

基于高通量技术的唐古特大黄叶绿体全基因组测序及应用研究

左文明, 曾 阳, 杨春芳, 李美睢, 李锦萍*, 刘力宽*

青海师范大学生命科学学院 青海省青藏高原药用动植物资源重点实验室, 青海 西宁 810008

摘要: 目的 获得野生唐古特大黄叶绿体全基因组信息特征, 并对相关物种亲缘关系进行研究。方法 本实验采用 Illumina 高通量测序技术构建了唐古特大黄叶绿体全基因组图谱。结果 唐古特大黄基因组大小为 161 054 bp, 大 (LSC)、小 (SSC) 单拷贝区大小分别为 86 441 bp 和 12 745 bp, 反向互补重复区 (IR) 大小为 30 934 bp, 共注释叶绿体基因 132 个, 包括 88 个蛋白编码基因, 36 个转运 RNA 基因和 8 个核糖体 RNA 基因, 其中每个 IR 区 19 个。结论 选取唐古特大黄在内的 7 个蓼科物种、4 个其他科物种构建系统发育树, 形态极其相似的唐古特大黄与掌叶大黄在分子上亲缘关系也最近, 但 *rpl32* 等基因存在差异位点, 对有效区分近缘物种提供新依据。

关键词: 蓼科; 唐古特大黄; 高通量; 叶绿体; 叶绿体全基因组

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2019)22-5545-09

DOI: 10.7501/j.issn.0253-2670.2019.22.024

High-throughput sequencing of complete chloroplast genome of *Rheum tanguticum* and its application in species identification

ZUO Wen-ming, ZENG Yang, YANG Chun-fang, Li Mei-ju, LI Jin-ping, LIU Li-kuan

Qinghai Key Laboratory of Medicinal Animal and Plant Resources in Qinghai-Tibetan Plateau, College of Biological Science, Qinghai Normal University, Xining 810008, China

Abstract: Objective To improve the survival status of wild *Rheum tanguticum*, which is threatened to meet the market demand, Illumina high-throughput sequencing technologies is used to bring new directions for relevant research. **Methods** The complete chloroplast genome of *Rheum tanguticum* was constructed with Illumina high-throughput sequencing technologies in this paper. **Results** The genome was 161 054 bp in length, and exhibited a typical quadripartite structure of the large (LSC, 86 441 bp) and small (SSC, 12 745 bp) single-copy regions, separated by a pair of inverted repeats (IRs, 30 934 bp each). The chloroplast genome contained 132 genes, including 88 protein-coding genes, 36 transfer RNA genes, eight ribosomal RNA genes and two pseudogenes, 19 genes located in each IR area. **Conclusion** The phylogenetic tree constructed with 11 sequences of seven Polygonaceae species and four other family species demonstrated a close relationship between *R. tanguticum* and *R. palmatum* in Polygonaceae, which was coincided with their morphological similarity, in addition, there were certain SNP sites in *rpl32* and other genes, which provided a new basis for the effective identification of related species.

Key words: Polygonaceae; *Rheum tanguticum*; high-throughput; chloroplast; complete chloroplast genome

唐古特大黄 *Rheum tanguticum* Maxim. ex Balf., 又名鸡爪大黄, 君木扎(藏名), 蓼科(Polygonaceae) 大黄属 *Rheum* L. 多年生高大草本植物。适生于林缘、林下沟谷、高山灌丛、边坡石砾等生境, 分布于我国青海大部分地区、甘肃南部、四川西北部及青海西藏交界一带^[1-2], 1998 年谢宗强^[3]对国产大黄

属 8 个组 40 余种植物生态地理分布规律进行了研究, 从垂直分布规律发现包括唐古特大黄在内的 11 种植物分布幅度超过 2 000 m, 显示出较强的生态适应性, 而本课题组调查发现野生唐古特大黄于平均海拔 3 000~4 000 m 以上的果洛地区分布较为广泛。唐古特大黄含有大黄酚、大黄素、大黄酸、芦

收稿日期: 2019-03-06

基金项目: 国家自然科学基金资助项目(31360068); 青海省科技厅自然科学基金项目(2017-ZJ-742); 教育部春晖计划(Z2019070); 本科生科技创新项目(qhnuxskj2019014)

作者简介: 左文明(1992—), 男, 河南新郑人, 研究生, 研究方向为中藏药生物化学与分子生物学研究。E-mail: 740886849@qq.com

*通信作者 李锦萍(1974—), 女, 硕士生导师, 教授, 博士学位, 研究方向为资源生理生态学及资源化学研究。E-mail: 2008lj@163.com

刘力宽(1969—), 男, 硕士生导师, 副教授, 博士学位, 研究方向为系统进化与资源评价。E-mail: 2008llk@163.com

荟大黄素、大黄素甲醚、番泻苷 A、番泻苷 B 等多种化学成分^[4-6]。若用不同方法对生鲜大黄进行炮制,如炭炒、酒炒、清蒸、酒蒸、制成饮片等,大黄化学成分会有所差异,主要表现在高温或长时间加热条件下,蒽醌苷类成分含量不同程度地降低和游离蒽醌、没食子酸等鞣质类含量不同程度地升高^[7-9]。其根部及根状茎粗壮,外表呈现棕褐色,有褶皱,里为黄色,为药用部位,具有泻下攻积、清热泻火、凉血解毒、逐瘀通经、利湿退黄等功效^[2,10-11]。

叶绿体是植物进行光合作用的主要场所,作为植物遗传物质所在细胞器之一,其基因组在进化过程中高度保守,且多为母系遗传。叶绿体基因组较核基因组测序更为简易,具有一定技术优势,对于药用植物来说,从叶绿体基因组研究其遗传关系和进化特点,结果更为准确。之前受技术条件和成本限制,叶绿体基因组技术应用多集中于条形码技术分子鉴定与遗传进化^[12-16]方面,在唐古特大黄方面有关研究也较多^[17-21]。但随着高通量测序技术的发展,短片段研究显出一定局限,访问 NCBI 网站,越来越多的植物叶绿体全基因组现已完整测出,全基因组测序逐渐成为一种更高要求趋势。

自 1987 年第一代测序仪被发明以来, Sanger 测序法^[22]被沿用至今,生物信息学研究也进入了一个全新的领域,而随着其通量低、速度慢等缺点日渐凸显,第 2 代测序技术应运而生,目前,有 3 种主流测序技术,分别为 Roche 公司的 454^[23]焦磷酸测序平台、Illumina 公司的 Solexa^[24]边合成边测序平台和 ABI 公司的 SOLID^[25]边连接边测序平台,其中, Illumina HiSeq 为使用最广泛的测序系统。本研究采用 Illumina HiSeq X ten 测唐古特大黄叶绿体全基因组主要基于以下几点原因: (1) 就测序读长而言, Roche 454 GS FLX Titanium 能读取 500~800 bp, 而 Illumina HiSeq X ten 只读到 150 bp, 前者具有明显优势, 但考虑到前者测序错误率高、单碱基重复检出率低的缺点, 选择了后者; (2) 阅读诸多文献, 发现 Roche 454 对多种植物叶绿体基因组的测序基本均先提取其叶绿体 DNA^[26-27], 考虑到唐古特大黄道地产区青海果洛与本研究实验室路程遥远, 返回途中又难以保持其足够新鲜, 为防止发霉现象, 采用了硅胶干燥叶片的保存方法, 而硅胶干燥叶片很难达到叶绿体 DNA 的提取要求, 另外, 本实验室利用改良的植物总 DNA 提取方法得到的硅胶干燥唐古特大黄总 DNA 浓度和质量较好, 故

选择 Illumina HiSeq 测序平台, 提取样本总 DNA 即可; (3) 实验前已知与唐古特大黄形态极其相似的掌叶大黄叶绿体全基因组序列已被报道, 且同样为 Illumina HiSeq 测序平台完成, 仍选择这一平台既为后期二者在分子特征比较上保证了同一技术条件, 又很好地完成了覆盖率为 100% 的测序, 且直接跳过了由第 1 代测序方法重新设计引物来完成最后补洞这一过程, 大大节约了时间等各项成本。

唐古特大黄收录于世界自然保护联盟 (IUCN) 红色名录 (<http://rep.iplant.cn/>), 根据其国际评估标准, 该物种受威胁等级为易危 (Vulnerable, VU), 致危因子为生境退化或丧失。近年来因市场需求量过大, 野生资源过度采挖, 生境破碎化趋于严重。针对唐古特大黄资源现存状态, 加强唐古特大黄野生种质资源的保护、以优质人工栽培大黄资源平衡市场需求与野生资源保有量之间的矛盾两方面工作势在必行。高通量测序技术为研究唐古特大黄遗传多样性和遗传进化等内容带来了新思路, 但目前为止, 在与唐古特大黄亲缘关系较近的大黄属物种中, 只有掌叶大黄的叶绿体全基因组被研究报道^[28], 研究唐古特大黄叶绿体全基因组对于唐古特大黄选育优良品种、提高栽培资源质量、走可持续发展道路等后续研究具有重要意义。

1 材料

样品新鲜叶片采自于其主产区青海果洛的达日县满掌乡, 由青海师范大学曾阳教授鉴定为唐古特大黄 *Rheum tanguticum* Maxim. ex Balf.

2 方法

2.1 总 DNA 提取及检测

取硅胶干燥的唐古特大黄叶片, 结合早期常规植物叶片 DNA 提取方法^[29], 采用改良后的博迈德植物基因组提取试剂盒法提取总 DNA, 改良主要内容: 将研磨好的植物组织粉末迅速转移到预先装有 400 μ L 溶液 A、20 μ L RNaseA (10 mg/mL) 和 5 μ L β -巯基乙醇的离心管后, 增加 65 $^{\circ}$ C 水浴过程, 时间为 1 h, 期间每隔 10 min 颠倒混匀 1 次。此法具有操作简便、快捷, 提取所得产物杂质少等优点。提取后的总 DNA NanoDrop2000 微量紫外分光光度计检测其浓度和纯度, 以便初步检验其是否符合高通量测序要求。

2.2 测序和序列拼接

样品基因组 DNA 经检测合格后, 用超声波方法对其进行片段化机械打断, 然后经片段化 DNA

纯化、末端修复、3'端加 A、测序接头连接、琼脂糖凝胶电泳进行片段大小选择、PCR 扩增、测序文库构建、文库质量检测等一系列程序,最终得到质检合格的 350 bp 的文库,并采用全基因组鸟枪法策略,用 Illumina HiSeq 系统进行双末端测序。以测序循环为单位,对单个样品所有 Reads 平行测序的碱基质量值^[30]做分布图,查看单个样品各个测序循环及整体的测序质量。

测序得到的原始图像数据文件经碱基识别(Base Calling)分析转化为以 FASTQ^[31]文件格式存储的原始数据(raw reads),其中包含测序序列(reads)的碱基信息以及其对应的测序质量信息。然后使用过滤软件 SOAPnuke1.3.0 对原始数据进行低质量序列、接头序列等的滤过,从而得到 clean reads,以保证信息分析结果的可靠性,结果仍然以 FASTQ 格式存储,用于后续分析、公共数据库提交、发表等。

使用 SPAdes3.10.1^[32]软件进行基因组拼接。将拼接结果与近缘物种掌叶大黄叶绿体基因组(NC_027728)进行 Blastn 比对,保留大于 500 bp 且覆盖度大于 90%的序列,确定候选序列。基于序列测序深度、PE reads 比对情况以及与近缘物种的比对顺序等,确定候选序列的连接关系,并进一步验证连接关系是否成环。最后,使用 Gapcloser1.12 补充序列中存在的 gap,最终得到完整的基因组序列。

2.3 基因组序列注释和提交

利用叶绿体的注释软件 DOGMA^[33](<http://dogma.ccb.utexas.edu/>)对唐古特大黄叶绿体全序列进行编码基因注释。根据叶绿体基因组注释结果利用 OGDRAW^[34]软件绘制唐古特大黄叶绿体全基因组基因图谱,并将序列数据和基因注释信息上传到 NCBI 网站数据库。

2.4 SSR 分析

利用 MISA1.0^[35]软件分析唐古特大黄叶绿体全基因组 SSR 位点。设置参数为单核苷酸单元的重复数 ≥ 8 ,二核苷酸、三核苷酸单元的重复数 ≥ 4 ,四核苷酸、五核苷酸和六核苷酸单元的重复数 ≥ 3 。

2.5 系统发育关系分析

从 NCBI 的叶绿体数据库(<https://www.ncbi.nlm.nih.gov/genome/bro-wse/>)下载蓼科大黄属掌叶大黄 *Rheum palmatum* L.: NC_027728.1; 蓼科大黄属天山大黄 *Rheum wittrockii* Lundstr.: NC_035950.1; 蓼科山蓼属中华山蓼 *Oxyria*

sinensis Hemsl.: NC_032031.1; 蓼科荞麦属苦荞麦 *Fagopyrum tataricum* (L.) Gaertn.: NC_027161.1; 蓼科荞麦属金荞麦 *Fagopyrum dibotrys* (D. Don) Hara.: NC_037705.1; 蓼科酸模属尼泊尔酸模 *Rumex nepalensis* Spreng.: AP017909.1; 桑科大麻属大麻 *Cannabis sativa* L.: KY084475.1; 石竹科剪秋罗属丝瓣剪秋罗 *Lychnis wilfordii* (Regel) Maxim.: NC_035225.1; 十字花科独行菜属玛卡 *Lepidium meyenii* Walp.: NC_034363.1 和藜科菠菜属菠菜 *Spinacia oleracea* L.: AJ400848.1 10 个物种的叶绿体全基因组,使用 mafft 7.397^[36]进行多序列比对,然后分别用 RAXML 8.2.4^[37]和 phyML^[38]软件建树并进行分析,在 RAXML 中,以 GTRCAT 为核酸模型,以 1 000 为重复次数,在 phyML 中,以 100 000 为重复次数,结果均为 nwk 格式,可用 mega^[39]软件打开。

3 结果与分析

3.1 总 DNA 提取

测得唐古特大黄总 DNA 质量浓度为 130.3 ng/ μ L, $A_{260/280}$ 为 1.88,质量浓度和纯度均较高,且琼脂糖凝胶电泳图中 DNA 条带较亮,无杂带,总体来看,本实验提取的唐古特大黄总 DNA 质量较好。

3.2 测序数据

利用高通量测序平台 Illumina HiSeq 测得唐古特大黄叶绿体基因组覆盖率达到 100%,共获得 6.71 G raw data 测序量,44 727 396 个原 bin 始 reads,GC 含量为 41.59%,经过数据滤过, clean data 数据量为 6.08 G,40 552 584 个 reads 参与基因组拼接,确定候选序列 128 256 bp,GC 含量为 36.19%,得到 9 个有效 contig, N50 为 38 999 bp, N90 为 7 388 bp,最大 contig 为 43 056 bp,最小 contig 为 419 bp,平均 contig 大小为 14 250.67 bp。确定候选序列连接关系后,共得到连接好的序列 161 047 bp, gap 数共有 16 个,补完洞最终得到 161 064 bp 序列,GC 含量为 37.35%。

3.3 唐古特大黄叶绿体基因大小

唐古特大黄叶绿体全基因组呈环状双链,序列全长为 161 054 bp,其中大单拷贝区(large single-copy region, LSC)长度及始末位置为 86 441 bp: 1~86 441,小单拷贝区(small single-copy region, SSC)长度及始末位置为 12 745 bp: 117 376~130 120,反向重复序列 A (inverted repeat, IRa)和反向重复序列 B (inverted repeat, IRb)长度均为

30 934 bp, 始末位置分别为 130 121~161 054 和 86 442~117 375。

3.4 唐古特大黄叶绿体基因注释、基因功能及分类

统计表明, 唐古特大黄叶绿体基因组共包含 132 (112 种) 个基因 (图 1), 主要分为 3 大类, 即 88 个 (79 种) 蛋白编码 (protein-coding gene, PCG) 基因, 36 个 (29 种) 转运 RNA (transfer RNA, tRNA) 基因, 8 个 (4 种) 核糖体 RNA (ribosomal RNA, rRNA) 基因。其中, 2 个反向重复序列区分别有 19 个基因, 为互相拷贝。在所有蛋白编码基因中, rps16、rpl20 2 个 (2 种) 基因因许多突

变而失去正常功能, 成为假基因 (pseudogene), 此外, 与光合作用相关 (photosystem I、photosystem II、cytochrome b/f complex、ATP synthase、NADH dehydrogenase、RubisCO large subunit) 基因有 45 个 (44 种)、其他蛋白编码 (RNA polymerase、ribosomal proteins-SSU、ribosomal proteins-LSU、Protease、maturase、Envelop membrane protein、c-type cytochrome synthesis gene、Translational initiation factor) 基因 35 个 (29 种)、未知功能基因 hypothetical chloroplast reading frames (ycf) 6 个 (4 种) (表 1)。

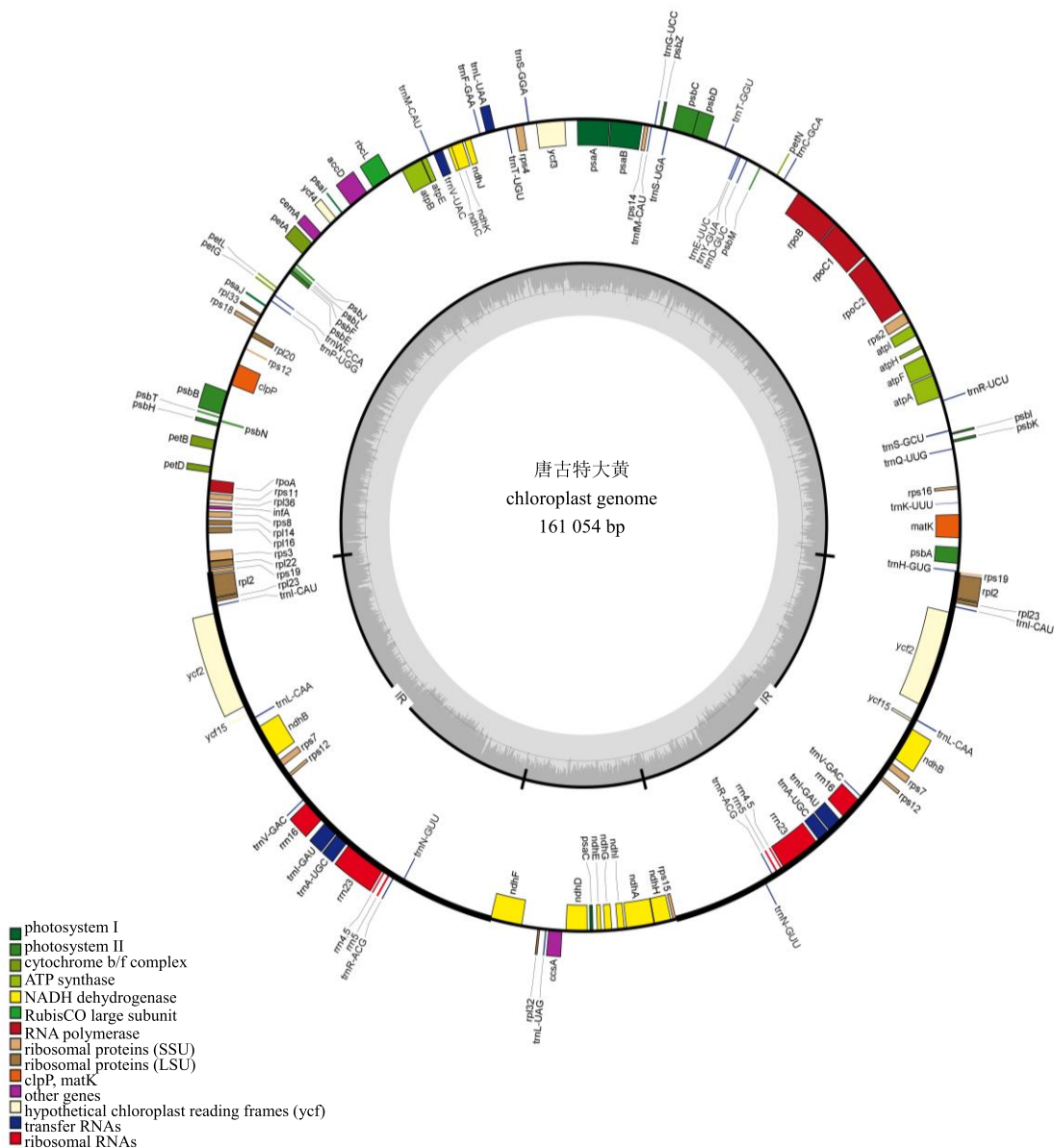


图 1 唐古特大黄叶绿体基因组图谱

Fig. 1 Gene map of *R. tanguticum* chloroplast genome

表 1 唐古特大黄叶绿体基因组基因分类

Table 1 List of genes of *R. tanguticum* chloroplast genome

基因组群	基因名称
photosystem I	psaJ、psaB、psaI、psaA、psaC
photosystem II	psbT、psbZ、psbI、psbJ、psbH、psbB、psbF、psbD、psbN、psbC、psbE、psbK、psbL、psbM、psbA
cytochrome b/f complex	petL、petA、petD、petB、petG、petN
ATP synthase	atpH、atpI、atpB、atpE、atpF、atpA
NADH dehydrogenase	ndhA、ndhB、ndhC、ndhD、ndhE、ndhF、ndhG、ndhH、ndhI、ndhJ、ndhK
rubisCO large subunit	rbcL
RNA polymerase	rpoB、rpoC1、rpoC2、rpoA
ribosomal proteins (SSU)	rps2、rps3、rps4、rps7、rps8、rps11、rps12、rps14、rps15、rps18、rps19
ribosomal proteins (LSU)	rpl2、rpl14、rpl16、rpl22、rpl23、rpl32、rpl33、rpl36
protease	clpP
maturase	matK
envelop membrane protein	cemA
subunit of acetyl-CoA-carboxylase	accD
c-type cytochrome synthesis gene	ccsA
hypothetical chloroplast reading frames (ycf)	ycf4、ycf2、ycf3、ycf15
transfer RNAs	trnM-CAU、trnH-GUG、trnC-GCA、trnF-GAA、trnN-GUU、trnD-GUC、trnI-CAU、trnE-UUC、trnS-GGA、trnA-UGC、trnV-GAC、trnM-CAU、trnK-UUU、trnL-UAG、trnS-UGA、trnT-GGU、trnQ-UUG、trnL-UAA、trnY-GUA、trnS-GCU、trnR-UCU、trnW-CCA、trnV-UAC、trnI-GAU、trnT-UGU、trnL-CAA、trnP-UGG、trnR-ACG、trnG-UCC
ribosomal RNAs	rrn4.5、rrn23、rrn5、rrn16
translational initiation factor	infA
pseudogene	rps16、rpl20

3.5 SSR 分析

叶绿体简单重复序列 (simple sequence repeats, SSR) 是一种高效的分子标记, 具有叶绿体基因组结构简单、相对保守、单亲遗传、标记数量丰富、共显性遗传、重复性高等优点, 被广泛应用于遗传图谱、物种鉴定、遗传多样性、遗传变异等研究^[40-41]。

在唐古特大黄叶绿体全基因组中, 共搜索到 243 个符合条件的 SSR 位点。其中包括 163 个单核苷酸重复基序, 60 个二核苷酸重复基序, 11 个三核苷酸重复基序, 8 个四核苷酸重复基序和 1 个五核苷酸重复基序, 未发现六核苷酸重复基序。在所有 SSR 中, 最多重复基序是 A/T, 然后依次是 AT/AT、AG/CT、AAT/ATT 和 AAAT/ATTT, 这些重复基序占总数的 95.47% (表 2)。另外, 从碱基组成来看, 多富含 AT, 与唐古特大黄叶绿体全基因组 AT 含量一致 (62.65%)。

3.6 石竹亚纲物种 IR 区边界比较分析

叶绿体基因组大小的差异主要体现在 IR 区边界的收缩与扩展^[42], 4 个 IR 区边界为 IRb-LSC、IRb-SSC、IRa-SSC 和 IRa-LSC。通过对石竹亚纲石竹目、蓼目、白花丹目 3 个目 6 个物种的叶绿体基因组 IR 区边界进行比较, 发现 IR 区长度与物种的叶绿体基因组全长呈正相关, 长度从大到小依次为蓝花丹、掌叶大黄、唐古特大黄、中华山蓼、丝瓣剪秋罗、头序麦瓶草。除丝瓣剪秋罗外, 其他物种的 IRb-LSC 边界均在 rps19 上, 且 rps19 基因序列大部分在 LSC 区, 大部分物种的 IRb-SSC 边界处于 ndhF 上, 且 ndhF 基因序列大部分在 SSC 区。在 IRa-SSC 边界处, 唐古特大黄的 ycf1 基因由于存在 SNP 位点, 测序未检测出该基因。除头序麦瓶草、唐古特大黄外, 其他物种的 ycf1 基因均在 IRa 区内, 且与 IRa-SSC 边界间隔较小, 为 1~282 bp, IRa-LSC 边界与 LSC 上的 trnH 连续或间隔极小, 只有 1~5 bp。示意图见图 2。

表 2 唐古特大黄叶绿体基因组中 SSR 位点类型及数量

Table 2 Type and number of SSR loci of *R. tanguticum* chloroplast genome

SSR 重复基序	长度/bp	SSR 数量/个	占比/%
单核苷酸			
A/T	8	85	99.39
	9	39	
	10	23	
	11	3	
	12	6	
	13	2	
	15	1	
	16	2	
	17	1	
C/G	8	1	0.61
二核苷酸			
AC/GT	4	2	3.33
	4	16	28.33
	5	1	
AT/AT	4	26	68.34
	5	14	
	6	1	
三核苷酸			
AAG/CTT	4	2	18.18
AAT/ATT	4	5	81.82
	5	3	
	6	1	
四核苷酸			
AAAT/ATTT	4	3	37.50
AACC/GGTT	4	1	12.50
AACT/AGTT	5	1	12.50
AATC/ATTG	4	1	12.50
AATT/AATT	4	1	12.50
ACAG/CTGT	4	1	12.50
五核苷酸			
AAGAT/ATCTT	4	1	100.00

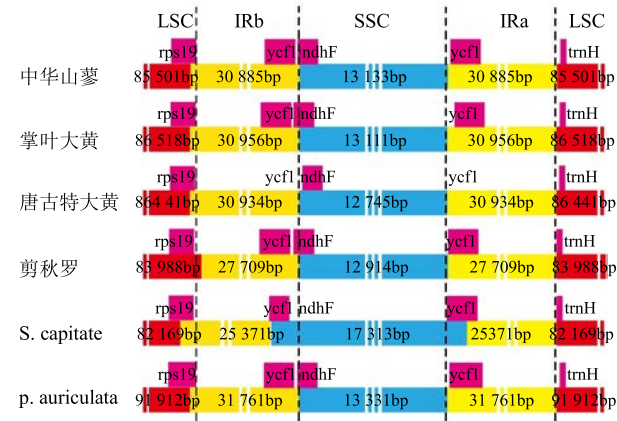


图 2 石竹亚纲 6 个物种叶绿体全基因组 IR 边界的比较示意图

Fig. 2 Comparison of large single copy, inverted repeat, and small single copy border regions among six Caryophyllidae chloroplast genomes

3.7 唐古特大黄与同属 2 物种叶绿体基因组的比较

唐古特大黄与同为大黄属的 2 个物种掌叶大黄和天山大黄在叶绿体基因组结构上基本相同, 除唐古特大黄与掌叶大黄的 LSC 长度类似, 且均比天山大黄长外, 3 物种的 SSC 区和 IR 区长度均相差不大, 组成基因数量与种类相差也不大 (表 3)。

3.8 亲缘关系分析

将唐古特大黄叶绿体基因组序列上传到 NCBI 网站基因库, 得到序列号 MK248631, 与下载的其他 10 个物种叶绿体全基因组序列构建系统发育树 (图 3、4)。此次分析选取了在分类上与唐古特大黄同为大黄属的 2 个物种掌叶大黄和天山大黄, 同为蓼科的 4 个物种苦荞麦、金荞麦、尼泊尔酸模和中华山蓼, 不同科的 4 个物种剪秋罗、大麻、玛卡和菠菜。从 2 个图中一致得出以下结论: (1) 所选其他 4 科物种均与蓼科距离较远, 说明亲缘关系较远, 这与植物志植物分类结果相一致; (2) 包括唐古特

表 3 大黄属 3 物种叶绿体基因组结构比较

Table 3 Comparison of chloroplast genome features of three species of *Rheum*

项目	唐古特大黄	掌叶大黄	天山大黄
总长度/bp	161 054	161 541	159 051
GC/%	37.35	37.3	37.6
LSC/bp	86 441	86 518	84 750
SSC/bp	12 745	13 111	12 999
IR/bp	30 934	30 956	30 651
基因总数	132	131	130
蛋白基因	86	84	84
rRNA 基因	8	8	8
tRNA 基因	36	37	36
Pseudogene	2	2	2

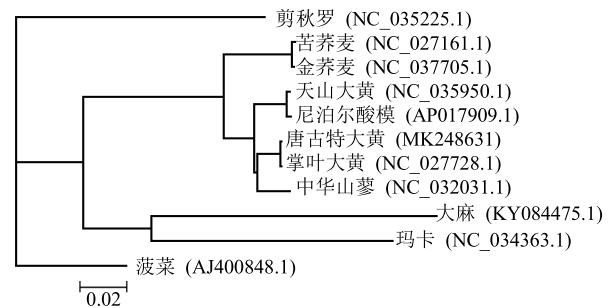


图 3 基于 11 个叶绿体全基因组序列对唐古特大黄系统发育树的构建 (RAXML 8.2.4 软件)

Fig. 3 Phylogenetic tree of *R. tanguticum* based on 11 complete chloroplast genomes (RAXML 8.2.4 software)

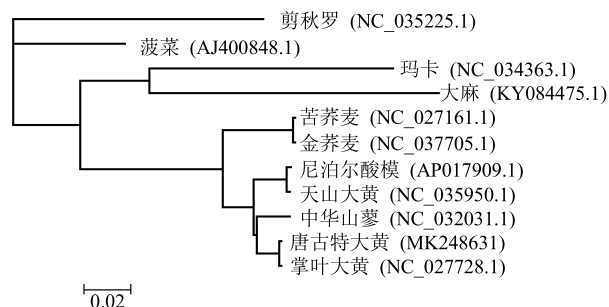


图 4 基于 11 个叶绿体全基因组序列对唐古特大黄系统发育树的构建 (phyML 软件)

Fig. 4 Phylogenetic tree of *R. tanguticum* based on 11 complete chloroplast genomes (phyML software)

大黄在内的 7 个蓼科物种集为一支, 且唐古特大黄与掌叶大黄的亲缘关系最近, 这与它们在形态上也极为相似相一致; (3) 同为大黄属的天山大黄却与酸模属的尼泊尔酸模集为一支, 与唐古特大黄亲缘关系相较较远, 分析可能与两物种所处的地理位置相差较大等因素有关, 相关结论有待进一步研究; (4) 相对于唐古特大黄和天山大黄的关系, 唐古特大黄与山蓼属中华山蓼亲缘关系较近, 这为相关物种系统分类提出了新要求。

4 讨论

不同测序平台、同一测序平台不同型号仪器、同一测序平台不同数据处理方式等均会造成不同程度数据差异, 在本研究基因注释、基因功能与分类这部分内容中主要存在以下几点讨论: (1) 测序平台在测序时, 遇到内含子和开放阅读框等结构时, 基因会被重新再计, 所以统计基因数量与种类时要对原始数据进行分析处理, 本研究起初测序得到 111 个 mRNA 基因、45 个 tRNA 基因和 8 个 rRNA 基因, 共 164 个基因, 而经过分析处理后实际得到 132 个, 前后相差较大, 其中, 实为拷贝的基因确实要再计 1 次, 此外, 还存在有些开放阅读框因编码功能未知而被单独列出的情况^[43], 这些均没有统一标准, 阅读者在阅读时要根据原始序列深入分析; (2) 测序完成后有些基因位置若包含在其他基因内部, 要进行整合取舍, 如在 28 858~28 929 序列标识为 *trnF*-GAA 基因, 而其区间 28 874~28 920 为 *trnF*-GAA 基因, 在计算基因种类和数量时便只计了 *trnF*-GAA 基因, 此类现象较多, 基因注释图谱等难免有所差异; (3) 核苷酸序列中终止密码子的提前出现或移码突变使得某些功能基因变成假基因, 但是不同的移码阅读起始得到的假基因不

同, 假基因的准确认定也需要研究者仔细分析序列信息。

从 NCBI 网站下载掌叶大黄叶绿体全基因组序列 (NC_027728.1、KR816224.1), 与高通量测序所得唐古特大黄叶绿体全基因组进行序列比对, 发现 *rpl32* 基因中 SNP 位点最多, *ndhF*、*rpl16*、*rpl22*、*rpl23*、*rps3*、*rpoA*、*rpoC2*、*petB* 等基因中较少, *petA*、*rps4*、*ndhK*、*accD*、*psbL*、*psbK*、*atpI* 等基因中只有 1 个, 剩余大多数基因在 2 个物种中序列完全一致, 没有差异。从上述结果中可得到 3 方面启示: (1) 可尝试通过 *rpl32* 基因的扩增测序来区分形态相似度较高的唐古特大黄和掌叶大黄植株, 在此基础上进行主要工作; (2) 可进行唐古特大黄和掌叶大黄系统进化亲缘关系研究, 看 2 个物种中是否存在变种关系; (3) 目前市面上多有唐古特大黄的混淆品或伪品, 而传统 DNA 条形码鉴别度不高, 相较单基因, 研究叶绿体全基因组能得到更高水平的种内差异, 这为唐古特大黄种群水平的物种鉴定、亲缘关系研究提供了巨大潜能及有力的技术支持。

另外, 叶绿体作为植物进行光合作用的场所, 在植物的整个生命过程中具有重要意义, 而研究其整个基因组, 可以为深入研究植物光合调控过程、抗性条件改善等内容奠定基础, 进而推动植物在分子育种、遗传转化等叶绿体基因工程方面的研究进展。

最后, 因唐古特大黄具有多种重要的药用价值, 市场需求量大, 造成唐古特大黄生境遭受不同程度的破坏, 野生资源匮乏, 研究其叶绿体全基因组, 一方面可为栽培高质量唐古特大黄物种提供分子依据, 另一方面可通过研究其分子结构与地理结构相关关系, 扩大野生唐古特大黄适宜生存范围, 进而达到保护野生唐古特大黄的目的。

参考文献

- [1] 中国科学院西北高原生物研究所. 青海植物志 [M]. 西宁: 青海人民出版社, 1997.
- [2] 中国药典 [S]. 一部. 2015.
- [3] 谢宗强. 国产大黄属植物的生态地理分布 [A] // 面向 21 世纪的中国生物多样性保护——第三届全国生物多样性保护与持续利用研讨会论文集 [C]. 昆明: 中国科学院生物多样性委员会, 1998.
- [4] Jin W, Ge R L, Huang Z Q, *et al*. Chemical Constituents of *Rheum tanguticum* Maxim. ex Balf. [J]. *J Chin Pharm Sci*, 2006, 15(4): 443-446.
- [5] 高亮亮, 许旭东, 南海江, 等. 唐古特大黄化学成分研

- 究 [J]. 中草药, 2011, 42(3): 443-446.
- [6] Wang X M, Ren Y. *Rheum tanguticum*, an endangered medicinal plant endemic to China [J]. *J Med Plants Res*, 2009, 3(13): 1195-1203.
- [7] 李 丽, 张 村, 肖永庆, 等. 大黄 5 种饮片化学成分的变化规律 [J]. 北京中医药大学学报, 2009, 32(12): 839-841.
- [8] 卫 昊, 冯改利, 郑 洁, 等. 清蒸和酒蒸对大黄中 9 种化学成分的影响分析 [J]. 中成药, 2013, 35(4): 777-780.
- [9] 祝婷婷, 刘 晓, 汪小莉, 等. 大黄不同方法炮制后药理作用及化学成分变化研究进展 [J]. 中国新药杂志, 2016, 25(8): 883-887.
- [10] 傅兴圣, 陈 菲, 刘训红, 等. 大黄化学成分与药理作用研究新进展 [J]. 中国新药杂志, 2011, 20(16): 1534-1538.
- [11] 杨 炆, 税丕先, 陈 滢, 等. 中药大黄在临床应用中的功效以及对其药理作用 [J]. 基因组学与应用生物学, 2017, 36(3): 1226-1231.
- [12] Chen T, Wang X R, Tang H R, *et al.* Genetic diversity and population structure of Chinese cherry revealed by chloroplast DNA trn Q-rps 16 intergenic spacers variation [J]. *Genet Resour Crop Evol*, 2013, 60(6): 1859-1871.
- [13] Guan-Mei Y I, Qiu Y J. Genetic diversity of *Torreya grandis* populations as revealed by chloroplast microsatellite markers [J]. *J Fruit Sci*, 2014, 31(4): 583-588.
- [14] 张靖国, 曹玉芬, 陈启亮, 等. 基于叶绿体 DNA 变异的湖北梨属种质系统进化及遗传多样性分析 [J]. 植物遗传资源学报, 2016, 17(4): 766-772.
- [15] 樊晓霞. 基于叶绿体 DNA 单倍型的银杏遗传多样性格局研究 [J]. 哈尔滨师范大学自然科学学报, 2014, 30(4): 104-107.
- [16] 刘 虹, 刘锡红, 刘秋宇, 等. 利用叶绿体基因组高变片段对 7 个软枣猕猴桃居群遗传多样性的研究 [J]. *Bot Res*, 2014, 3(6): 238-248.
- [17] Yang D Y, Fushimi H, Cai S Q, Komatsu K. Molecular analysis of *Rheum* species used as *Rhei Rhizoma* based on the chloroplast matK gene sequence and its application for identification [J]. *Biol Pharm Bull*, 2004, 27(3): 375-383.
- [18] 张晓芹, 刘春生, 闫兴丽, 等. 多基原药材大黄叶绿体 matK 基因序列分析及鉴定研究 [J]. 药学学报, 2013, 48(11): 1722-1728.
- [19] Zhang X Q, Liu Y, Gu X, *et al.* Genetic diversity and population structure of *Rheum tanguticum* (Dahuang) in China [J]. *Chin Med*, 2014, 9(1): 26-34.
- [20] Yang M H, Zhang D M, Liu J Q, *et al.* A molecular marker that is specific to medicinal rhubarb based on chloroplast trnL/trnF sequences [J]. *Planta Med*, 2001, 67(8): 784-786.
- [21] Wang A L, Li W W. Genetic diversity of *Rheum tanguticum* (Polygonaceae), an endangered species on Qinghai-Tibetan Plateau [J]. *Biochem Syst Ecol*, 2016, 69(1): 132-137.
- [22] Ner S S, Goodin D B, Pielak G J, *et al.* A rapid droplet method for Sanger dideoxy sequencing [J]. *Biotechniques*, 1988, 6(5): 408-412.
- [23] Marcel M, Michael E, Altman W E, *et al.* Genome sequencing in microfabricated high-density picolitre reactors [J]. *Nature*, 2005, 437: 376-380.
- [24] Shaffer C. Next-generation sequencing outpaces expectations [J]. *Nat Biotechnol*, 2007, 25(2): 149-149.
- [25] Ondov B D, Anjana V, Karla D P, *et al.* Efficient mapping of applied biosystems SOLiD sequence data to a reference genome for functional genomic applications [J]. *Bioinformatics*, 2008, 24(23): 2776-2777.
- [26] Dang Y Y, Yang Y, Li Q, *et al.* Complete chloroplast genome sequences of poisonous and medicinal plant datura stramonium: organizations and implications for genetic engineering [J]. *Chin Med*, 2014, 9(11): 110656-110656.
- [27] Wu M L, Li Q, Xu J, *et al.* Complete chloroplast genome of the medicinal plant *Amomum compactum*: Gene organization, comparative analysis, and phylogenetic relationships within Zingiberales [J]. *Chin Med*, 2018, 13(1): 10.
- [28] Fan K, Sun X J, Huang M, *et al.* The complete chloroplast genome sequence of the medicinal plant *Rheum palmatum* L. (Polygonaceae) [J]. *Mitochondrial DNA*, 2015, 27(4): 2935-2936.
- [29] Doyle J, Doyle J. A Rapid DNA isolation procedure from small quantities of fresh leaf tissues [J]. *Phytochem Bull*, 1987, 19(1): 11-15.
- [30] Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities [J]. *Genome Res*, 1998, 8(3): 186-194.
- [31] Cock P J A, Goto N. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants [J]. *Nucl Acids Res*, 2010, 38(6): 1767-1771.
- [32] Bankevich A, Nurk S, Antipov D, *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing [J]. *J Comput Biol*, 2012, 19(5): 455-477.
- [33] Wyman S K, Jansen R K, Boore J L. Automatic

- annotation of organellar genomes with DOGMA [J]. *Bioinformatics*, 2004, 20(17): 3252-3255.
- [34] Lohse M, Drechsel O, Kahlau S, *et al.* Organellar Genome DRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. *Nucl Acids Res*, 2013, 41(Web Server issue): 575-581.
- [35] Biggeri A, Baccini M, Bellini P, *et al.* Meta-analysis of the Italian studies of short-term effects of air pollution (MISA), 1990-1999 [J]. *Int J Occup Environm Health*, 2004, 11(1): 107-122.
- [36] Katoh K, Misawa K, Kuma K, *et al.* MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform [J]. *Nucl Acids Res*, 2002, 30(14): 3059-3066.
- [37] Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies [J]. *Bioinformatics*, 2014, 30(9): 1312-1313.
- [38] Guindon S, Delsuc F, Dufayard J, *et al.* Estimating maximum likelihood phylogenies with PhyML [J]. *Methods Mol Biol*, 2009, 537: 113-137.
- [39] Tamura K, Peterson D, Peterson N, *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods [J]. *Mol Biol Evol*, 2011, 28(10): 2731-2739.
- [40] Jiao Y, Jia H M, Li X W, *et al.* Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*) [J]. *BMC Genomics*, 2012, 13(1): 201.
- [41] Yang H, Li X, Liu D, *et al.* Genetic diversity and population structure of the endangered medicinal plant *Phellodendron amurense* in China revealed by SSR markers [J]. *Biochem System Ecol*, 2016, 66: 286-292.
- [42] Ravi V, Khurana J P, Tyagi A K, *et al.* An update on chloroplast genomes [J]. *Plant Syst Evol*, 2008, 271(1/2): 101-122.
- [43] Bodin S S, Kim J S, Kim J H. Complete chloroplast genome of *Chionographis japonica* (Willd.) Maxim. (Melanthiaceae): Comparative genomics and evaluation of universal primers for liliales [J]. *Plant Mol Biol Rep*, 2013, 31(6): 1407-1421.