉴定 48 个重复序列和 263 个简单重复序列（simple sequence repeat，SSR）位点；密码子偏好性分析显示大黄属密码子存在显著 A/T 偏好，第 3 位倾向 A/U 结尾，自然选择是影响密码子偏差的主要因素，共有最优密码子为 UUU 和 UUA。基因组结构比较发现，大黄属 CP 整体结构保守，反向重复序列（inverted repeat，IR）区域没有明显的扩张和收缩现象；核苷酸多样性（Pi）分析筛选 3 个显著变异位点区域，即 *trnH-GUG-psbA*、*rpoC1-CDS1-rpoB*、*trnG-UCC-rps14*，可作为候选 DNA 条形码序列。系统进化分析显示喜马拉雅大黄 *R. webbianum* 与心叶大黄 *R. acuminatum*，塔黄 *R. nobile*、菱叶大黄 *R. rhomboideum* 和穗序大黄 *R. spiciforme* 聚为一支，节点支持率为 100%，亲缘关系较近。**结论**　首次公布了喜马拉雅大黄 *CP* 基因组，注释并解析了基因组结构特征；筛选 3 个高变异位点区域可作为大黄属物种鉴定的候选 DNA 条形码；明确其亲缘关系，丰富了大黄属 *CP* 基因组数据库。

关键词：喜马拉雅大黄；叶绿体基因组；密码子偏好；基因组结构比较；系统进化分析

中图分类号：R282.6　　**文献标志码：**A　　**文章编号：**0253 - 2670(2026)22 - 9078 - 13

DOI: 10.7501/j.issn.0253-2670.2026.22.021

Characteristics and phylogenetic analysis of chloroplast genome of *Rheum webbianum*

MYIMA Pantok[1], DAWA Pantok[1], KELSANG Wendeng[2], REN Zeng[1], LUOSANG Tajie[1], LUOSANG Jiangcai[1], XU Lijun[1], TSE Ring[1]

1. University of Tibetan Medicine, Lhasa 850000, China
2. Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: Objective　*Rheum webbianum*, a unique ethnic medicinal resource of the Qinghai-Tibet Plateau, possesses prominent medicinal value and development potential. However, its chloroplast (CP) genome has not yet been reported. **Methods**　Using *R. webbianum* as material, high-throughput sequencing was performed to sequence its CP genome for the first time, followed by analyses of repetitive sequences and codon usage bias. Comparative genomic and phylogenetic analyses were further conducted using publicly available CP genome data of congeneric species. **Results**　The *R. webbianum* CP genome exhibits a typical circular structure, with a total length of 162 150 bp and a GC content of 37.40%. A total of 125 genes were annotated in the genome, including 34 transfer RNA (tRNA), 8 ribosomal RNA (rRNA), and 83 protein coding genes (PCGs). Repeat sequence analysis identified 48 repeat sequences and 263 simple-sequence repeat (SSR) loci. Codon-usage bias analysis revealed a pronounced A/T bias in codons of *Rheum*, with a

收稿日期：2026-05-05

基金项目：西藏自治区藏药管理局重点项目（JJKT2023013）；西藏藏医药大学科研创新团队项目（KYTD2303）；2024 年度博士点建设项目-重点学科建设项目（BSDJS-XKJS-24-05）

作者简介：米玛潘多，研究员，主要从事中藏植物资源研究。E-mail: pentokmy@163.com

***通信作者：**徐立军，研究员，主要从事藏药资源开发利用。E-mail: 623740161@qq.com

次　仁，副教授，主要从事藏医药研究。E-mail: zangyiciren@163.com

preference for A/U at the third codon position. Natural selection is the primary factor influencing codon-usage bias, and the shared optimal codons were UUU and UUA. Comparative genomic analysis showed that the overall structure of the *Rheum* CP genome was conserved, with no obvious expansion or contraction in the inverted repeat (IR) regions. Three highly divergent regions, namely *trnH-GUG-psbA*, *rpoC1-CDS1-rpoB* and *trnG-UCC-rps14*, were screened via nucleotide diversity (Pi) analysis and proposed as candidate DNA barcode sequences. Phylogenetic analysis revealed that *R.webbianum* forms a well-supported clade (100% bootstrap value) with *R. acuminatum*, *R. nobile*, *R. rhomboideum and R. spiciforme*, indicating a close genetic relationship. **Conclusions** For the first time, the complete CP genome of *R. webbianum* was sequenced, annotated, and its structural characteristics were systematically analyzed. Three highly variable regions were identified as candidate DNA barcodes for species identification within the genus *Rheum*. The present findings clarify the genetic relationship of *R. webbianum* and enrich the CP genomic resource database of the genus *Rheum*.
Key words: *Rheum webbianum* Royle; chloroplast genome; codon usage bias; comparative genomics; phylogenetic analysis

喜马拉雅大黄 *Rheum webbianum* Royle 是蓼科（Polygonaceae）大黄属 *Rheum* L.多年生高大草本植物，具有显著的形态特征和重要的药用价值[1-2]。其茎生小叶呈卵形，成熟叶片基部呈圆形或心形；圆锥花序，开浅黄色小花。果实为宽椭圆形或近圆形蒴果。根茎粗壮，表面暗褐色，断面橙黄色，是传统藏药系统中的重要药材[3-4]。它自然分布于喜马拉雅山脉周边地区，包括中国青藏高原、印度北部和巴基斯坦等地。在我国主要集中分布在海拔 3 500～4 660 m 的青藏高原地区，这种特殊的生长环境可能与其活性成分的积累有密切关系[5-6]。大黄的药用价值主要源于其根部富含的次生代谢产物，特别是蒽醌类化合物如大黄酸、大黄素、芦荟大黄素、大黄素甲醚和大黄酚等，具有抗肿瘤、改善肾功能和降血脂等多重功效[7]。近年研究发现其提取物能通过血脑屏障，具有改善老年认知功能障碍、增强记忆力的潜力，这为开发抗阿尔茨海默病药物提供了新思路[8]。随着现代研究技术的深入，喜马拉雅大黄展现出越来越广阔的药用开发前景。

叶绿体（chloroplast，CP）是植物和某些真核生物细胞中的关键细胞器，也是光合作用的主要场所，还参与氨基酸、脂肪酸、核苷酸与植物激素等多种物质的代谢合成过程，在植物能量代谢与物质合成网络中占据枢纽地位[9-10]。最新研究表明，CP 还是关键的“环境感受器”与“信号调控中枢”，通过逆向信号向细胞核传递自身发育与胁迫状态，精准调控核基因表达，重塑植物生长发育与抗逆响应[11]。其基因组通常为双链环状，少数呈线状，大小因类群而异。维管植物多在 120～160 kb，蕨类为 140～150 kb，而藻类变异幅度较大，为 37～2 Mb [12-13]。CP 基因组结构相对保守，多数具有典型四分体构型，包括一个大单拷贝区（large single copy，LSC）、一个小单拷贝区（small single copy，SSC）以及一对反向重复区（inverted repeat，IR）。因其结构保守、进化速率适中，*CP* 基因组在研究分子进化和系统发育中具有重要价值，也为通过 CP 工程改良作物品质和提高抗逆性提供了理论和应用依据[14]。

目前，大黄属部分物种的叶绿体基因组已有报道，如锐齿大黄 *R. acuminatum* Hook. f. & Thomson（161 306 bp）[15]、掌叶大黄 *R. palmatum* L.（161 136 bp）和药用大黄 *R. officinale* Baill.（161 093 bp）[16]等。但是，喜马拉雅大黄作为青藏高原特有的高海拔（3 500～4 660 m）药用植物，在大黄属植物进化和遗传研究中具有重要地位[17]，其 *CP* 基因组至今未见报道。本研究以喜马拉雅大黄为对象，通过高通量测序和注释，系统解析其 *CP* 基因组的结构与特征，开展与近缘种的比较基因组分析。研究结果将为喜马拉雅大黄资源的保护、分类鉴定及系统发育研究提供重要的数据支撑。

1 材料

本研究喜马拉雅大黄样品采集于西藏自治区拉萨市堆龙德庆区古荣镇楚布寺前方山上（29°43′49″N，90°35′59″E），经西藏自治区高原研究所央金卓嘎研究员鉴定为喜马拉雅大黄 *Rheum webbianum* Royle，现存于西藏藏医药大学藏医药研究所。于 2023 年 10 月采集种子，在本校藏药研发中心实验室进行栽培研究。待长出叶子后采集新鲜幼嫩叶片，于−80 ℃低温保存，用于后续实验。

2 方法

2.1 DNA 提取和测序

采用改良 CTAB 法提取 DNA[18]。利用 1%琼脂糖电泳检查 DNA 完整性好，分光光度计测定 DNA 的浓度与纯度。将合格的 DNA 样品，送至南京集思慧远有限公司高通量测序平台完成 DNA 测序。

2.2　基因组组装及注释

测序获得的原始数据通过 NGS QC Toolkit 进行质量过滤[19]，去除低质量 reads 后获得高质量序列（clean reads）。采用 SPAdes v3.11.1 对质控后的 clean reads 进行拼接[20]。组装序列与 NCBI 数据库（https://www.ncbi.nlm.nih.gov/）中已发表的大黄属近缘物种叶绿体基因组序列进行同源比对，选择最佳参考基因组，并下载 GenBank 文件。使用 Prodigal v2.6.3 软件对叶绿体基因组进行注释[21]，HMMER v3.1 软件扫描 tRNA 基因，Aragorn v1.2.3 软件识别 rRNA 基因[22-24]。将注释后的喜马拉雅大黄 *CP* 基因组上传到 NCBI 数据库中（登录号：PQ963207）。使用 Organellar Genome DRAW v1.3.1 生成基因组图谱[23]。

2.3　重复序列分析

采用 vmatch v2.3.0 软件检测重复结构[25]，包括正向（forward，F）、反向（reverse，R）、互补（complement，C）和回文（palindromic，P）。利用 MISA 软件鉴定基因组的简单重复序列（simple sequence repeats，SSR）[26]，包括单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸及六核苷酸重复类型。最小重复次数分别设置为 10、6、5、5、5 和 5。

2.4　密码子偏好性分析

密码子偏好性参考郑梦迪等[27]的方法，使用 CodonW 和 mEMBOSS 计算密码子参数，软件 ChiPlot 分析相关性及显著性。

2.5　基因组结构比较分析

以喜马拉雅大黄 *CP* 基因组为参考，对已发表的 9 个 *CP* 基因组进行注释校正，通过 IRscope 软件解析 IR 边界扩张/收缩动态，生成 *CP* 基因组 IR 区结构示意图[28]。

采用 K_aK_s_calculator v3.0 软件对 10 个大黄属物种 *CP* 基因组蛋白编码基因进行两两比对，计算获得各基因的同义突变频率（K_s）、非同义突变频率（K_a）以及 K_a/K_s 值，用于评估基因的选择压力[29]。使用 DnaSP 软件确定 Pi，设置步长为 200 bp，窗口长度为 600 bp[30]。

2.6　系统发育分析

从 NCBI GenBank 数据库中下载 25 个相关物种的叶绿体基因组序列，以 *Afrobrunnichia erecta*（MH286316.1）作为外类群，使用 mafft v7.505 软件进行多序列比对[31]。采用 RAxML v8.2.12 构建最大似然进化树，参数设置为 1 000 次 bootstrap 迭代，模型为 GTRGAMMA[32]。

3　结果与分析

3.1　喜马拉雅大黄 *CP* 基因组的结构和基因注释

喜马拉雅大黄 *CP* 基因组呈现典型的四分体结构，全长 162 150 bp，包括 1 个 LSC（87 366 bp）、1 个 SSC（12 832 bp）以及 1 对反向重复区（IRa/b，各 30 976 bp），GC 含量为 37.40%（图 1）。基因组

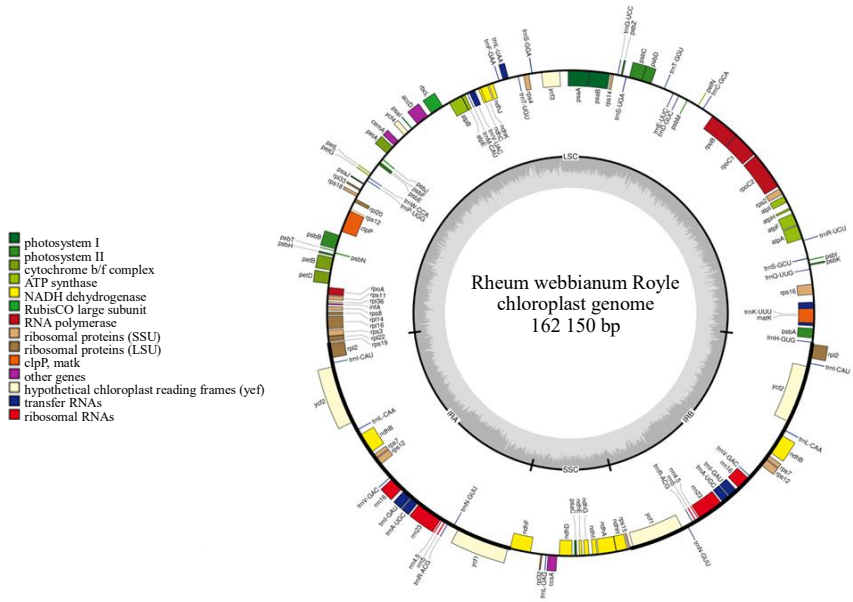


图 1　喜马拉雅大黄 *CP* 基因组结构图

Fig. 1　Structure chart of *CP* genome of *R. webbianum*

总共注释到 125 个基因，包括 34 个 tRNA，8 个 rRNA 和 83 mRNA，其中拥有内含子基因 11 个，基因 *ndhA*、*ndhB*、*petB*、*petD*、*rpl16*、*rpl2*、*rpoC1*、*rps12*、*rps16* 拥有 1 个内含子；*clpP* 和 *ycf3* 拥有 2 个内含子。多拷贝基因有 10 个且均具有 2 个拷贝，基因分别为 *ndhB*、*rpl2*、*rrn16*、*rrn23*、*rrn4.5*、*rrn5*、*rps12*、*ycf2*、*ycf1*、*rps7*（表 1）。

表 1 喜玛拉雅大黄 *CP* 基因组中基因预测结果

Table 1 Gene prediction results of *R. webbianum CP* genome

基因类型	基因名称
ATP 合成酶	*atpA*、*atpB*、*atpE*、*atpF*、*atpH*、*atpI*
光系统II亚基	*psbA*、*psbB*、*psbC*、*psbD*、*psbE*、*psbF*、*psbH*、*psbI*、*psbJ*、*psbK*、*psbM*、*psbN*、*psbT*、*psbZ*、*ycf3*
NADH 脱氢酶亚基	*ndhA**、*ndhB**(2)、*ndhB*、*ndhC*、*ndhD*、*ndhE*、*ndhF*、*ndhG*、*ndhH*、*ndhI*、*ndhJ*、*ndhK*
细胞色素b/f复合物亚基	*petA*、*petB**、*petD**、*petG*、*petL*、*petN*
光系统I亚基	*psaA*、*psaB*、*psaC*、*psaI*、*psaJ*
核酮糖二磷酸羧化酶	*rbcL*
核糖体大亚基	*rpl14*、*rpl16**、*rpl2**(2)、*rpl20*、*rpl22*、*rpl23*、*rpl32*、*rpl33*、*rpl36*
DNA 依赖的 RNA 聚合酶	*rpoA*、*rpoB*、*rpoC1**、*rpoC2*
核糖体RNA	*rrn16*(2)、*rrn23*(2)、*rrn4.5*(2)、*rrn5*(2)
核糖体小亚基	*rps11*、*rps12***(2)、*rps14*、*rps15*、*rps16**、*rps18*、*rps19*、*rps2*、*rps3*、*rps4*、*rps7*(2)、*rps8*
乙酰辅酶A 羧化酶亚基	*accD*
c 型细胞色素合成基因	*ccsA*
外膜包膜蛋白	*cemA*
蛋白酶	*clpP***
成熟酶	*matK*
翻译起始因子	*infA*
保守开放阅读框	*ycf1*(2)、*ycf2*(2)、*ycf3***、*ycf4*

*-含 1 个内含子的基因；**-含 2 个内含子的基因；（2）-多拷贝基因的拷贝数量。

*-gene with one intron; **-gene with two introns; (2)-number of copies of multi-copy genes.

3.2 重复序列分析

在喜玛拉雅大黄 *CP* 基因组中检测到 48 个重复序列，包括 22 个正向重复、26 个回文重复、没有发现反向重复和互补重复（图 2）。绝大多数重复序列的长度范围为 30～50 bp，其中以 30 bp（15 个）和 31 bp（12 个）的序列数量最多，为最主要的重复类型。长度为 32 bp 和 39 bp 的最少，各仅有 1 个序列。值得注意的是，基因组中发现存在一条长度为 30 976 bp 的超长回文重复序列，位于反向重复区（IR）结构中，这可能与被子植物叶绿体基因组 IR 区高度保守性具有相关性。

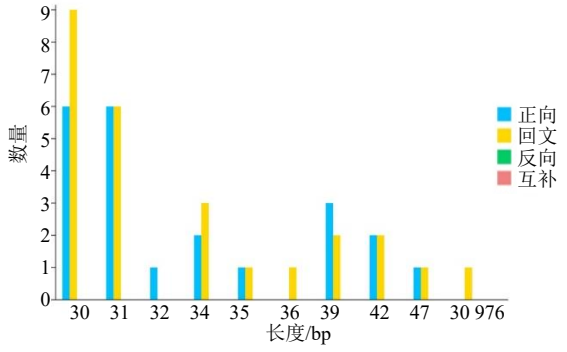


图 2 喜玛拉雅大黄 *CP* 基因组中的长重复序列的长度与类型分布

Fig. 2 Length and type distribution of long repeat sequences in *R. webbianum CP* genome

SSR 以 1～6 个碱基为重复单元，串联重复形成短序列，广泛分布于叶绿体基因组中。通过 MISA 软件在喜玛拉雅大黄 *CP* 基因组中发现 263 个 SSRs。这些 SSRs 分属 6 种不同类型，其中最主要的是单核苷酸重复（171 个，65.019%），其次是三核苷酸重复（71 个，26.996%）和二核苷酸重复（13 个，4.94%）。另外 3 种 SSRs 出现频率较低：四核苷酸（5 个，1.901%）、五核苷酸（1 个，0.38%）和六核苷酸重复（2 个，0.76%）（图 3）。此外，绝大多数 SSRs 属 A/T 类型（68.44%），表明短 A 或 T 重复序列在喜玛拉雅大黄 *CP* 基因组 SSR 中占比最高。

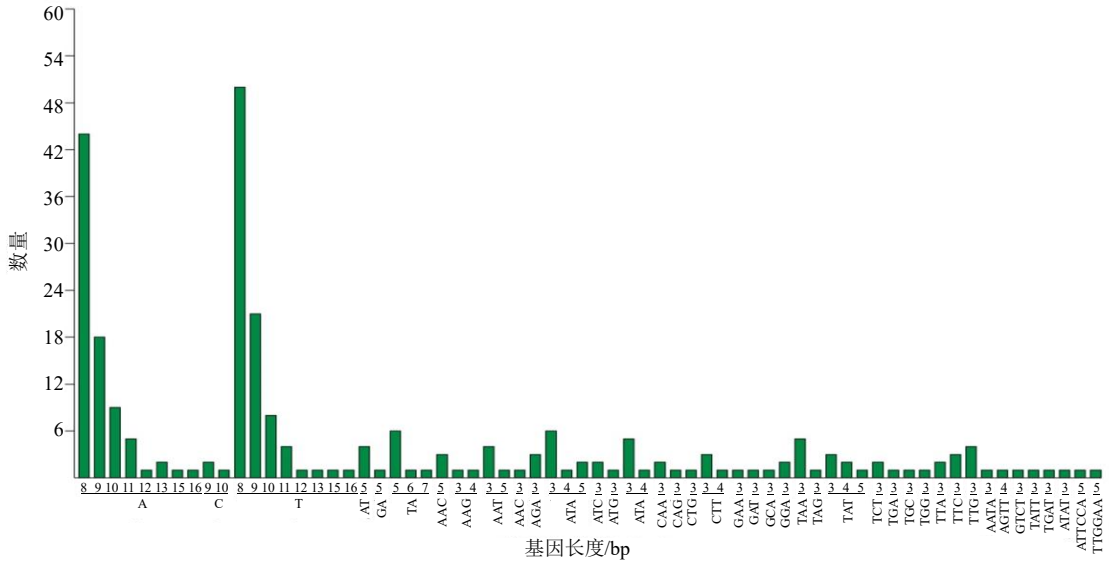


图 3　喜玛拉雅大黄 *CP* 基因组 SSR 重复的类型和数量

Fig. 3　Types and numbers of SSR in *R. webbianum* CP genome

3.3　密码子偏好性分析

喜马拉雅大黄 *CP* 基因组编码序列分析结果显示，共有 20 571 个密码子参与蛋白编码。对 10 种大黄属植物叶绿体 CDS 序列开展密码子偏好性分析（表 2），结果表明喜马拉雅大黄密码子数量处于中等水平，其中密码子数量最多的是唐古特大黄，为 21 100 个，最少是河套大黄，为 19 077 个。各物种平均 GC 含量为 38.24%～38.55%，3 位点 GC 含量为 GC_1（46.36%～46.88%）＞GC_2（38.13%～38.44%）＞GC_3（30.22%～30.37%），均小于 50%，表明密码子构成整体偏向 A/T，第 3 位碱基更倾向于 A/U 结尾。有效密码子数（effective number of codons，ENC）平均值在 48.28～48.41，均大于 45，说明密码子使用呈弱偏好。CAI 值介于 0.197～0.200，整体 CAI 数值偏低。相关性分析（图 4）表明，GC 含量与 GC_1（0.83～0.84）、GC_2（0.79～0.83）和 GC_3（0.48～0.60）呈显著正相关；GC_1 与 GC_2 显著正相关（0.47～0.51），二者与 GC_3 相关性均偏弱（0.14～0.25），说明密码子第 1、2 位碱基组成保守且存在协同进化，第 3 位碱基则相对独立、变异较大。ENC 与 GC_{3s} 呈显著正相关（0.33～0.45），表明密码子偏好性与第 3 位 GC 含量密切相关。

表 2　大黄属 *CP* 基因组 CDS 密码子组成

Table 2　Codon composition of *CP* genome CDSs of *Rheum*

物种	拉丁名称	密码	GC_1	GC_2	GC_3	GC_{all}	ENC	GC_{3s}	CAI
喜马拉雅大黄	*R. webbianum*	20571	46.75	38.29	30.28	38.44	48.37	30.3	0.198
总序大黄	*R. racemiferum*	20576	46.68	38.29	30.22	38.4	48.28	30.2	0.199
华北大黄	*R. franzenbachii*	20576	46.67	38.29	30.22	38.39	48.28	30.2	0.199
河套大黄	*R. hotaoense*	19077	46.62	38.22	30.37	38.40	48.38	30.4	0.200
药用大黄	*R. officinale*	20665	46.63	38.15	30.22	38.33	48.32	30.2	0.199
掌叶大黄	*R. palmatum*	20595	46.68	38.28	30.30	38.42	48.35	30.3	0.199
歧穗大黄	*R. przewalskyi*	20097	46.88	38.44	30.33	38.55	48.41	30.3	0.199
小大黄	*R. pumilum*	20601	46.72	38.3	30.27	38.43	48.36	30.3	0.198
心叶大黄	*R. acuminatum*	20079	46.36	38.13	30.24	38.24	48.32	30.2	0.198
唐古特大黄	*R. tanguticum*	21100	46.55	38.24	30.29	38.36	48.31	30.3	0.197

为探究密码子偏好性的形成机制，采用中性绘图、ENC-plot 和 PR2-plot 进行综合分析。10 个物种基因分布均偏离对角线（斜率 0.240～0.494），GC_{12} 与 GC_3 相关性弱，R^2_{adj} 值极低（0.011～0.100），表明突变对密码子偏好性影响微弱，自然选择压力的作用显著（图 5）。ENC-plot 显示，各物种基因点

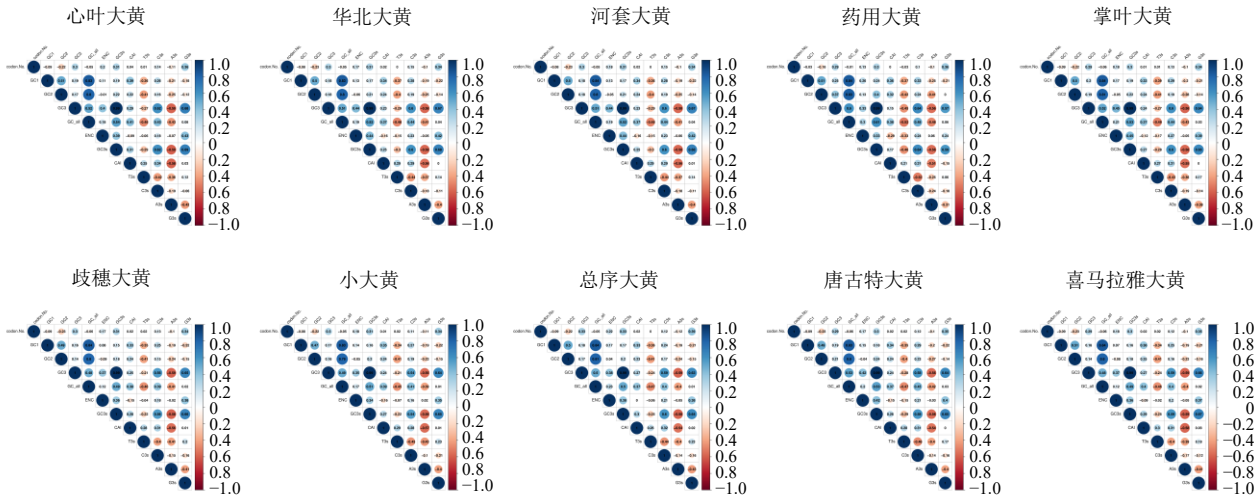


图 4　大黄属植物 *CP* 基因组密码子使用指标的相关分析热图

Fig. 4　Correlation analysis heatmaps of *CP* genome codon usage indicators in *Rheum* species

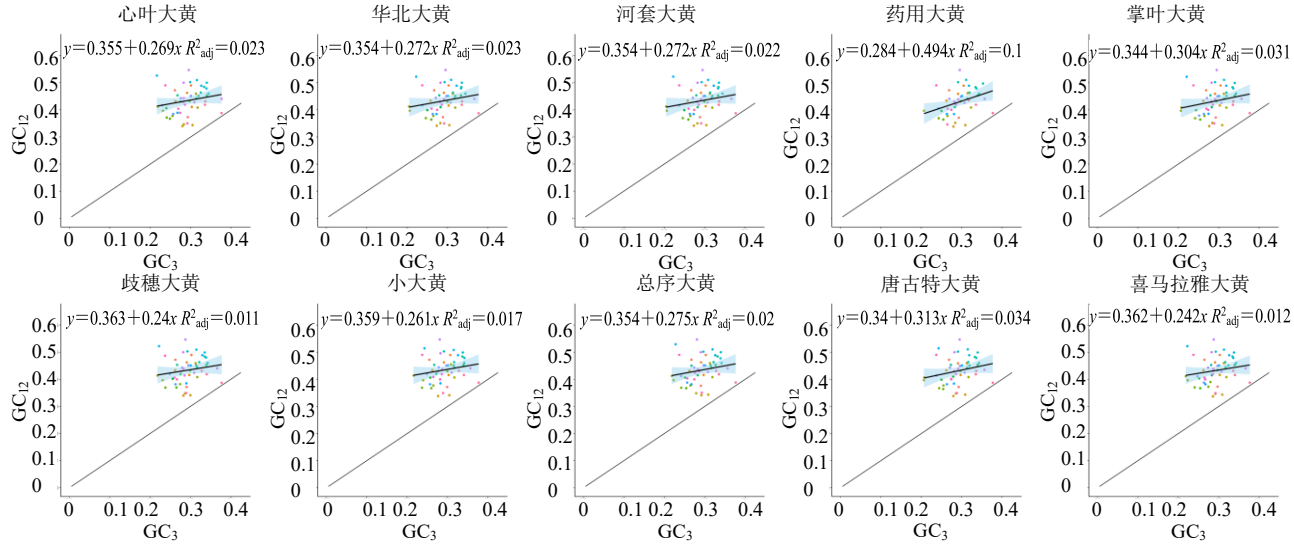


图 5　大黄属植物 *CP* 基因的密码子中性图谱分析

Fig. 5　Codon neutrality plot analysis of *CP* genome in *Rheum* species

ENC 值均＞35，绝大多数位于标准曲线下方（91.3%～96.2%），56.9%～67.4%的基因靠近标准曲线，32.6%～43.1%远离，进一步证实自然选择占主导（图 6）。PR2-plot 分析显示，第 3 位碱基分布不均，多数位于下半部分（第 3、4 象限和右半部分（第 1、4 象限），表明 A/T 碱基更倾向 T、G/C 碱基更倾向 G，即密码子第 3 位碱基整体倾向 G/T（图 7）。结果表明大黄属密码子偏好性由突变压力与自然选择共同作用，自然选择为主导，各物种受自然选择影响程度排序为歧穗大黄＞喜马拉雅大黄＞小大黄＞总序大黄＞河套大黄＞心叶大黄＞华北大黄＞掌叶大黄＞唐古特大黄＞药用大黄。

基于相对同义密码子使用度（RSCU）值和 ENC 值筛选最优密码子（ΔRSCU＞0.08 且 RSCU＞1），10 个物种共筛选出 24～28 个最优密码子数，均以 U/A 结尾，与叶绿体基因组第 3 位碱基偏好 A/T 一致（图 8）。其中喜马拉雅大黄和掌叶大黄、唐古特大黄、小大黄、总序大黄最多，各有 28 个。心叶大黄、华北大黄和河套大黄各有 26 个，歧穗大黄和药用大黄最少，各有 24 个。10 个物种共享的最优密码子为 UUU（Phe）和 UUA（Leu），显示大黄属密码子使用策略高度保守。RSCU 值分布在

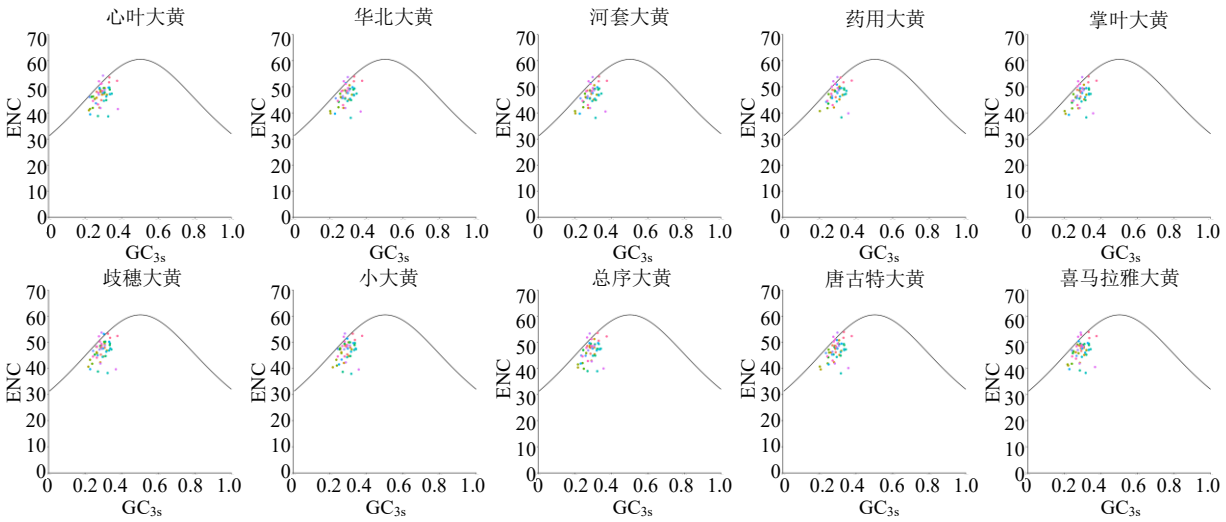


图 6　大黄属植物 *CP* 基因组中 ENC 和 GC_3 的关系

Fig. 6　Relationship between ENC and GC_3 in *CP* genome of *Rheum* species

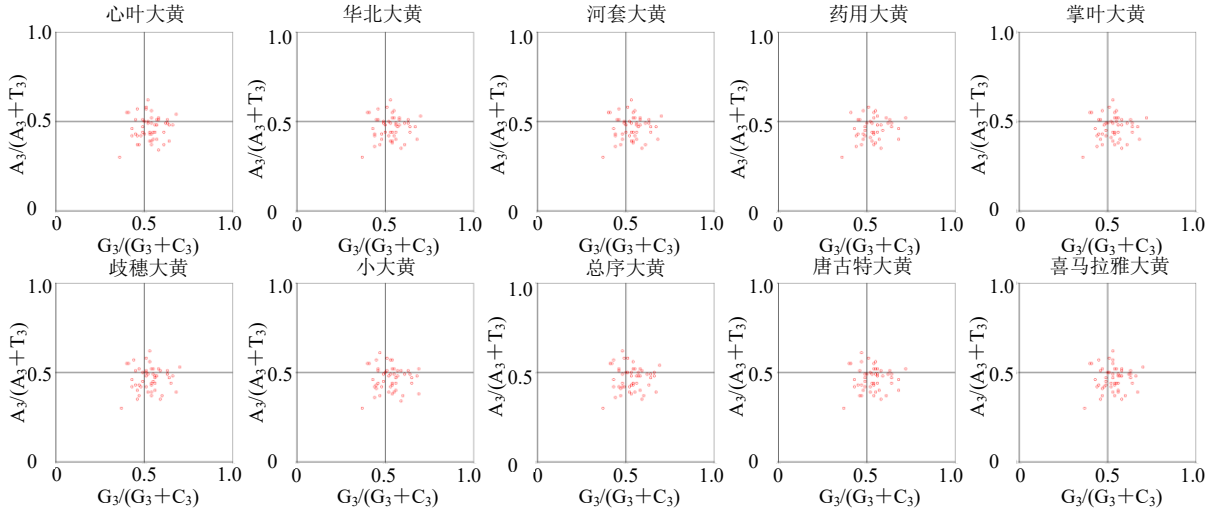


图 7　大黄属植物的 PR2-plot 分析

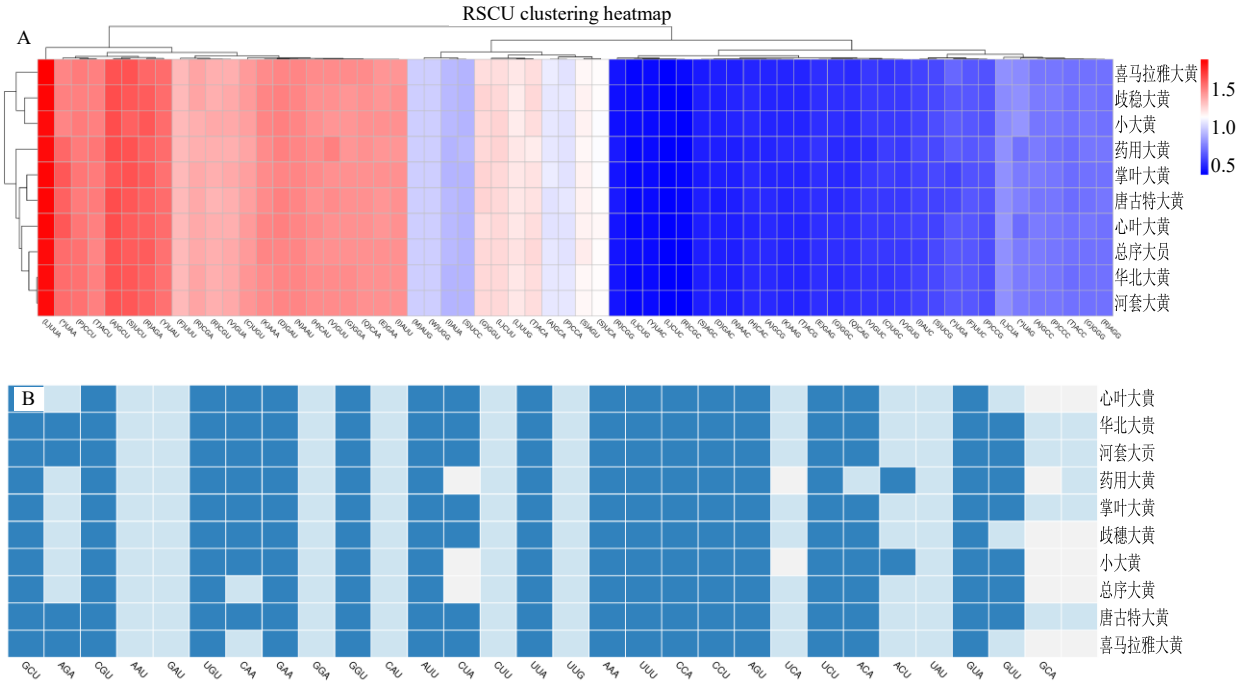
Fig. 7　PR2-plot analysis of *Rheum* species

0.5～1.8，高频密码子（RSCU＞1）28 个，多以 U/A 结尾（占比 90%以上），仅 UUG（Leu）以 G 结尾，其中 RSCU 值最高的密码子为 UUA（Leu），其次为 AUU（Ile）和 UUU（Phe）。

3.4　基因组结构特征比较分析

根据已发表 9 个大黄属 CP 进行比较，喜马拉雅大黄及其近缘种的叶绿体基因组大小在 161 093～162 150 bp，变异幅度较小。这 10 个叶绿体 IR 边界结构分析显示（图 9），*rps19* 基因总长度为 279 bp，横跨 LSC 和 IRb 区域，且在 IRb 区域的长度均为 108 bp。*ndhF* 基因总长度除了药用大黄，掌叶大黄和唐古特大黄为 2 235 bp，其余 7 个大黄属均为 2 244 bp，均横跨 IRb 和 SSC 区域。在 IRb 区域，除了小大黄为 98 bp，歧穗大黄为 95 bp，其余均为 95 bp。*ycf1* 和 *rpl2* 均位于 IR 区域，没有出现横跨。说明大黄属 IR 比较保守，没有明显的扩张和收缩现象。

Pi 分析如下图所示（图 10），对于编码区，整体核苷酸多样性水平较低（Pi＝0～0.027）。*rps*12、*ndhH*、*rps15*、*psbZ*、*matK*、*ndhF*、*rpl32*＞0.01（超过平均值 3 倍），变异程度较高的热点区，其中 *rps12* 的 Pi 值最高为 0.027 最高显示出 Pi＞0.025（超过平均值 7.88 以上）变异程度最高。相比之下，大部分编码区基因的 Pi 值接近 0，表现出高度的序列保守性。基因间隔区的整体核苷酸多样性水平显著高于编码区（Pi＝0～0.107 59），平均值为 0.010 6，有 7 个区 *trnH-GUG-psbA*、*rpoC1-CDS1-rpoB*、*trnG-UCC-rps14*、*ccsA-*



深蓝色背景：RSCU＞1 和 ΔRSCU≥0.08；浅蓝色背景：RSCU＞1

Dark blue background: RSCU > 1 and ΔRSCU ≥ 0.08; lighr blue background: RSCU > 1

图 8 RSCU 热图 (A) 和最优密码子 (B)

Fig. 8 RSCU heatmap (A) and optimal codons (B)

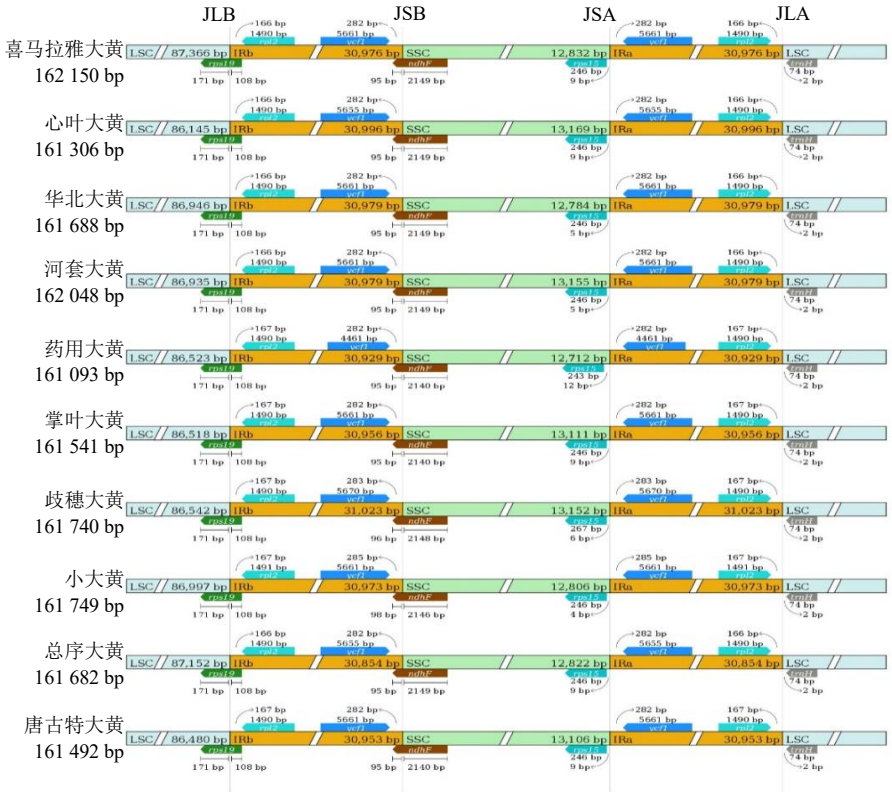


图 9 喜玛拉雅大黄及其近缘种 *CP* 基因组比较

Fig. 9 Comparative analysis on *CP* genomes of *R. webbianum* and its closely related species

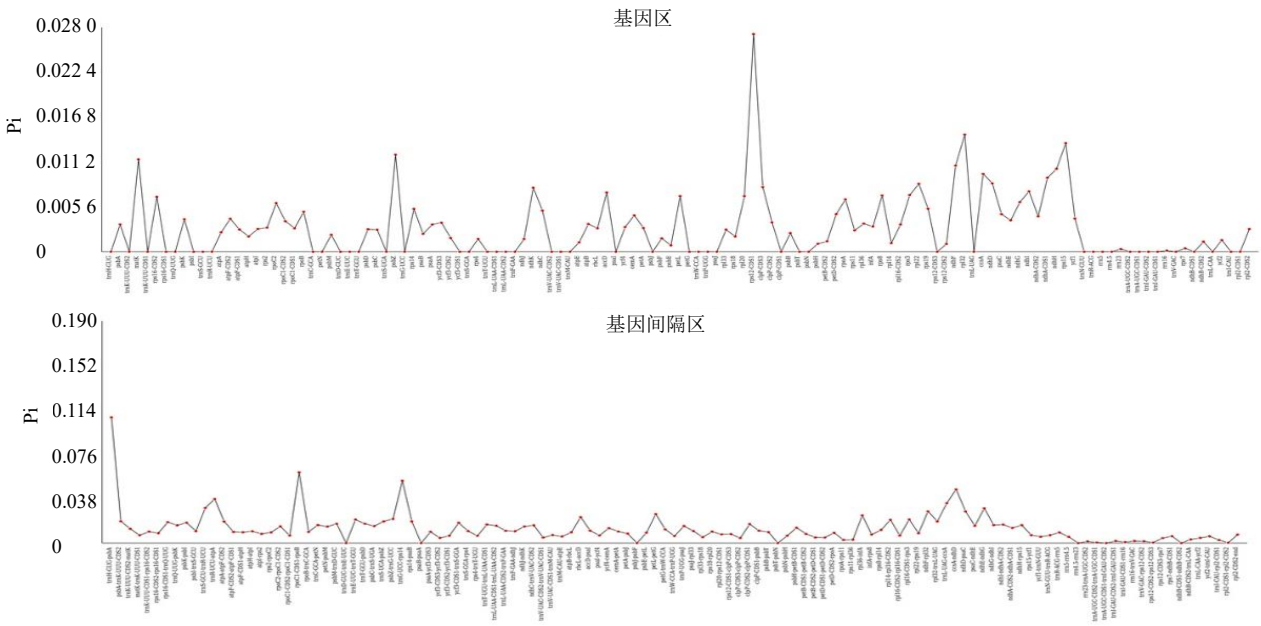


图 10　大黄属 *CP* 基因组滑动窗口分析

Fig. 10　Sliding window analysis of *CP* genomes of *Rheum* species

ndhD、*trnR-UCU-atpA*、*trnL-UAG-ccsA*、*trnS-GCU-trnR-UCU*＞0.03（超过平均值 3 倍），为变异热点区，其中 3 个 *trnH-GUG-psbA*、*rpoC1-CDS1-rpoB*、*trnG-UCC-rps14*＞0.05（超过平均值 5 倍以上）为显著的变异热点。这些高变区间序列差异显著，可为大黄属物种鉴定、系统发育分析提供潜在的分子标记。

选择压力分析结果如图（图 11）所示。大部分叶绿体基因的中位数 K_a/K_s 显著小于 1，整体受到强烈的纯化选择，核心功能基因在进化上高度保守；部分基因如 *ycf2*、*matK*、*ycf1*、*rpoA* 等，在个别物种中出现 K_a/K_s＞1 的离群值，其中 *ndhF* 异常值接近 9，暗示这些基因选择压力发生改变，存在适应性进化的潜在信号。经过 P≤0.05 显著性过滤后（图 12），*ycf1*、*ycf2*、*matK*、*rpoA* 等基因的 K_a/K_s 中位数相对较高（＞0.25），表明这些基因在物种间存在显著的进化速率差异。值得注意的是，*rps12* 的 K_a/K_s 中位数约为 0.55，明显高于其他显著基因，反映该基因纯化选择压力发生明显松弛，可能发生功能分化。此外，*psbA*、*rbcL*、*atpA*、*atpB*、*psbF* 等光合作用核心基因的 K_a/K_s 中位数均低于 0.3，说明这类

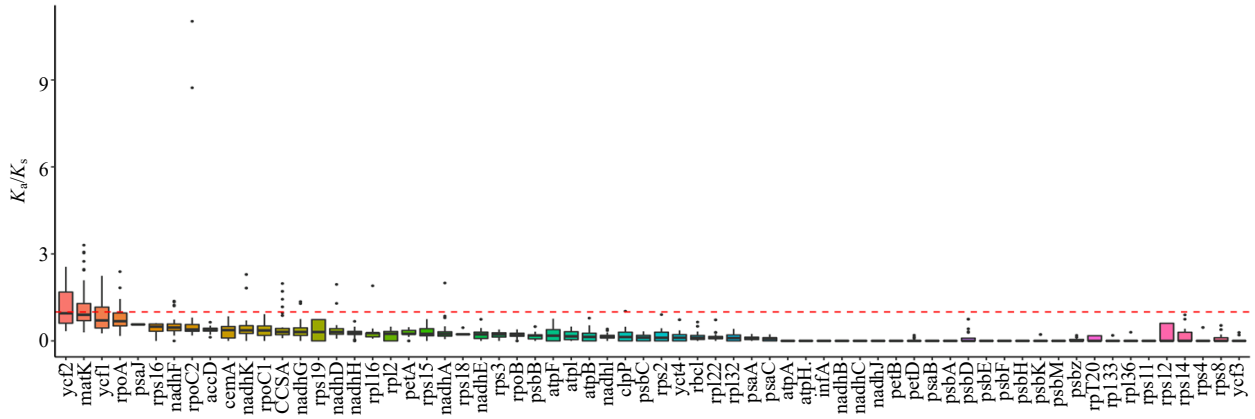


图 11　大黄属 CP 蛋白编码基因 K_a/K_s 值分布

Fig. 11　Distribution of K_a/K_s ratios for CP protein coding genes in *Rheum*

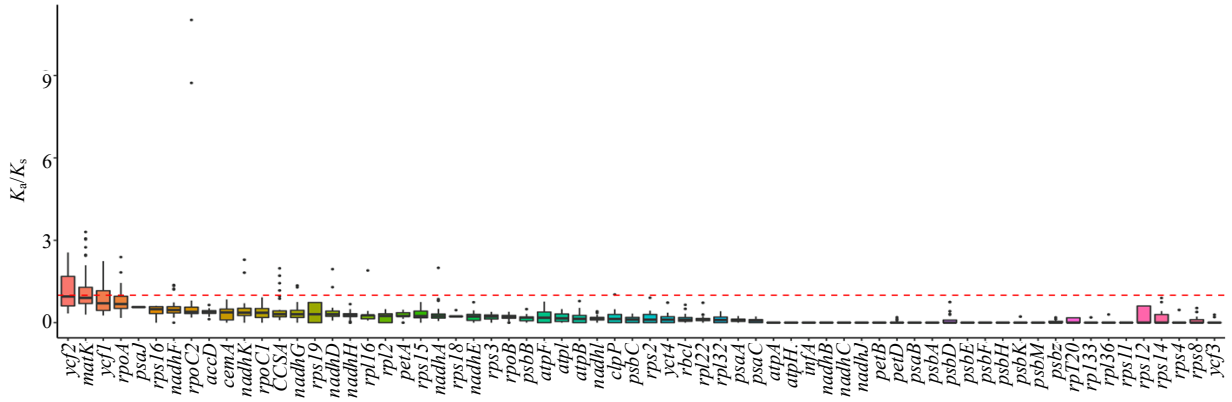


图 12　差异基因 K_a/K_s 值分布

Fig. 12　Distribution of K_a/K_s ratios for differential genes

基因在维持核心生理功能、受纯化选择约束的同时，可能保留了与高光强、低温胁迫相关的低水平适应性变异。*ndh* 家族多数基因 K_a/K_s 整体处于较低水平，但 *nad hF*、*nad hK*、*nad hD*、*nad hH* 等部分成员存在离群点，提示该基因家族内部的选择压力可能发生了分化。

3.5　系统发育分析

系统发育分析结果显示，22 个大黄属植物形成两大独立分支，拓扑结构稳定、多数节点支持度高（图 13）。喜马拉雅大黄与心叶大黄、塔黄、穗序大黄、菱叶大黄构成一个亚支系；该支系进一步与小大黄、歧穗大黄、单脉大黄及滇边大黄等聚类，共同构成大黄属基部类群。该类群多数物种主要分布于青藏高原及喜马拉雅极端高海拔原始生境，植株低矮或无茎，叶基生、近革质，根状茎粗壮，保留了大黄属祖先的形态与生态特征[33]。

另一分支进一步形成 2 大亚支，第一亚支包括河套大黄、华北大黄、波叶大黄、苞叶大黄，亲缘关系较近；另一亚支包括唐古特大黄、掌叶大黄、药用大黄及总序大黄等。与上述基部类群不同，该

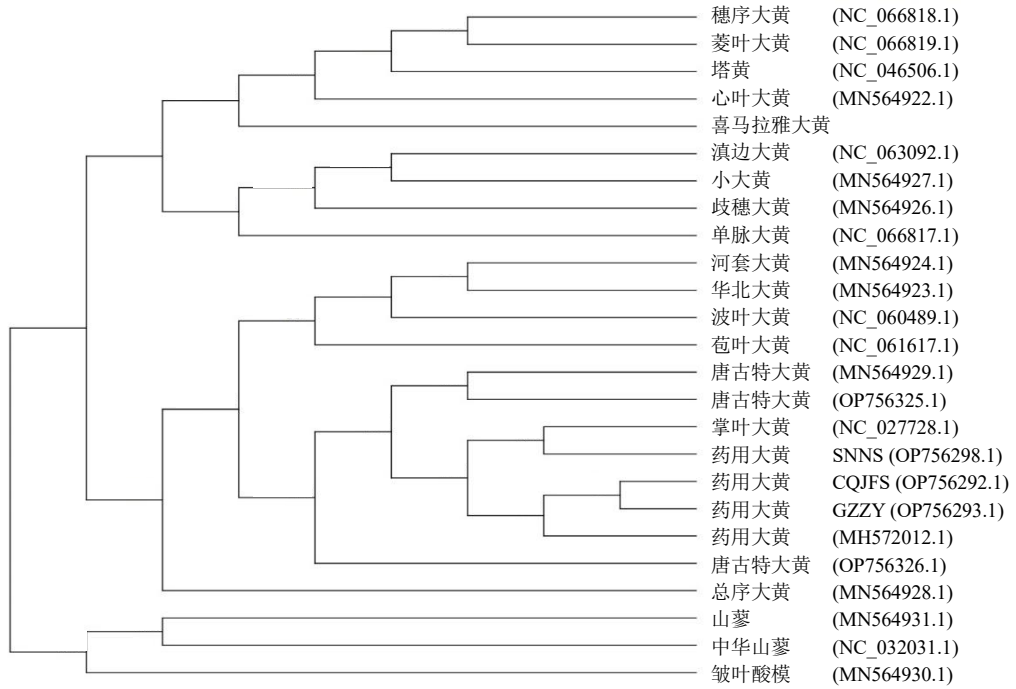


图 13　*CP* 基因组的系统进化分析

Fig. 13　Phylogenetic analysis of *CP* genomes

分支类群主要分布于横断山区、青藏高原东缘及华北、西北的温带至高山地带，相对低海拔地区；形态上表现出植株粗壮高大、叶片宽阔且多掌裂等特征，是为适应相对温和及多样化生境演化形成的特化性状[34]。结果暗示大黄属植物在物种演化过程中形成了青藏高原-喜马拉雅极端高海拔原始类群向横断山-温带广布类群拓展的多元分化格局。

4 讨论

CP 基因组因其结构紧凑、序列保守及重组率低等特点，成为研究植物进化与系统发育的理想工具[12]。被子植物 *CP* 基因组通常为环状四分体结构，长度介于 115～165 kb，但在不同科、属间存在差异[35]。本研究获得的喜玛拉雅大黄 *CP* 基因组全长 162 150 bp，注释基因 125 个。对比显示，其长度大于多个近缘种（如药用大黄的 161 093 bp 和掌叶大黄的 161 541 bp），属于大黄属中基因组规模较大的类型。

由于密码子的简并性，同一氨基酸可由多个同义密码子编码，而不同物种对这些密码子的使用频率存在显著差异，这种偏好性称为 RSCU[36]。本研究发现喜玛拉雅大黄共有 20 571 个密码子参与蛋白编码，与其他 9 种大黄属植物叶绿体比较，密码子数量处于中等水平，整体呈现显著的 A/T 偏向及第 3 位 A/U 结尾特征。利用中性绘图、PR2-plot 及 ENC-plot 进一步分析显示自然选择是塑造密码子使用偏好的主导因素，最优密码子为 UUU（Phe）和 UUA（Leu）。RSCU 分析显示，最高的密码子为 UUA（Leu），其次为 AUU（Ile）和 UUU（Phe），进一步证实首选密码子强烈倾向以 U/A 结尾。

本研究在喜马拉雅大黄 *CP* 基因组中鉴定出 48 个重复序列，以正向重复和回文重复为主，未发现反向重复与互补重复。在 SSR 方面，共检测到 263 个 SSRs 位点，其类型分布呈现明显偏斜：单核苷酸重复占主导，其次为三核苷酸和二核苷酸重复，四、五及六核苷酸重复数量较少，且绝大多数 SSR 属于 A/T 类型，占比达 68.44%。这一分布特征与大黄属唐古特大黄、药用大黄、掌叶大黄等叶绿体 SSRs 特征一致[16]。叶绿体 SSR 具有单亲遗传特征，在种质资源评价与遗传多样性研究中应用价值巨大，为物种鉴定与遗传研究提供了有效工具[37]。

基因组结构比较发现大黄属 IR 区比较稳定，没有明显的扩张和收缩现象，说明该 CP 整体上较为保守。基因组比较显示，大黄属 IR 区稳定，无明显扩张或收缩，且边界基因（*rps19*、*ndhF*、*ycf1*、*rpl2*）位置高度一致，说明 CP 整体保守。绝大多数基因 K_a/K_s 远小于 1，说明这种稳定性主要依赖结构保守和强烈的纯化选择，而非 IR 区的动态重组。光合核心基因 psbA 和 rbcL 的 K_a/K_s 均较低，表明核心功能高度保守，但可能积累少量与高光强、低温相关的适应性氨基酸替换。ndh 家族基因编码叶绿体 NADH 脱氢酶复合体，参与循环电子传递，缓解强光抑制，对抵御高海拔紫外损伤至关重要[38]，注释显示喜马拉雅大黄 ndh 家族完整保留，未缺失或假基因化；选择压力分析发现多数 *ndh* 基因受纯化选择，但 *ndhF*、*ndhK*、*ndhD*、*ndhH* 出现高 K_a/K_s 离群值，提示家族内选择压力分化。完整的 *ndh* 基因维持循环电子传递通路，推测部分成员的序列变异有助于适应高原低温强光胁迫。Pi 分析显示大黄属基因间隔区的整体核苷酸多样性水平显著高于编码区，发现 3 个显著变异热区域，分别为 *trnH-GUG-psbA*、*rpoC1-CDS1-rpoB*、*trnG-UCC-rps14*，可根据热点区序列开发分子标记，为大黄属物种鉴定和系统发育分析提供支持。

青藏高原是全球面积最大、平均海拔最高的独特高原地貌，孕育了极其丰富的高山植物资源，是研究植物物种起源、分化与适应性进化的关键区域[39-40]。已有研究表明青藏高原是大黄属植物的主要起源与分化中心，高原隆升与气候波动显著推动了该属物种的辐射演化与地理扩散[41-42]。本研究基于已发表的 22 个大黄属 *CP* 基因组构建系统发育树，显示大黄属植物存在清晰的谱系分化格局，呈现多元分化特征。高海拔基部类群以喜马拉雅大黄为代表，与小大黄、心叶大黄、塔黄等形成一大分支，长期生长高海拔、高寒、强紫外、低温贫瘠的极端生境[5]，保留了大黄属祖先性状，进化速率相对缓慢，体现了原始物种对高原极端环境的稳态适应策略。值得注意的是，单脉大黄虽分布于海拔 1 100～2 300 m 的温带干旱区，但仍嵌套于该基部类群内部，提示高海拔祖先已向低海拔干旱地带发生独立拓展。与之相比，另一大分支包括河套大黄、华北大黄、波叶大黄、唐古特大黄、掌叶大黄、药用大黄及总序大黄等。该类群分布范围更广、生境温湿条件更优，在长期地理扩散与生境适应过程中发生了明显形态特化。

从时间维度看，大黄属冠群的分化时间约为中

新世中期至晚期（20～10 Ma），与青藏高原主隆升期高度吻合，暗示该属的起源与高原隆升驱动的生境异质化及气候冷却密切相关[42-44]。进一步的研究表明，大黄属最近共同祖先的最可能分布区为青藏高原高海拔区域，而向横断山-西南（如药用大黄支系）、西北（如掌叶大黄支系）及中东部（如唐古特大黄支系）的多次扩散事件分别发生于晚中新世至更新世[45-46]，对应高原隆升后期及东亚季风气候的强化阶段，在此过程中，大黄属逐步向周边扩散形成次级多样化中心，呈现多元化格局，出现较高的物种丰富度。说明晚新生代青藏高原阶段性快速隆升及其引发的气候变化，是驱动大黄属物种谱系分化、地理扩散与形态适应性进化的关键动力。但本研究系统发育分析显示少数分支节点的自展支持率偏低（63%、68%、78%），部分近缘类群演化关系仍存在不确定性。

本研究首次解析了喜马拉雅大黄 *CP* 基因组，进行了基因组注释、密码子偏好性分析、重复序列分析及基因组结构比较等，鉴定 3 个显著变异热区域，明确了其亲缘关系。研究完善了大黄属药用植物质体基因组数据资源，为探究该属的起源、多样化及种质资源开发利用提供参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 中国科学院中国植物志编辑委员会. 中国植物志 [M]. 北京: 科学出版社, 1998: 176-177.

[2] 西藏植物志编辑委员会. 西藏植物志 (第 1 卷) [M].北京: 科学出版社, 1983: 599.

[3] 国家中医药管理局《中华本草》编委会.中华本草 [M]. 上海: 上海科学技术出版社, 2002: 128-129.

[4] Ahmad Wani I, Verma S, Ahmad P, *et al*. Reproductive biology of *Rheum webbianum* royle, a vulnerable medicinal herb from alpines of north-western Himalaya [J]. *Front Plant Sci*, 2022, 13: 699645.

[5] Ahmad Wani I, Verma S, Kumari P, *et al*. Ecological assessment and environmental niche modelling of Himalayan rhubarb (*Rheum webbianum* Royle) in northwest Himalaya [J]. *PLoS One*, 2021, 16(11): e0259345.

[6] 王彦青, 廖志新. 大黄属藏药植物资源与药用价值 [J]. 中国中药杂志, 2010, 35(17): 2305-2310.

[7] Kumai G P, Baskota M, Aryal P, *et al*. Study of phytochemical constituent and biological activities of methanolic extract of rhizomes of *Neopicrorhiza scrophulariiflora* and roots of *Rheum australe* collected from the Alpine region of Nepal [J]. *J Nepal Chem Soc*, 2023, 43(2): 130-140.

[8] Mao M J, Cao X Q, Liang Y H, *et al*. Neuroprotection of rhubarb extract against cerebral ischaemia-reperfusion injury via the gut-brain axis pathway [J]. *Phytomedicine*, 2024, 126: 155254.

[9] 邢少辰, Clarke J H. 叶绿体基因组研究进展 [J]. 生物化学与生物物理进展, 2008, 35(1): 21-28.

[10] Neuhaus H E, Emes M J. Nonphotosyntheticmetabolism inplastids [J]. *Annu Rev Plant Physiol Plant Mol Biol*, 2000, 51: 111-140.

[11] van Veen E, Küpers J J, Gommers C M M. Plastids in a pinch: Coordinating stress and developmental responses through retrograde signalling [J]. *Plant Cell Environ*, 2025, 48(9): 6897-6911.

[12] Wicke S, Schneeweiss G M, DePamphilis C W, *et al*. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function [J]. *Plant Mol Biol*, 2011, 76(3): 273-297.

[13] 杜新宇, 卢金梅, 李德铢. 石松类和蕨类植物质体基因组结构演化研究进展 [J]. 生物多样性, 2019, 27(11): 1172-1183.

[14] 赵志常, 高爱平, 黄建峰, 等. 叶绿体基因组在植物系统发育与基因工程中的应用 [J]. 植物学报, 2020, 55(3): 321-336.

[15] Zhao K H, Li L Q, Lu Y Z, *et al*. Characterization and comparative analysis of two *Rheum* complete chloroplast genomes [J]. *BioMed Res Int*, 2020, 2020: 6490164.

[16] 李冉郡, 武立伟, 辛天怡, 等. 大黄药材基原物种叶绿体基因组分析与特异 DNA 条形码开发 [J]. 药学学报, 2022, 57(5): 1495-1505.

[17] Paudel B R, Subedi C K, Ghimire S K, *et al*. Impacts of climate change on the distribution pattern of Himalayan Rhubarb (*Rheum australe* D. Don) in Nepal Himalaya [J]. *PLoS One*, 2025, 20(5): e0323755.

[18] 李荣华, 夏岩石, 刘顺枝, 等. 改进的 CTAB 提取植物 DNA 方法 [J]. 实验室研究与探索, 2009, 28(9): 14-16.

[19] Patel R K, Jain M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data [J]. *PLoS One*, 2012, 7(2): e30619.

[20] Bankevich A, Nurk S, *et al*. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing [J]. *J Comput Biol*, 2012, 19(5): 455-477.

[21] Hyatt D, Chen G L, Locascio P F, *et al*. Prodigal: Prokaryotic gene recognition and translation initiation site identification [J]. *BMC Bioinformatics*, 2010, 11: 119.

[22] Laslett D, Canback B. ARAGORN a program to detect tRNA genes and tmRNA genes in nucleotide sequences [J].

Nucleic Acids Res, 2004, 32(1): 11-16.

[23] 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(Web Server issue): W575-W581.

[24] Peden J F. Analysis of codon usage [D]. Nottingham: University of Nottingham, 1999.

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

[26] Thiel T, Michalek W, Varshney R, *et al*. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) [J]. *Theor Appl Genet*, 2003, 106(3): 411-422.

[27] 郑梦迪, 徐笑雨, 黄美琳, 等. 薯蓣属叶绿体基因组特征、密码子偏好性与机器学习分子鉴别研究 [J]. 中草药, 2026, 57(2): 640-651.

[28] 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.

[29] 刘巧珍, 代江鹏, 朱鹏坚, 等. 栽培奇楠沉香叶绿体基因组特征及系统发育分析 [J]. 中国中药杂志, 2023, 48(20): 5531-5539.

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

[31] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability [J]. *Mol Biol Evol*, 2013, 30(4): 772-780.

[32] Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies [J]. *Bioinformatics*, 2014, 30(9): 1312-1313.

[33] Zhang Y L, Sun H, Deng T, et al. Molecular phylogeny and biogeography of Rheum (Polygonaceae) in China [J]. *J Systemat Evol*, 2021, 59 (4): 789-802.

[34] Chen X, Wang A L, Liu J Q, *et al*. Origin of medicinal traits in *Rheum officinale* and *R. palmatum* [J]. *J Ethnopharmacology*, 2022, 285: 114823.

[35] Daniell H, Lin C S, Yu M, *et al*. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 2016, 17(1): 134.

[36] Wald N, Alroy M, Botzman M, *et al*. Codon usage bias in prokaryotic pyrimidine-ending codons is associated with the degeneracy of the encoded amino acids [J]. *Nucleic Acids Res*, 2012, 40(15): 7074-7083.

[37] Tan J, Li M, Wang H, *et al*. Development of a highly polymorphic chloroplast SSR set in *Abies grandis* [J]. *Ecol Evolut*, 2024, 14(6): e11593.

[38] 闫海锋, 韦毅刚. 被子植物叶绿体 *ndh* 基因的研究进展 [J]. 广西植物, 2024, 44(5): 769-782.

[39] Sun H, Zhang J W, Deng T, *et al*. Plant diversity and conservation in the Qinghai-Tibet Plateau [J]. *Biod Conservat*, 2014, 23(12): 3105-3122.

[40] Fortelius M, Eronen J, Liu L P, *et al*. Late Miocene and Pliocene large land mammals and climatic changes in Eurasia [J]. *Palaeogeogr Palaeoclimatol Palaeoecol*, 2006, 238(1/2/3/4): 219-227.

[41] Wang A L, Yang M H, Liu J Q. Molecular phylogeny, recent radiation and evolution of gross morphology of the rhubarb genus *Rheum* (Polygonaceae) inferred from chloroplast DNA trnL-F sequences [J]. *Ann Bot*, 2005, 96(3): 489-498.

[42] 孙航, 周浙昆. 青藏高原隆升与新生代植物演化、物种分化研究进展 [J]. 植物生态学报, 2011, 35(11): 1191-1202.

[43] Drummond A J, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees [J]. *BMC Evol Biol*, 2007, 7(1): 214.

[44] Favre A, Päckert M, Pauls S U, *et al*. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas [J]. *Biol Rev*, 2015, 90(1): 236-253.

[45] Ree R H, Smith S A. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis [J]. *Syst Biol*, 2008, 57(1): 4-14.

[46] Xing Y W, Ree R H. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot [J]. *Proc Natl Acad Sci USA*, 2017, 114(17): E3444-E3451.

[责任编辑 时圣明]