

虎杖叶绿体基因组结构与变异分析

孙孟涛^{1,2}, 张峻鑫^{1,2}, 黄体冉^{1,2,3}, 杨明峰^{1,2}, 马兰青^{1,2}, 段留生⁴

1 北京农学院 林木分子设计育种高精尖创新中心 北京 102206

2 农业农村部华北都市农业重点实验室 北京 102206

3 中国科学院大学 生命科学学院, 北京 100049

4 北京农学院 植物科学技术学院, 北京 102206

孙孟涛, 张峻鑫, 黄体冉, 杨明峰, 马兰青, 段留生. 虎杖叶绿体基因组结构与变异分析. 生物工程学报, 2022, 38(5): 1953-1964.

SUN MT, ZHANG JX, HUANG TR, YANG MF, MA LQ, DUAN LS. Genome structure and variation of *Reynoutria japonica* Houtt. chloroplast genome. Chin J Biotech, 2022, 38(5): 1953-1964.

摘要: 虎杖 (*Reynoutria japonica* Houtt.) 为蓼科 (Polygonaceae) 蓼族 (Polygoneae) 虎杖属 (*Reynoutria* Houtt.) 植物, 是一种传统的中草药, 具有利湿退黄、清热解毒、散瘀止痛、止咳化痰的功效。本研究采取高通量测序技术获得 5 个虎杖品种的叶绿体全基因组序列, 并与 NCBI 已公布的蓼族何首乌 (*Fallopia multiflora*) 和金线草 (*Antenoron filiforme*) 等植物的叶绿体全基因组序列进行了基因组学和系统发育分析。通过基因组学分析发现, 5 种虎杖的叶绿体基因组大小有 163 376 bp 和 163 371 bp 两种情况, 并呈现出典型的环状四分体结构, 85 784 bp 的一条较长的单拷贝区 (large single-copy region, LSC), 18 616 bp 的一条较短的单拷贝区 (small single-copy region, SSC), 还有两条长度一致的反向重复区, 分别为 IRa 区和 IRb 区相间隔分布。通过注释得到 161 个基因, 其中蛋白编码基因 106 个, rRNA 编码基因 10 个, tRNA 编码基因 45 个。总 GC 含量为 36.7%, 其中 LSC、SSC 和 IR 区域的 GC 含量分别为 34.8%、30.7% 和 42.7%。对比不同虎杖品种叶绿体的全基因组分析发现, *trnk-UUU*、*rpoC1*、*petD*、*rpl16*、*ndhA*、*rpl12* 基因的编码区发生了变异。系统发育分析显示, 虎杖 5 个品种聚为一枝, 处在已知 11 种蓼族材料最原始的位置, 与何首乌构成姐妹群。

关键词: 中草药; 叶绿体全基因组; 基因组学分析; 环状四分体

Received: November 12, 2021; **Accepted:** March 2, 2022; **Published online:** March 10, 2022

Supported by: Beijing Municipal Education Commission under Grant (KM201910020016)

Corresponding authors: HUANG Tiran. Tel: +86-10-80799294; E-mail: huangtr151@bua.edu.cn
MA Lanqing. Tel: +86-10-80799046; E-mail: lqma@bac.edu.cn

基金项目: 北京市教育委员会科研计划一般项目 (KM201910020016)

Genome structure and variation of *Reynoutria japonica* Houtt. chloroplast genome

SUN Mengtao^{1,2}, ZHANG Junxin^{1,2}, HUANG Tiran^{1,2,3}, YANG Mingfeng^{1,2}, MA Lanqing^{1,2}, DUAN Liusheng⁴

1 Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing University of Agriculture, Beijing 102206, China

2 Key Laboratory for Northern Urban Agriculture of Ministry of Agriculture and Rural Affairs, Beijing University of Agriculture, Beijing 102206, China

3 College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

4 College of Plant Science and Technology, Beijing University of Agriculture, Beijing 102206, China

Abstract: *Reynoutria japonica* Houtt., belonging to Polygoneae of Polygonaceae, is a Chinese medicinal herb with the functions of draining dampness and relieving jaundice, clearing heat and detoxifying, dispersing blood stasis and relieving pain, and relieving cough and resolving phlegm. In this study, we carried out high-throughput sequencing for the chloroplast genome sequences of five cultivars of *R. japonica* and analyzed the genome structure and variations. The chloroplast genomes of the five *R. japonica* cultivars had two sizes (163 376 bp and 163 371 bp) and a typical circular tetrad structure composed of a large single-copy (LSC) region of 85 784 bp, a small single-copy (SSC) region of 18 616 bp, and a pair of inverted repeat (IR) regions (IRa/IRb) which are spaced apart. A total of 161 genes were obtained by annotation, which consisted of 106 protein-coding genes, 10 rRNA-coding genes, and 45 tRNA-coding genes. The total GC content was 36.7%. Specifically, the GC content in the LSC, SSC, and IR regions were 34.8%, 30.7%, and 42.7%, respectively. Comparison of the whole chloroplast genome among the five cultivars showed that *trnk-UUU*, *rpoC1*, *petD*, *rpl16*, *ndhA*, and *rpl12* in coding regions had sequence variations. In the phylogenetic tree constructed for the 11 samples of Polygoneae, the five cultivars of *R. japonica* clustered into one clade near the root and was a sister group of *Fallopia multiflora* (Thunb.).

Keywords: Chinese medicinal herb; whole chloroplast genome; genomics analysis; circular tetrad structure

虎杖 (*Reynoutria japonica* Houtt.) 又名酸筒杆、酸桶芦、大接骨和斑庄根, 为蓼科 (Polygonaceae) 蓼亚科 (Subfam.), 蓼族 (Trib. Polygoneae), 虎杖属 (*Reynoutria* Houtt.) 多年生草本植物。曾经被归入蓼属 (*Polygonum* L.), 命名为 *Polygonum cuspidatum* Sieb. & Zucc. 并在很长一段时间里被广泛使用。1998 年李安仁

在编纂《中国植物志》中将虎杖从蓼属中分离出来, 归为虎杖属^[1]。作为一种传统中草药, 虎杖具有除风、除湿、止痛、缓解咳嗽和减少痰量^[2]等功效, 而我国又是虎杖主要分布区, 因此, 开展虎杖系统发育定位和叶绿体基因结构的研究可为开发新型中草药资源提供理论依据。

系统发育是指生物的起源和进化。系统发育学是一门利用数学模型探索生物起源和进化关系的学科,进化关系主要以系统发育树的形式出现。构建出一个可信度高、涵盖面广的系统发育关系,为我们了解地球上生物多样性及演化关系提供了先决条件^[3-6]。20世纪60年代,Fitch 等对线粒体中细胞色素 C 蛋白的分析,找出动物和真菌间的不同之处,进而构建了动物和真菌间的系统发育关系^[7],这是以蛋白质序列作为构建系统发育关系的分子证据。随着科学技术的发展,聚合酶链式反应 (polymerase chain reaction, PCR) 技术逐渐为多数人接受,获取了大量的蛋白质序列和核酸 (DNA) 序列,以大批量的蛋白质序列和 DNA 序列作为分子证据构建系统发育树成为了主流方法^[8]。

叶绿体是植物细胞内最重要、最普遍的细胞器。叶绿体属于半自主性细胞器,具有自身的 DNA 和 RNA,能进行半自主复制,和线粒体一样具有一个相对独立的遗传体系。早在 1962 年,Ris 和 Plaut 在对莱茵衣藻 (*Chlamydomonas reinhardtii*) 的研究中首次观察到了叶绿体 DNA 的存在^[9],随后在 1963 年 Chun 等^[10]报道也发现了叶绿体 DNA 的存在。1986 年 Shinozaki 等成功提取烟草 (*Nicotiana tabacum* L.) 的 DNA,并获得叶绿体基因组全序列^[11]。2020 年杨嘉鹏等对 3 种石豆兰属植物的叶绿体基因组进行了组装注释,为系统发育分析和物种鉴定提供了理论依据^[12]。

随着对叶绿体基因组研究的深入,发现叶绿体基因组是裸露的环状双链 DNA 分子,以共价闭合形式存在,大多数植物的叶绿体基因组是由 4 个区域组成:间隔分布为一个长单拷贝序列区 LSC 区,相邻的是一个反向重复区 IRb 区,

再相邻的是一个短单拷贝序列区 SSC 区,最后接一个反向重复区 IRa 区^[13-14],构成一个封闭的环状结构。由于叶绿体基因组序列分子量小且分子进化速率慢、序列保守、不易发生基因突变等特点,使其能较好地应用于 DNA 条形码的开发、在植物的亲缘关系和遗传多样性及系统发育等方面发挥着重要作用^[15],现已成为分析系统发育关系最有效的手段之一。随着第三代测序技术的开发与应用,测定叶绿体基因组的全序列也越来越普遍^[16]。

2019 年 Zhang 等^[17]获取并组装了一个虎杖 (*Polygonum cuspidatum* Sieb. & Zucc.) 的基因组草案,但虎杖的叶绿体基因组还未见报道。为了更好地了解虎杖的生物学特征,寻找可替代野生中药资源,本研究拟通过高通量测序技术获取 5 个虎杖品种的叶绿体全基因组序列,进行基因组结构特征和序列变异分析,并将虎杖和同属蓼族的金线草 (*Antenoron filiforme* (Thunb.) Rob. et Vaut)、何首乌 (*Fallopia multiflora* (Thunb.) Harald.)、金荞麦 (*Fagopyrum dibotrys* (D. Don) Hara)、苦荞麦 (*Fagopyrum tataricum* (L.) Gaertn.)、荞麦 (*Fagopyrum esculentum* Moench)、小野荞麦 (*Fagopyrum leptopodum* (Diels) Hedb.) 的叶绿体全基因组进行了系统发育分析,为蓼族植物的系统归类和叶绿体基因组的结构变异分析提供理论依据。

1 材料与方法

1.1 材料

虎杖样本取自从东明格鲁斯生物科技有限公司引进并移栽于北京农学院苗圃的虎杖植株,取陕西 1 号、陕西 2 号、湖北 1 号、湖南 1 号、湖南 2 号植株叶片材料 (图 1)。

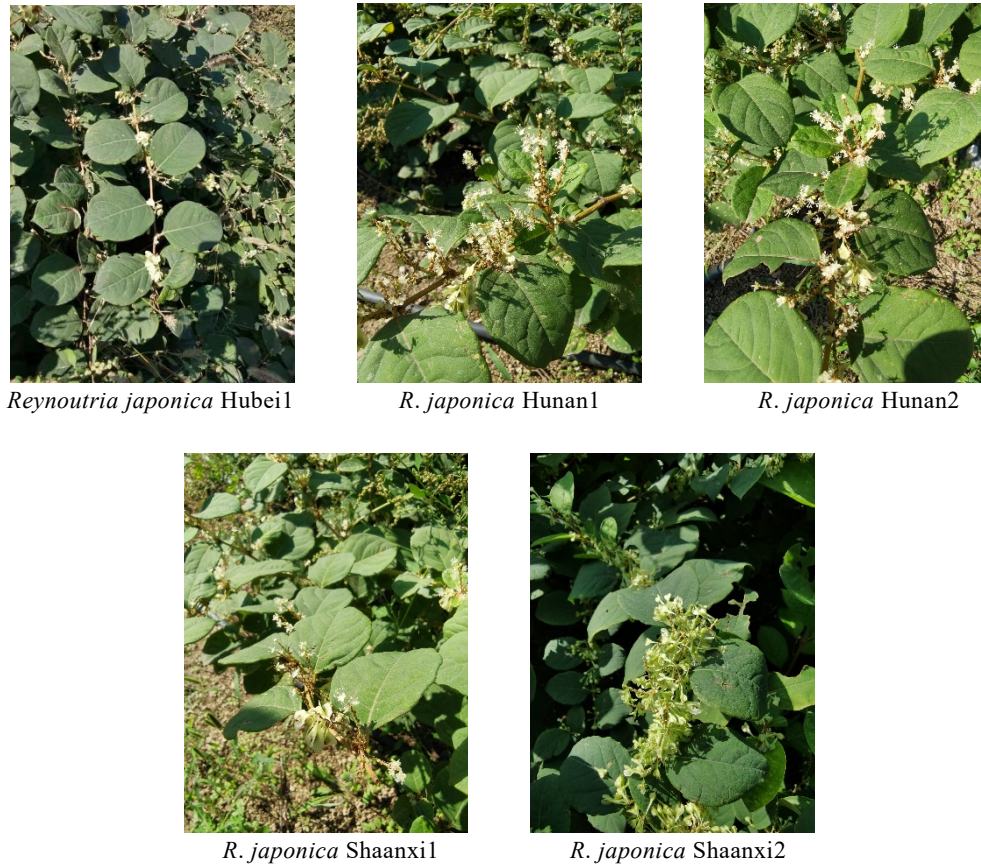


图1 虎杖5个品种栽培植株
Figure 1 Five cultivars of *Reynoutria japonica*.

1.2 DNA 提取与测序

用 DNeasy plant Kit of Aidlab 试剂盒提取叶绿体 DNA, 在 Illumina nova-seq 6 000 (Illumina Inc., San Diego, CA) 平台完成测序, 获得湖北 1、湖南 1、湖南 2、陕西 1、陕西 2 共 5 个不同品种虎杖的完整叶绿体基因组序列。将 5 个品种的虎杖 DNA 样本送至天津诺禾致源生物信息科技有限公司进行质量检验, 以安捷伦完整性检测。检测结果总体结论评价均为 A。A 类样本表示质量满足建库测序质量要求, 且总量满足 1 次或 1 次以上建库需要 (《生物工程学报》网络版附表 1)。

1.3 组装注释与统计

通过在线 GenBank 数据库 ([http://www.](http://www.ncbi.nlm.nih.gov/genbank)

[ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank)) 检索蓼族所有已发表的叶绿体全基因组序列, 共获得蓼族金线草 (*Antenoron filiforme* (Thunb.) Rob. et Vaut)、何首乌 (*Fallopia multiflora* (Thunb.) Harald.)、金荞麦 (*Fagopyrum dibotrys* (D. Don) Hara)、苦荞麦 (*Fagopyrum tataricum* (L.) Gaertn.)、荞麦 (*Fagopyrum esculentum* Moench)、小野荞麦 (*Fagopyrum leptopodum* (Diels) Hedb.) 6 个物种叶绿体全基因组序列, 物种名称、基因组登录号详见表 1。

以何首乌 (*Fallopia multiflora*, GenBank 登录号为 MK330002.1) 的叶绿体基因组作为对照, 将 5 个品种的虎杖和 6 种蓼族植物的叶绿体全基因组序列提交到 GeSeq ([https://chlorobox.](https://chlorobox)

mpimp-golm.mpg.de/geseq.html) 网站进行了基因组注释。

将注释结果用Geneious7.1.4软件 (<https://www.Geneious.com/>) 统计各序列的 GC 含量, LSC、IRb、SSC、IRa 区序列长度信息以及叶绿体基因组中 tRNA、rRNA、CDS 基因等各类型基因数目。

1.4 序列与结构变异分析

使用在线工具 OGDRAW (<https://chlorobox.mpimp-golm.Mpg.de/OGDraw.html>)^[18]绘制虎杖 5 个品种和金线草、何首乌、金荞麦、苦荞麦、荞麦、小野荞麦的叶绿体基因组结构图谱。利用 R 语言本地化运行 IRscope (<https://ir-scope.shinyapps.io/irapp/>)^[19]绘制叶绿体基因组边界分析图, 分析叶绿体基因组 4 个区连接位点的差异。利用本地化 python 程序获取叶绿体基因组 GB 文件的注释文件, 再将注释文件和叶绿体基因组序列提交到 mVISTA (<http://genome.lbl.gov/vista/index.sht-ml>)^[9]程序, 在 Shuffle-LANGAN 模式下做叶绿体全基因组比对, 并根据图分析叶绿体基因组序列变异情况。

1.5 系统发育分析

以蓼亚科的木蓼族 (Trib. Atraphaxideae Damm.), 沙拐枣属 (*Calligonum* L.), 基翅组 (Sect. *Calligonum* Borszcz.), 褐色沙拐枣 (*Calligonum*

colubrinum Borszcz., GenBank 登录号为 MN202602.1) 为外类群, 开展基于叶绿体全基因组序列的蓼族植物系统发育分析。利用 Geneious 7.1.4 软件将包含外类群在内的叶绿体全基因组进行序列比对, 将比对结果以 MEGA (*meg) 格式导出, 在 MEGA 7 中打开, 以 NJ 法 (neighbor-joining) 进行系统发育树的构建, 利用步长进行重复检验, 重复 5 000 次得到系统发育树。

2 结果与分析

2.1 叶绿体基因组结构的基本特征

虎杖叶绿体基因组图谱见图 2, 蓼族何首乌等 6 种植物的叶绿体基因组图谱见附图 1 (详见《生物工程学报》网络版附图 1)。蓼族现已发布 6 种植物 (表 1) 和实验室测定 5 个品种的虎杖 (图 1) 的叶绿体全基因组均具有典型的 4 分体结构, 排列分布为 1 个 LSC 区、1 个 IRb 区、1 个 SSC 区、1 个 IRa 区。

根据叶绿体基因组注释统计结果 (表 2) 分析, 5 个虎杖品种差异较小, 湖南 1、湖南 2 叶绿体基因组大小均为 163 371。湖北 1、陕西 1、陕西 2 均为 163 376, 相差 5 个 bp。CDS 基因、rRNA 基因和内含子基因数目均相同, 分别为 106、10、28。5 个虎杖品种的 tRNA 数目均为 45 个。GC 含量均为 37.5%。与蓼族其他 6 种植物相比, 何首乌的叶绿体基因组大小与虎杖相同, 金线草、金荞麦、苦荞麦、小野荞麦、荞麦的叶绿体基因组大小均小于虎杖, 最低的是苦荞麦为 159 272 bp, 与虎杖相差 4 104 bp。tRNA 数和 CDS 基因数与虎杖差异在 3 以内, rRNA 数与虎杖相同均为 10。四分体区长度波动较大, LSC 区长度为 84 397–88 340 bp, SSC 区长度为 13 073–13 887 bp, IR 区长度为 30 685–31 128 bp。GC 含量由小到大为 37.5%、37.8%、37.9%和 38.0%, 最大差异为 0.5%。

表 1 蓼族 6 种材料及叶绿体基因组 GenBank 登录号

Table 1 GenBank accession number of the chloroplast genomes in 6 species of Polygoneae

Species	GenBank No.
<i>Antenoron filiforme</i> (Thunb.) Rob. et Vaut.	MW915468.1
<i>Fallopia multiflora</i> (Thunb.) Harald.	MK330002.1
<i>Fagopyrum dibotrys</i> (D. Don) Hara	MF491390.1
<i>Fagopyrum tataricum</i> (L.) Gaertn.	KM201427.1
<i>Fagopyrum esculentum</i> Moench	EU254477.1
<i>Fagopyrum leptopodum</i> (Diels) Hedb.	MW017633.1

由叶绿体基因组注释统计结果 (表 2) 可以得出叶绿体基因组的基因数量较为保守。CDS 基因数为 104-106 个。转运 RNA (tRNA) 数量为 45-46 个, 苦荞麦最多为 46 个, 其余均为 45 个。内含子数为 25、27 或 28 个, 何首乌 25 个, 金线草和荞麦 27 个, 其余均为 28 个。

2.2 叶绿体基因组边界分析

叶绿体基因组是具有一个 LSC 区、一个 SSC 区、一个 IRa 区和一个 IRb 区相间分布构成的环状四分体结构, 同时存在 JLA (IRa-LSC)、JLB (LSC-IRb)、JSA (IRa-SSC) 以及 JSB (IRb-SSC) 4 个不同的边界分布。在物种

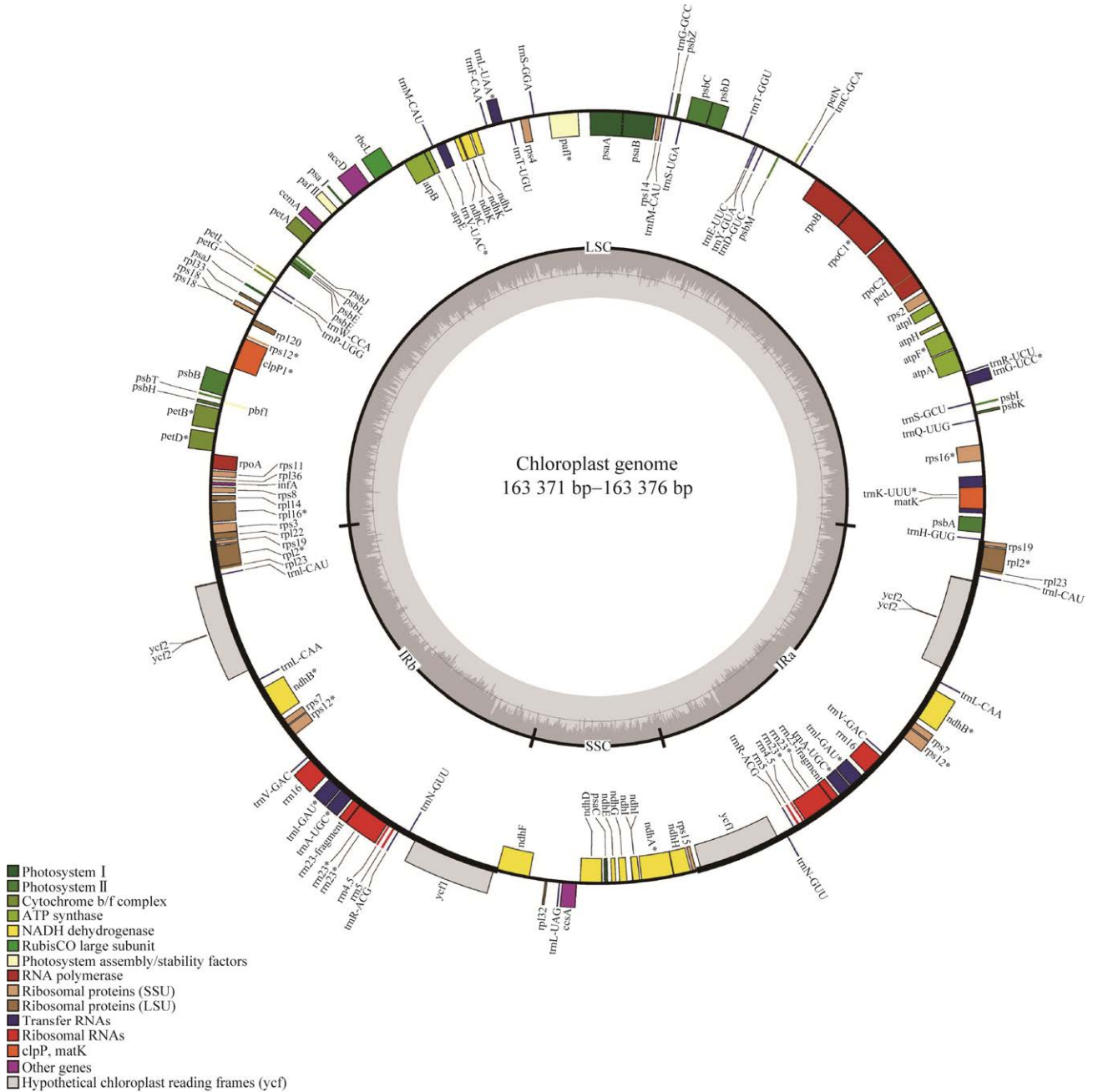


图 2 虎杖叶绿体基因组图谱
Figure 2 Chloroplast genome map of *Reynoutria japonica*.

表 2 叶绿体基因组注释统计数据表

Table 2 Chloroplast genome annotation statistics

Species	Genome size (bp)	GC content (%)	rRNA (No.)	tRNA (No.)	Intron (No.)	CDS genes (No.)	LSC (bp)	SSC (bp)	IRa (bp)	IRb (bp)
<i>Antenoron filiforme</i> (Thunb.) Rob. et Vaut.	159 741	37.8	10	45	27	105	84 432	13 073	31 118	31 118
<i>Fallopia multiflora</i> (Thunb.) Harald.	163 376	37.5	10	45	25	105	88 340	13 887	30 773	30 773
<i>Fagopyrum dibotrys</i> (D. Don) Hara	159 919	37.9	10	45	28	104	85 134	13 309	30 738	30 738
<i>Fagopyrum tataricum</i> (L.) Gaertn.	159 272	37.9	10	46	28	104	84 397	13 241	30 817	30 817
<i>Fagopyrum esculentum</i> Moench	159 599	38.0	10	45	27	104	84 885	13 344	30 685	30 685
<i>Fagopyrum leptopodum</i> (Diels) Hedb.	159 375	37.8	10	45	28	106	84 454	13 225	30 848	30 848
<i>Reynoutria japonica</i> Houtt. Hubei1	163 376	37.5	10	45	28	106	87 549	13 571	31 128	31 128
<i>Reynoutria japonica</i> Houtt. Hunan1	163 371	37.5	10	45	28	106	87 571	13 558	31 121	31 121
<i>Reynoutria japonica</i> Houtt. Hunan2	163 371	37.5	10	45	28	106	87 571	13 558	31 121	31 121
<i>Reynoutria japonica</i> Houtt. Shaanxi1	163 376	37.5	10	45	28	106	87 576	13 558	31 121	31 121
<i>Reynoutria japonica</i> Houtt. Shaanxi2	163 376	37.5	10	45	28	106	87 576	13 558	31 121	31 121

进化过程中叶绿体基因组也会发生部分变异，相对应的边界会发生扩张或收缩。尽管已有材料叶绿体基因组相对保守，在 GC 含量、序列长度方面发生变异较少，但不同分区边界在收缩扩张过程中还是发生了一定的改变。

叶绿体基因组边界分析显示 (图 3), 5 个虎杖品种 IRb-SSC 边界完全相同, 均位于 *ndhF* 基因上, 并将其分为 IRb 62 bp 和 LSC 2 182 bp 两部分。何首乌的 IRb-SSC 边界位于 *ndhF* 基因左侧, *ndhF* 基因完全位于 SSC 区域, 且 IRb-SSC 边界与 *ndhF* 基因间隔 56 bp。除何首乌之外的其他蓼族材料与虎杖保持一致, 均位于 *ndhF* 基因编码区上, IRb-SSC 边界将 *ndhF* 基因分为 IRb 区域、LSC 区域两部分, 金线草 IRb 62 bp、LSC 2 188 bp, 金荞麦 IRb 74 bp、

LSC 2 176 bp, 苦荞麦 IRb 90 bp、LSC 2 160 bp, 小野荞麦 IRb 54 bp、LSC 2 184 bp, 荞麦 IRb 73 bp、LSC 2 183 bp (图 3)。

5 个虎杖品种的 IRb-LSC 边界均位于 *rps19* 和 *rpl22* 的基因间隔区, 陕西 1 号、陕西 2 号、湖南 1 号、湖南 2 号的边界与 *rps19* 基因相差 25 bp。湖北 1 号的边界与 *rps19* 基因相差 31 bp。与蓼族其他材料比较显示, 金线草和 5 个品种虎杖材料的 IRb-LSC 边界均位于 *rps19* 基因和 *rpl22* 基因的基因间隔区。金线草的边界与 *rpl22* 基因相隔 25 bp。何首乌和荞麦、金荞麦、苦荞麦、小野荞麦的 IRb-LSC 边界与虎杖不同, 位于 *rps19* 基因上, 距离相差 5 bp 以内, 较为保守。

5 个虎杖品种的 IRa-SSC 边界均位于 *rps15*

基因与 *ycf1* 基因的基因间隔区, 陕西 1 号、陕西 2 号、湖南 1 号、湖南 2 号边界均距 *rps15* 基因 54 bp, 湖北 1 号距 *rps15* 基因 67 bp。除荞麦外, 其余 5 种材料和虎杖的 IRa-SSC 边界保持一致, 位于 *rps15* 基因与 *ycf1* 基因的基因间隔区, 何首乌与 *ycf1* 基因间隔 135 bp, 其余 4 种材料和虎杖的 IRa-SSC 边界与 *rps15* 基因间隔 3、4、27、37、54、67 bp 不等。荞麦的 IRa-SSC 边界位于 *rps15* 基因编码区上, 与 *rps15* 基因尾部间隔 2 bp (图 3)。

IRa-LSC 边界具有较大的差异性, 虎杖和其余 6 种材料的 IRa-LSC 边界均位于 *rps12* 基因与 *trnH* 基因之间。其中, 苦荞麦、何首乌、小野荞麦的 IRa-LSC 边界与 *trnH* 基因间隔 2 bp。金荞麦、荞麦则是在 *rps12* 基因与 *trnH* 基因之间又多了 *rps1* 基因, 且 *rps1* 基因与 *trnH* 基因紧密相连, IRa-LSC 边界位于两基因之间。5 个虎杖品种和金线草则是在 *rps12* 基因与 *trnH* 基因之间多了 *rps19* 基因, 同样与 *trnH* 基因紧密相连, 并以两基因之间作为 IRa-LSC 边界 (图 3)。

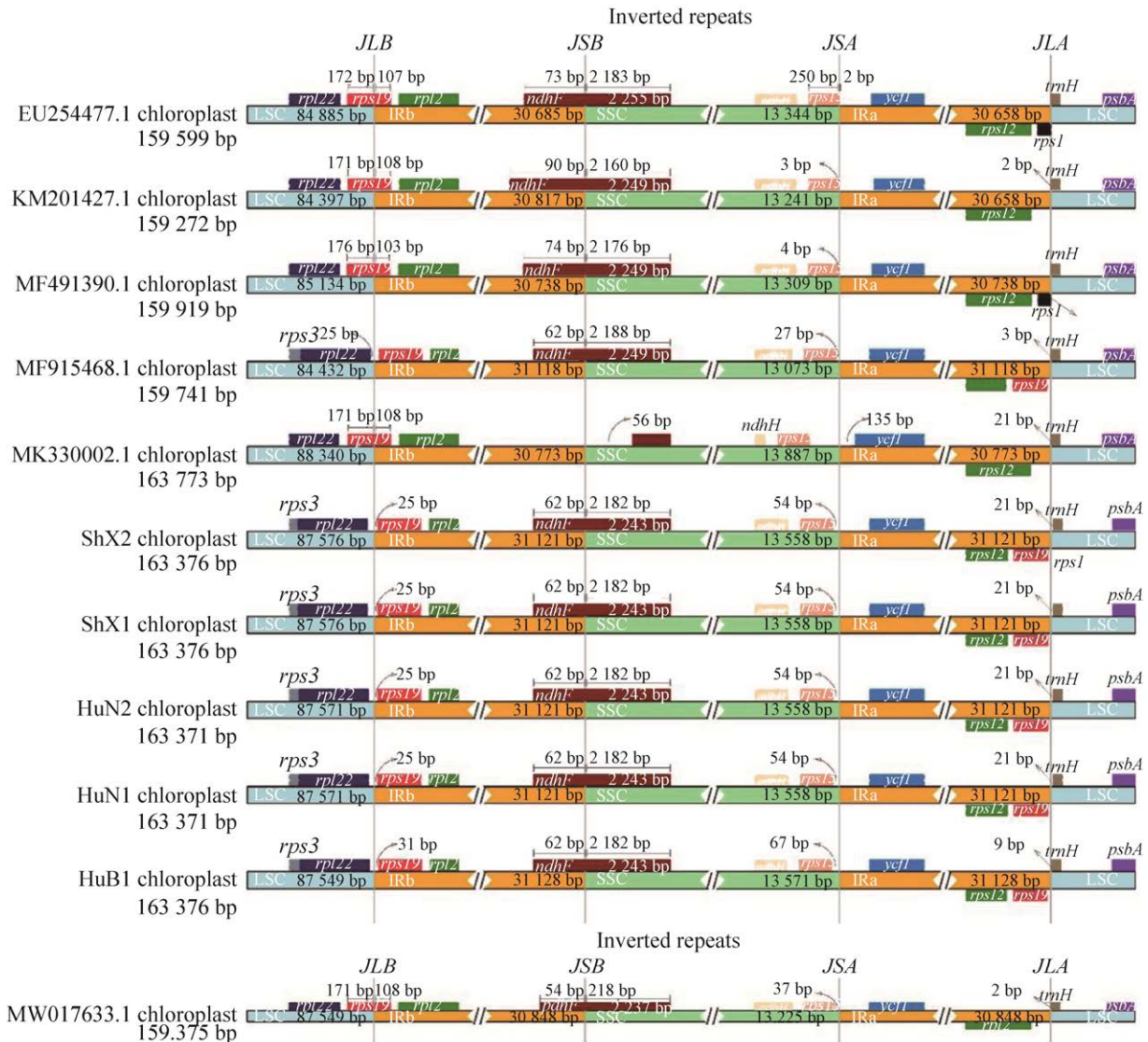


图 3 叶绿体基因组边界分析

Figure 3 Chloroplast genome boundaries.

2.3 叶绿体基因组序列变异分析

本试验利用在线工具 mVISTA 对 5 个虎杖品种和荞麦、金荞麦、苦荞麦、小野荞麦、金线草的叶绿体基因组序列比较结果进行可视化分析 (图 4), 以何首乌的叶绿体基因组序列为对照。分析结果发现, 各材料叶绿体基因组排列顺序基本一致, 具有较高保守性, 基因编码区基本一致, 而非编码区可明显看出参差不齐, 存在较明显的差异, 变异程度较高。在叶绿体的四分体结构中 LSC 区长度最大, 同时变异程度也最高, SSC 区长度最小, 但变异程度不低。而 IRa 区长度比 SSC 区长, 但变异程度最低。

5 个虎杖品种中 *trnk-UUU*、*rpoC1*、*petD*、

rpl16、*ndhA* 和 *rpl12* 基因的编码区发生变异, 对应可视化峰图出现差异。虎杖湖北 1 号品种相比于其余 4 个虎杖品种在编码区基因 *trnk-UUU* 3 847–3 852 bp 间发生突变, 插入了 ATAGTT 6 个碱基; 编码区基因 *rpoC1* 在 23 945–23 952 bp 间发生突变, 缺失了 ATTTATT 7 个碱基; 编码区基因 *petD* 在 80 914–80 918 bp 间发生突变, AAAAC 碱基突变为 CAAAA 碱基; 编码区基因 *rpl16* 在 87 572–87 600 bp 间发生突变, AATTA 碱基突变为 CATT 碱基; 编码区基因 *rpl12* 在序列尾部多出 AATTTG 6 个碱基。虎杖湖南 1 号和 2 号品种相比于其余 3 个虎杖品种, 编码区基因 *ndhA* 在 129 268 bp 发生突变,

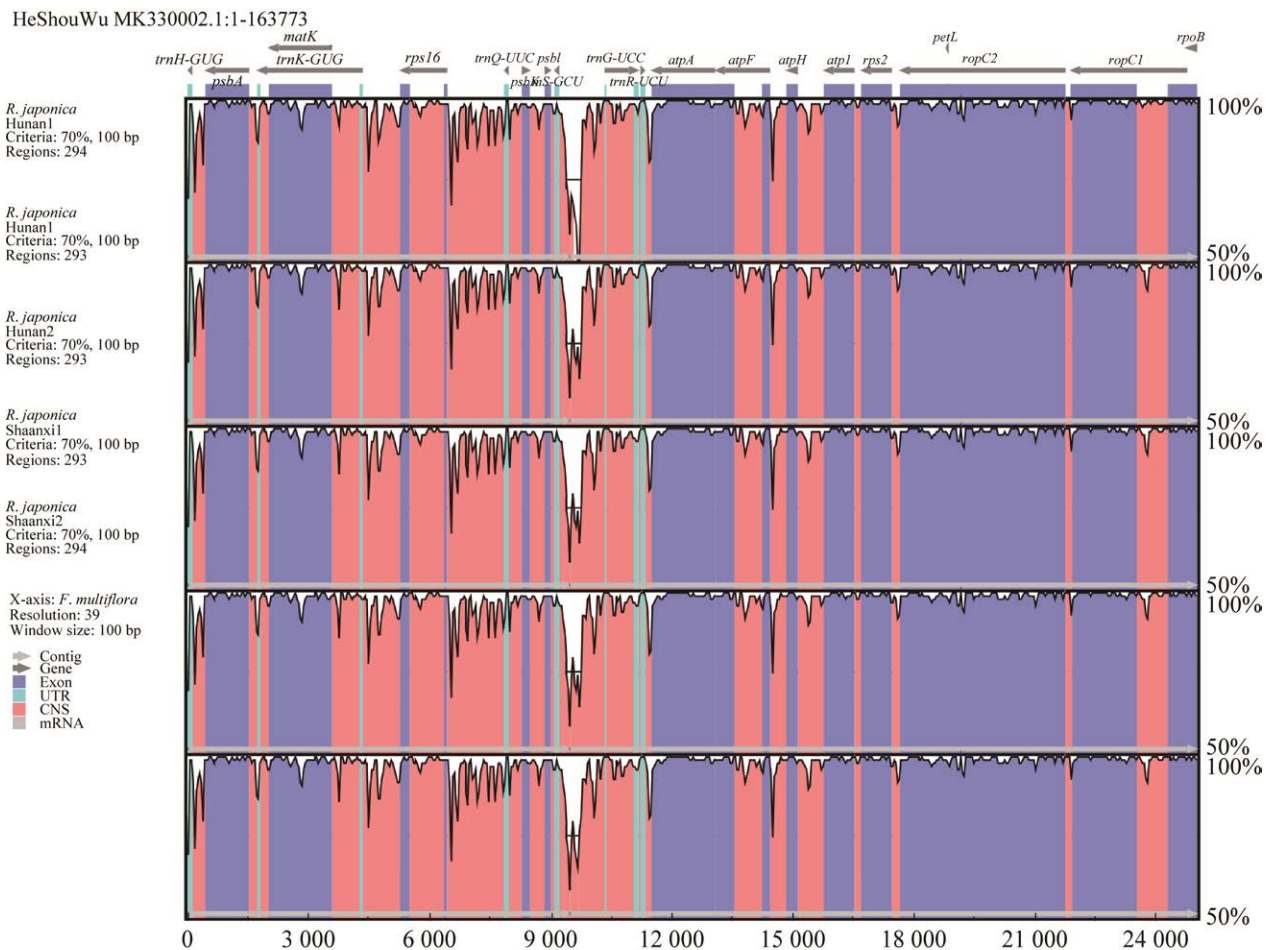


图 4 叶绿体基因组序列变异分析

Figure 4 Variations of the chloroplast genome sequence.

T 碱基突变为 A 碱基。5 个虎杖品种中, 还有非编码区变异和单碱基变异, 不影响基因的结构和功能。与虎杖相比, 其余 6 种材料的叶绿体基因组基因的编码区出现较多的差异。*trnK-UUU*、*rps16*、*rpoC2*、*rpl22*、*rps3*、*ycf2*、*ycf1*、*ndhF*、*ndhD* 和 *ycf1* 等基因的编码区差异明显, 可视化峰图差异较大。*trnK-UUU-rps16*、*rps16-trnQ-UUG*、*trnS-GCU-trnG-UCC*、*trnC-GCA-petN*、*petN-psbM*、*psaA-pafI*、*rps4-ndhJ*、*ndhK-atpB*、*accD-psaI*、*petA-psbL*、*rps7-ycf1*、*ndhF-rpl32*、*ycf1-rps7* 和 *ndhB-ycf2* 等基因间隔区的非编码区峰图参差不齐, 差异较大 (见图 4 和《生物工程学报》网络版附图 2)。

2.4 系统发育分析

以褐色沙拐枣 (*Calligonum colubrinum*,

GenBank 登录号为 MN202602.1) 为外类群, 用蓼族 11 种植物的叶绿体全基因组序列构建系统发育树 (图 5)。结果显示, 11 种植物大致分为两大分支, 一大分枝包括荞麦属的荞麦、金荞麦、苦荞麦、小野荞麦, 以及金线草属的金线草, 它们遗传距离较近; 另一大分支包括虎杖属的 5 个不同品种, 以及何首乌属的何首乌。由图 5B 分析, 5 个品种的虎杖中, 湖南 1 和湖南 2 以及陕西 1 和陕西 2 遗传距离差异为 0, 说明其根源基因相同。陕西 1 号和 2 号以及湖南 1 号和 2 号 4 个虎杖品种遗传距离 (0.000 12) 相较于湖北 1 号 (0.000 18) 小, 说明湖北 1 号与陕西 1 号和 2 号以及湖南 1 号和 2 号的亲缘关系相对较远。

蓼族内部系统发育分析显示, 虎杖是蓼族较为原始类群, 遗传距离为 0.0 294。何首乌 (遗

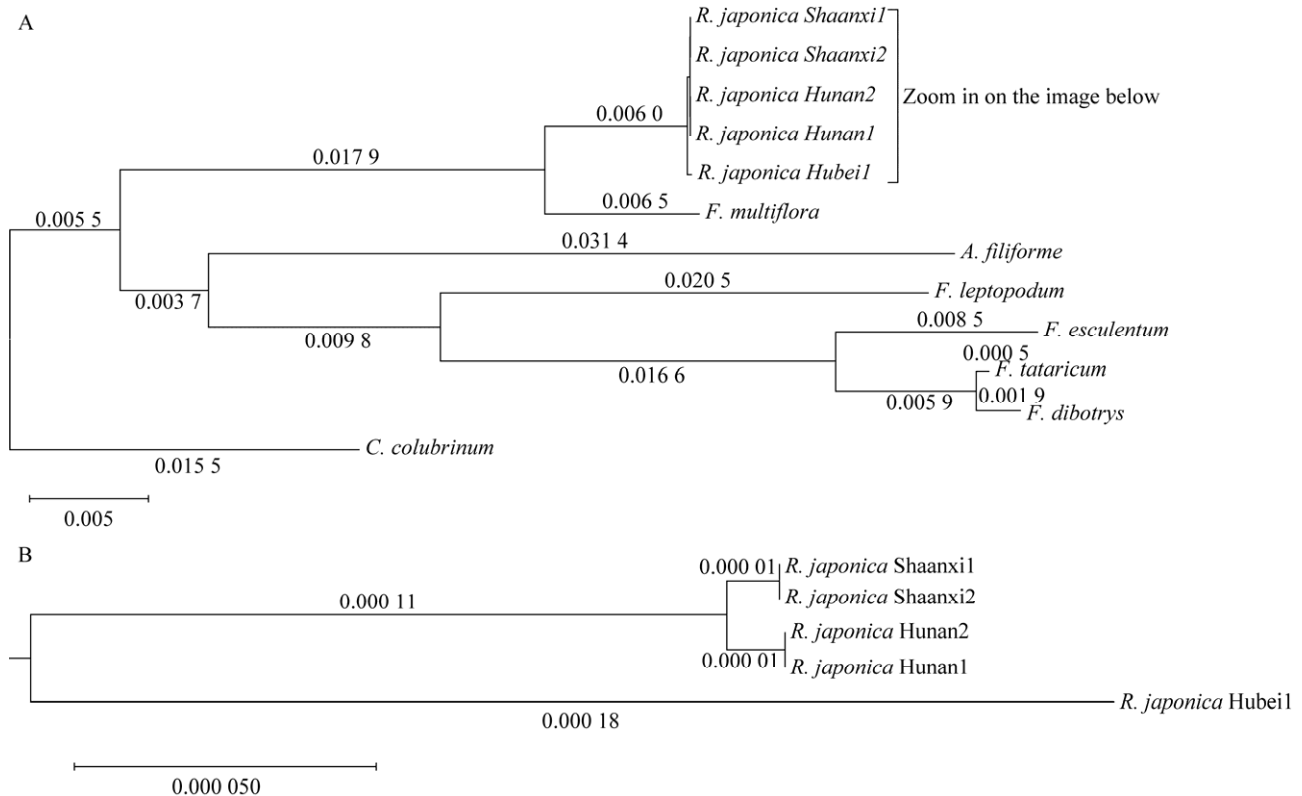


图 5 11 种植物叶绿体基因组序列系统发育分析

Figure 5 Phylogenetic tree. (A) Phylogenetic tree of the chloroplast genome sequences of 11 samples. (B) A enlarged part for *R. japonica* in the tree.

传距离 0.029 9, 下同)、小野荞麦 (0.039 5)、金线草 (0.041 6)、苦荞麦 (0.042)、金荞麦 (0.043 4)、荞麦 (0.044 1) 进化程度依次增加。

3 讨论

虎杖 5 个品种的植株以完全相同培养条件栽培于北京农学院东大地中, 芽叶长势良好, 并没有因品种不同而长势出现差别。从表型来看, 5 个品种的虎杖没有明显差异。

蓼族植物属于被子植物门。被子植物的叶绿体基因组大小在 120–160 kb 上下, 且叶绿体基因组的序列和结构具有高度的保守性^[20-21]。蓼族 11 种植物的叶绿体基因组都符合常见的双链环状结构, 由一个 SSC 区、一个 LSC 区和两个 IR 区构成的典型四分体结构, 叶绿体基因组长度均在 160 kb 左右。其中虎杖叶绿体基因组最长, 为 163 376 bp, 苦荞麦的叶绿体基因组最短, 为 159 272 bp, 二者相差为 4 104 bp。

一般而言, 植物叶绿体基因组的 LSC 区长度在 80–90 kb, SSC 区长度波动幅度最大, 为 16–27 kb, IR 区长度波动幅度最小, 为 20–28 kb^[21]。虎杖叶绿体基因组长度约为 163 kb, 大于 120–160 kb 的区间。其中 LSC 区长度为 87.5 kb, 在 80–90 kb 之间。SSC 区域长度为 13 kb, 小于 16–27 kb 的范围。IR 区长度为 31 kb, 超出 20–28 kb 的范围。可见, 虎杖叶绿体基因组较大主要是由具有较大的 IRa 和 IRb 区域导致的。

研究表明, 较长的 IR 区在维持叶绿体基因组稳定中起到重要作用^[22-23], 较长的 IR 区, 可使叶绿体基因组减少受到结构重排的影响^[24]。叶绿体基因组序列变异分析显示, IR 区的基因突变率较低, 相对应叶绿体基因组的保守性就较高。SSC 区为基因变异较多的编码区, 但是虎杖中较短的 SSC 区也减少了其叶绿体基因组基因变异的几率, 所以虎杖叶绿体基因组的保

守性较高。

研究中 11 种蓼族植物均只有 LSC 区域长度符合 80–90 kb 的范围, 长度为 84–89 kb。而 SSC 区域长度均在 13–14 kb 之间, 远短于 16–27 kb 的范围。IR 区长度在 30–32 kb 之间, 长于 20–28 kb 的范围。可见, 11 种蓼族植物均具有较大的 IR 区域。虽然 IR 区的基因突变率较低, 但是绝大部分的被子植物的 IR 区收缩和扩张与基因和内含子的缺失一样具有高突变性^[25]。这可能是导致蓼族植物 IR 区和 LSC 区和 SSC 区边界位置多样性的原因。

通过对 5 个虎杖品种以及蓼族其他 6 种植物叶绿体全基因组序列的对比分析, 发现 5 个虎杖品种叶绿体基因组具有较高的保守性, 虎杖湖北 1 号编码区变异相对于其余 4 个虎杖品种较多, *trnk-UUU*、*rpoCl*、*petD*、*rpl16* 和 *rpl12* 基因发生变异, 可为开发新的 DNA 条形码提供参考。虎杖湖南 1 号和 2 号品种的编码区基因 *ndhA* 发生单碱基突变, 并不影响基因的表达和蛋白质的合成。其余 6 种蓼族植物叶绿体基因组同样具有较高的保守性, 基因排列顺序基本一致, 没有大规模的基因缺失、突变、倒位和易位的现象。但在 IR 区与单拷贝区的边界上存在较明显差异, 基因编码区和非编码区的基因均存在一定的变异, 但非编码区基因变异较大。基于叶绿体全基因组序列的系统发育分析表明, 虎杖是蓼族植物中相对原始类群, 并且与何首乌属的何首乌亲缘关系较近, 而与何首乌属的小野荞麦、金荞麦、苦荞麦以及荞麦等亲缘关系较远。因此, 虎杖编码区基因变异对其生理功能的影响以及何首乌属内系统发育关系还有待进一步研究。

REFERENCES

- [1] 李安仁. 中国植物志(第 25 卷第一分册). 北京: 科学出版社, 1998: 1-118.

- Li AR. Chinese plant ambition (Volume 25, Volume I). Beijing: Science Press, 1998: 1-118 (in Chinese).
- [2] Peng W, Qin RX, Li XL, et al. Botany, phytochemistry, pharmacology, and potential application of *Polygonum cuspidatum* Sieb. et Zucc.: a review. *J Ethnopharmacol*, 2013, 148(3): 729-745.
- [3] Janko K, Kotlík P, Ráb P. Evolutionary history of asexual hybrid loaches (*Cobitis*: Teleostei) inferred from phylogenetic analysis of mitochondrial DNA variation. *J Evol Biol*, 2003, 16(6): 1280-1287.
- [4] Fry BG, Wüster W. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol Biol Evol*, 2004, 21(5): 870-883.
- [5] Pramual P, Nanork P. Phylogenetic analysis based on multiple gene sequences revealing cryptic biodiversity in *Simulium multistriatum* group (Diptera: Simuliidae) in Thailand. *Entomol Sci*, 2012, 15(2): 202-213.
- [6] Wen DQ, Yu Y, Hahn MW, et al. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Mol Ecol*, 2016, 25(11): 2361-2372.
- [7] Fitch WM, Margoliash E. Construction of phylogenetic trees. *Science*, 1967, 155(3760): 279-284.
- [8] Hillis DM, Huelsenbeck JP, Cunningham CW. Application and accuracy of molecular phylogenies. *Science*, 1994, 264(5159): 671-677.
- [9] Ris H, Plaut W. Ultrastructure of DNA-containing areas in the chloroplast of *Chlamydomonas*. *J Cell Biol*, 1962, 13(3): 383-391.
- [10] Chun EHL, Vaughan MH Jr, Rich A. The isolation and characterization of DNA associated with chloroplast preparations. *J Mol Biol*, 1963, 7(2): 130-141.
- [11] Shinozaki K, Ohme M, Tanaka M, et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J*, 1986, 5(9): 2043-2049.
- [12] 杨嘉鹏, 朱紫乐, 范雅娟, 等. 三种石豆兰属药用植物的叶绿体基因组比较分析及其在物种鉴定中的意义. *药学学报*, 2020, 55(11): 2736-2745.
Yang JP, Zhu ZL, Fan YJ, et al. Comparative plastomic analysis of three *Bulbophyllum* medicinal plants and its significance in species identification. *Acta Pharm Sin*, 2020, 55(11): 2736-2745 (in Chinese).
- [13] Raubeson LA, Jansen RK. Chloroplast genomes of plants. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. Wallingford: CABI, 2005: 45-68.
- [14] Jansen RK, Ruhlman TA. *Plastid Genomes of Seed Plants. Genomics of Chloroplasts and Mitochondria*[M]. Dordrecht: Springer, 2012: 103.
- [15] 付涛, 王志龙, 钱萍仙, 等. 高等植物 DNA 条形码最新研究进展及其应用. *核农学报*, 2016, 30(5): 887-896.
Fu T, Wang ZL, Qian PX, et al. The latest research progress and application of the DNA barcode in higher plants. *J Nucl Agric Sci*, 2016, 30(5): 887-896 (in Chinese).
- [16] 于惠敏. 植物的叶绿体基因组. *植物生理学通讯*, 2001, 37(5): 483-488.
Yu HM. Chloroplast genome of plant. *Plant Physiol J*, 2001, 37(5): 483-488 (in Chinese).
- [17] Zhang YH, Zheng LL, Zheng Y, et al. Assembly and annotation of a draft genome of the medicinal plant *Polygonum cuspidatum*. *Front Plant Sci*, 2019, 10: 1274.
- [18] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 1987, 4(4): 406-425.
- [19] Studier JA, Keppler KJ. A note on the neighbor-joining algorithm of Saitou and Nei. *Mol Biol Evol*, 1988, 5(6): 729-731.
- [20] Chumley TW, Palmer JD, Mower JP, et al. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol*, 2006, 23(11): 2175-2190.
- [21] Serrano M, Wang BJ, Aryal B, et al. Export of salicylic acid from the chloroplast requires the multidrug and toxin extrusion-like transporter EDS5. *Plant Physiol*, 2013, 162(4): 1815-1821.
- [22] Guisinger MM, Kuehl JV, Boore JL, et al. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol*, 2011, 28(1): 583-600.
- [23] Hirao T, Watanabe A, Kurita M, et al. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biol*, 2008, 8: 70.
- [24] Palmer JD, Thompson WF. Rearrangements in the chloroplast genomes of mung bean and pea. *PNAS*, 1981, 78(9): 5533-5537.
- [25] Dempewolf H, Kane NC, Ostevik KL, et al. Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass.—the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. *Mol Ecol Resour*, 2010, 10(6): 1048-1058.

(本文责编 陈宏宇)