

• 生物技术与方法 •

基于全基因组从头测序技术盐肤木叶绿体基因组的测序分析

左瑞华^{1,2}, 蒋平^{1,2}, 孙传伯^{1,2}, 陈存武^{1,2}, 娄新建¹

1 皖西学院 生物与制药工程学院, 安徽 六安 237012

2 安徽省中药资源保护与持续利用工程实验室, 安徽 六安 237012

左瑞华, 蒋平, 孙传伯, 等. 基于全基因组从头测序技术盐肤木叶绿体基因组的测序分析. 生物工程学报, 2020, 36(4): 772-781.

Zuo RH, Jiang P, Sun CB, et al. Analysis of the chloroplast genome characteristics of *Rhus chinensis* by *de novo* sequencing. Chin J Biotech, 2020, 36(4): 772-781.

摘要: 盐肤木是一种重要的经济树种, 可为医药和工业染料提供原料。盐肤木具有较强的抗旱、耐寒、耐盐, 可在温带、暖温带和亚热带地区生长。本研究首次对盐肤木叶绿体基因组进行从头测序 (*de novo sequencing*) 组装研究。结果表明, 盐肤木叶绿体基因组长度为 159 082 bp, 具有典型的四部分结构, 两个单拷贝区被一对反向重复区分隔。LSC 和 SSC 的长度分别为 85 394 bp 和 18 663 bp。叶绿体基因组总共编码 126 个基因, 其中包括 88 个蛋白编码基因, 8 个 rRNA 基因, 30 个 tRNA 基因。在叶绿体基因组中, 61.97% 的序列为基因编码区。在盐肤木叶绿体基因组中, 只有 8 个基因含有内含子, 除 *ycf3* 基因 (2 个内含子) 外, 其余均含有 1 个内含子。盐肤木叶绿体基因组总共存在 755 个 SSR 位点。SSR 主要由二核苷酸和单核苷酸组成, 分别占 60% (453) 和 28.74% (217)。聚类分析结果表明, 漆树科与盐肤木最为接近, 其次为槭树科和无患子科。本研究为盐肤木的分类提供了分子基础。本研究是关于盐肤木叶绿体基因组的首次报道, 对了解其光合作用、进化和叶绿体转基因工程具有重要意义。

关键词: 盐肤木, 叶绿体, 基因进化

Received: August 8, 2019; **Accepted:** February 2, 2020

Supported by: National Technical system of Chinese Medicinal Materials Industry (No. CARS-21), Anhui Provincial Quality of Undergraduate project (Nos. 2016sjjd079, 2018zygc075, 2018jyxm1153, 2018jyxm1155), Key Natural Science Projects of West Anhui University (No. WXZR201932), Teaching and Research Projects of West Anhui University (Nos. wxy2018026, wxy2017078), Pilot Project on the Reform of the Training Mode of Innovative Entrepreneurship in West Anhui University (No. wxy2017116).

Corresponding author: Ping Jiang. Tel: +86-564-3305043; E-mail: 837264050@qq.com

国家中药材产业技术体系 (No. CARS-21), 安徽省本科质量工程项目 (Nos. 2016sjjd079, 2018zygc075, 2018jyxm1153, 2018jyxm1155), 皖西学院自然科学重点项目 (No. WXZR201932), 皖西学院教学研究项目 (Nos. wxy2018026, wxy2017078), 皖西学院创新创业人才培养模式改革试验区项目 (No. wxy2017116) 资助。

网络出版时间: 2020-01-30

网络出版地址: <http://kns.cnki.net/kcms/detail/11.1998.Q.20200309.1046.002.html>

Analysis of the chloroplast genome characteristics of *Rhus chinensis* by *de novo* sequencing

Ruihua Zuo^{1,2}, Ping Jiang^{1,2}, Chuanbo Sun^{1,2}, Cunwu Chen^{1,2}, and Xinjian Lou¹

¹ College of Biological and Pharmaceutical Engineering, West Anhui University, Lu'an 237012, Anhui China

² Anhui Engineering Laboratory for Conservation and Sustainable Utilization of Traditional Chinese Medicine Resources, Lu'an 237012, Anhui China

Abstract: *Rhus chinensis* is an important economic species, which could provide raw materials for pharmaceutical and industrial dyes. *Rhus chinensis* is famous for its resistance to drought, cold, and salt. It grows in temperate, warm temperate, and subtropical regions. We report here *Rhus chinensis* chloroplast genomes by *de novo* sequencing. The results show that the length of *Rhus chinensis* was 159 082 bp, exhibiting a typical four-part structure with two single-copy regions (long single copy [LSC] and short single copy [SSC] sections) separated by a pair of inverted repeats (IRs). The length of LSC and SSC was 85 394 bp and 18 663 bp, respectively. The genomes contained 126 genes, including 88 protein encoding genes, 8 rRNA and 30 tRNA genes. In the chloroplast genome, 61.97% of the sequence were gene coding region. In the sequence of gene encoding region, the vast majority of sequences were protein encoding region, accounting for 86.65%, followed by rRNA (10 620 bp, 10.77%) and tRNA (2 540 bp, 2.58%). In *Rhus chinensis* chloroplast genome, only 8 genes contain introns, all containing 1 intron except *ycf3* gene (2 introns). The *Rhus chinensis* chloroplast genome contains 755 SSR locies. SSR mainly consists of dinucleotide and mononucleotide, accounting for 60% (453) and 28.74% (217) respectively. The clustering results show that *Anacardiaceae* were closest to *Rhus chinensis*, followed by *Aceraceae* and *Sapindaceae*. This study provides a molecular basis for the classification of *Rhus chinensis*.

Keywords: *Rhus chinensis*, chloroplast, genetic evolution

盐肤木 (*Rhus chinensis*) 隶属于漆树科盐肤木属, 是中国分布最为广泛和重要的经济树种之一, 是重要的制药、榨油和工业染料的原料^[1]。同时盐肤木的根、叶、花及果实均可作为中药材, 具有活血化瘀、舒经活络、涩肠止泻及止血的功效^[1], 此外, 寄生在盐肤木上的五倍子蚜虫便是我国传统中药五倍子^[2-3]。由于盐肤木具有抗旱、耐寒和耐盐碱等特性^[4-5], 广泛分布于东亚及东南亚各国, 是一种优良的经济树种^[6-7]。

叶绿体是植物细胞质中重要的细胞器, 可发生光合作用, 为植物的生长发育提供能量, 是植物生长发育的基础及生产力构成的主要因素^[8-9]。在植物中, 叶绿体基因组包含大量的遗传信息, 且具有高度保守的特性; 其基因组的自我复制和进化相对于物种而言, 保持着相对独立。因而, 叶绿体基因组常被用于探索植物基因组学和生物信息学的发生、发展和进化^[10-11]。到目前为止, 已经有超过 1 000 多种植物完成了叶绿体基因组

测序, 极大地推进了叶绿体基因组学的发展。然而到目前为止, 盐肤木叶绿体基因组的相关研究未见报道。因此, 本研究以盐肤木叶绿体为研究对象, 对其进行基因组学研究, 确定盐肤木的叶绿体基因组的特征及其在盐肤木系统进化地位, 并为盐肤木的系统进化和遗传多样性分析等提供参考。

1 材料与方法

1.1 试验材料

新鲜盐肤木叶由皖西学院生物制药工程学院中药种植基地提供, 将新鲜叶片用无菌水清洗干净后, 立即进行干冰冻存, -80℃冰箱保存, 用于后续分析。

1.2 试验方法

1.2.1 叶绿体基因组测序

采用高盐低 pH 法提取叶绿体基因组, 即将获得的新鲜叶片经匀浆过滤和纯化处理以获得盐肤木叶绿体, 同时使用显微镜对盐肤木叶绿体的完整性进行检测。为了消除细胞核及线粒体等其

他 DNA 的干扰, 使用 DNase 消化。之后使用裂解液充分裂解盐肤木叶绿体, 经离心、抽提、纯化后获得 cpDNA。cpDNA 检测合格后, 由北京源宜基因科技有限公司负责对其进行高通量测序。

1.2.2 原始测序数据质控统计

为了提高原始测序数据(Raw Data)的质量,

避免 Raw Data 的一部分 Reads(即带有测序引物、接头等人工序列)对实验分析产生干扰, 需要先对 Raw Data 进行过滤, 去除影响数据质量和低质量序列, 即去除接头、含有模糊碱基 N 及长度小于 60 bp 的序列, 并使用 BWA 算法去除序列两端的低质量区域, 质量阈值为 30。

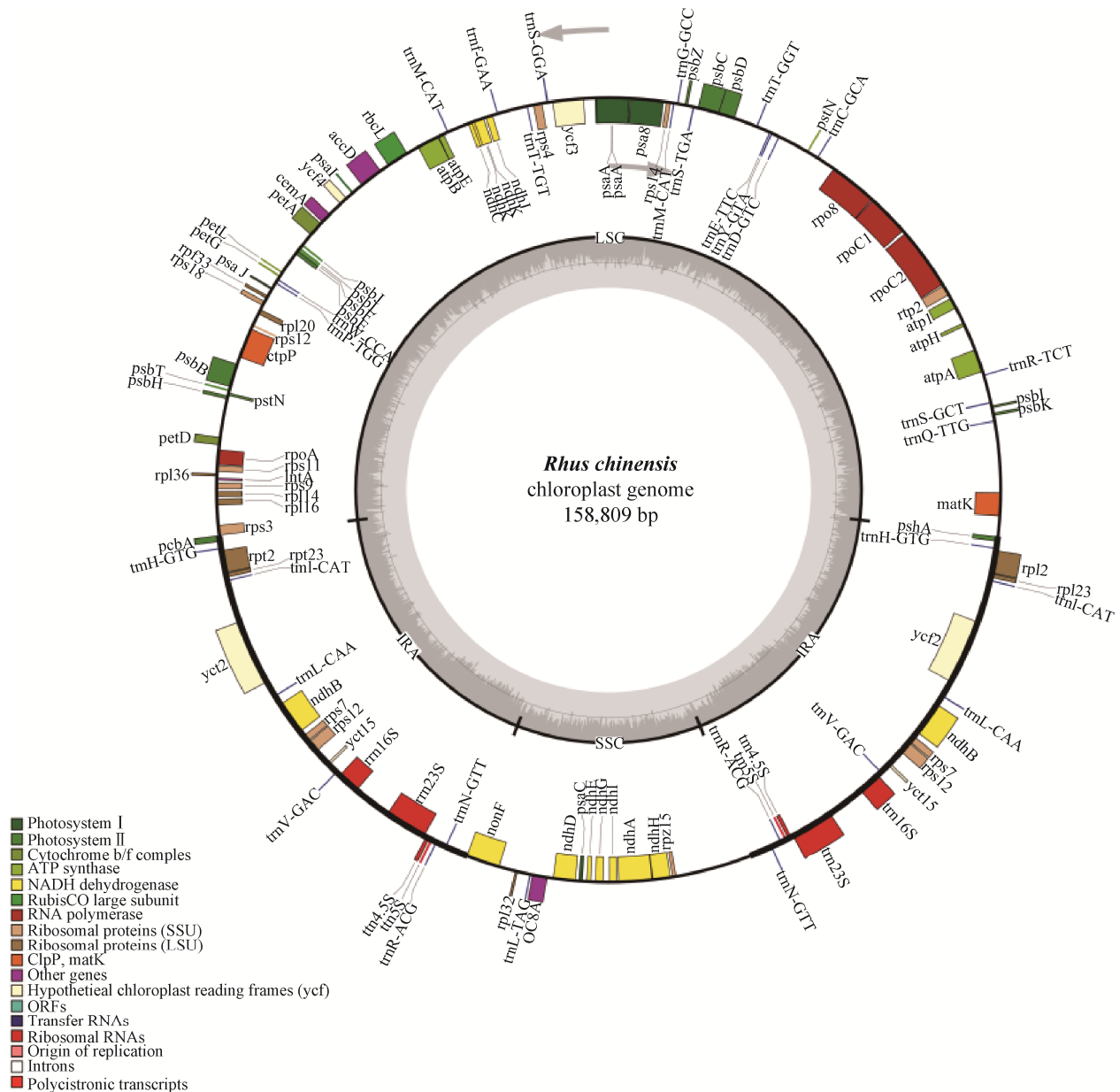


图 1 盐肤木叶绿体基因组环状基因图 (内侧基因顺时针转录, 外侧基因逆时针转录。不同功能的基因以不同颜色)

Fig. 1 Gene map of the *Rhus chinensis* chloroplast genome. Genes drawn inside the circle are transcribed clockwise, and those outside are counterclockwise. Genes belonging to different functional groups are color-coded.

1.2.3 叶绿体基因组组装及注释

对原始数据 Raw Data 进行过滤处理后得到高质量数据 Clean Data。首先,为了得到 contig 序列,使用 SOAP *de novo* (<http://soap.genomics.org.cn/soapdenovo.html>) 对 Clean Data 进行初步组装。其次,使用 BLAT 软件将其与近缘物种的叶绿体参考基因组进行定位比对,获得 contig 序列之间的相对位置进行拼接并校正组装错误,以获得 cpDNA 全长框架图。之后,利用 GapCloser 软件将高质量的短序列填补到框架图序列上的 gap,并利用一代测序对剩余的 gap 及可疑区域进行补充并确认,之后验证 LSC、SSC、IR 区域连接处,最终得到一条环形的 cpDNA 完成图序列,并使用 CpGAVAS 软件对 cpDNA 序列进行注释。同时采用 DNAMAN 和 DNASTAR 软件对 cpDNA 序列进行 GC 和 AT 含量统计,采用 MEGA 软件对密码子使用情况进行统计分析。

1.2.4 叶绿体基因组组装及注释

使用 MISA 在线工具 (<http://pgrc.ikp-gatersleben.de/misa>) 对盐肤木 cpDNA 进行 SSR 位点搜索并设置单核苷酸 (8 次)、二核苷酸 (3 次)、三核苷酸 (3 次)、四核苷酸 (3 次)、五核苷酸 (3 次)、SSR 侧翼序列长度 (≥ 50 bp) 和不同位点间最小间距 (100 bp) 等搜索参数。根据搜索结果统计分析盐肤木叶绿体 SSR 位点的数量和分布规律。

1.2.5 系统发育分析

根据盐肤木叶绿体基因组注释信息,以及从 NCBI 下载的其他 10 种无患子目植物叶绿体基因组,从基因组 gb 文件中分别提取其注释的所有基因。通过比对,选择 66 个共有蛋白编码基因用于系统发育分析,首先将所有基因按照顺序进行拼接,再用 ClustalW 软件进行多序列全局比对,利用 MEGA 5.0 软件构建系统发育树,使用最大似然法 (Maximum likelihood, ML) 进行系统进化树构建,进化树用 1 000 次重复 bootstrap 检验,构建系统发育树,分析盐肤木的系统进化关系。

2 结果与分析

2.1 盐肤木叶绿体的基本特征

对盐肤木叶绿体基因组进行分析可知 (图 1),盐肤木和其他植物叶绿体基因组类似,其叶绿体基因组为典型的四段式结构,即包括一个大单拷贝区(LSC)、一个小单拷贝区(SSC)和两个反向重复区(IR)。如表 1 所示,盐肤木叶绿体基因组的全长为 158 809 bp,其中 LSC 区长度为 85 394 bp,SSC 长度 27 376 bp,两个 IR 区长度为 27 376 bp。基因组的 GC 含量为 37.86%,与其他植物叶绿体基因组 GC 含量接近。盐肤木 cpDNA 共编码 126 个基因,其中蛋白编码基因 88 个,rRNA 编码基因 8 个和 tRNA 编码基因 30 个。在叶绿体基因组中,有 48.80%的序列为基因编码区。在基因编码区序列中,绝大多数序列为蛋白质编码区,占 84.90% (65 799 bp),其次是 rRNA (12.13%, 9 400 bp) 和 tRNA (2.97%, 2 302 bp)。非编码区占叶绿体总数的 51.20%,大部分为间质间隔序列 (IGS),其余为内含子(12 898 bp, 15.86%)。

由表 2 可知,在盐肤木 cpDNA 功能分类过程中发现,绝大多数基因都与盐肤木的光合作用

表 1 盐肤木叶绿体基因组基本特性

Table 1 Basic characteristics of *Rhus chinensis* chloroplasts

Chloroplasts feature	Numerical value
Length (bp)	158 809
GC content (%)	37.86
AT content (%)	62.14
LSC length (bp)	85 394
SSC length (bp)	18 663
IR length (bp)	27 376
Gene number	126
Gene number in IR regions	35
Protein-coding gene number	88
Protein-coding gene (%)	69.84
rRNA gene number	8
rRNA (%)	6.35
tRNA gene number	30
tRNA (%)	23.81

(Genes for photosynthesis) 和自我复制 (Self replication) 功能有关。光合作用相关的基因主要聚集在 ATP 合酶亚基 (Subunits of ATP synthase)、NADH-脱氢酶亚基 (Subunits of NADH-dehydrogenase)、细胞色素 b/f 复合物亚基 (Subunits of cytochrome b/f complex)、光系统 I 亚基 (Subunits of photosystem I)、光系统 II 亚基 (Subunits of photosystem II) 和二磷酸核酮糖羧合酶/羧化酶亚基 (Subunit of rubisco) 这些类别中, 其中有 15 个基因聚集在光系统 II 亚基中, 其相关基因的数量最多。而与盐肤木叶绿体自我复制相关的基因主要聚集在核糖体大亚基 (Large subunit of ribosome)、DNA 依赖性 RNA 聚合酶 (DNA dependent RNA polymerase)、核糖体小亚基 (Small subunit of ribosome)、rRNA 基因

(rRNA genes) 和 tRNA 基因 (tRNA genes) 中, 其中 tRNA 基因 (tRNA genes) 家族包含的基因最多, 有 30 个基因。

2.2 盐肤木叶绿体基因内含子分析

对盐肤木叶绿体的蛋白编码基因进行分析可知, 盐肤木叶绿体白编码基因绝大多数基因均不含有内含子, 这与其他植物类似。盐肤木叶绿体中总共有 8 个基因含有内含子, 除了 *ycf3* 基因含有 2 个内含子, 其余 7 个基因 (*rpoC1*、*psaA*、*rpl2*、*ndhA*、*rpl2*、*rps12+*、*rps12-*) 均含有 1 个内含子。不同基因的内含子长度差异较大, 其长度在 539–2 815 bp 范围内变化, *rpoC1* 基因内含子最长, *RPS12* 基因内含子最短 (图 2)。

表 2 盐肤木叶绿体基因组表达基因

Table 2 Genes in *Rhus chinensis* chloroplast genome

Functions	Family name	Code	List of genes
Genes for photosynthesis	Subunits of ATP synthase	<i>atp</i>	<i>atpA</i> , <i>atpH</i> , <i>atpI</i> , <i>atpE</i> , <i>atpB</i>
	Subunits of NADH-dehydrogenase	<i>ndh</i>	<i>ndhJ</i> , <i>ndhK</i> (pseudogene), <i>ndhK</i> , <i>ndhC</i> , <i>ndhB</i> (pseudogene), <i>ndhF</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhG</i> , <i>ndhI</i> , <i>ndhA</i> , <i>ndhH</i> , <i>ndhB</i> (pseudogene)
	Subunits of cytochrome b/f complex	<i>pet</i>	<i>petN</i> , <i>petA</i> , <i>petL</i> , <i>petG</i> , <i>petD</i>
	Subunits of photosystem I	<i>psa</i>	<i>psaB</i> , <i>psaA</i> , <i>psaA</i> , <i>psaI</i> , <i>psaJ</i> , <i>psaC</i>
	Subunits of photosystem II	<i>psb</i>	<i>psbK</i> , <i>psbI</i> , <i>psbD</i> , <i>psbC</i> , <i>psbZ</i> , <i>psbJ</i> , <i>psbL</i> , <i>psbF</i> , <i>psbE</i> , <i>psbB</i> , <i>psbT</i> , <i>psbN</i> , <i>psbH</i> , <i>psbA</i> , <i>psbA</i>
Self replication	Subunit of rubisco	<i>rbc</i>	<i>rbcL</i>
	Small subunit of ribosome	<i>rps</i>	<i>rps2</i> , <i>rps14</i> , <i>rps4</i> , <i>rps18</i> , <i>rps11</i> , <i>rps8</i> , <i>rps3</i> , <i>rps7</i> , <i>rps15</i> , <i>rps7</i> , <i>rps12</i> , <i>rps12</i>
	rRNA genes	<i>rrn</i>	<i>rrn16S</i> , <i>rrn23S</i> , <i>rrn4.5S</i> , <i>rrn5S</i> , <i>rrn5S</i> , <i>rrn4.5S</i> , <i>rrn23S</i> , <i>rrn16S</i>
	Large subunit of ribosome	<i>rpl</i>	<i>rpl33</i> , <i>rpl20</i> , <i>rpl36</i> , <i>rpl14</i> , <i>rpl16</i> , <i>rpl2</i> , <i>rpl23</i> , <i>rpl32</i> , <i>rpl23</i> , <i>rpl2</i>
	DNA dependent RNA polymerase	<i>rpo</i>	<i>rpoC2</i> , <i>rpoC1</i> , <i>rpoB</i> , <i>rpoA</i>
tRNA genes	<i>trn</i>	<i>trnQ</i> -TTG, <i>trnS</i> -GCT, <i>trnR</i> -TCT, <i>trnC</i> -GCA, <i>trnD</i> -GTC, <i>trnY</i> -GTA, <i>trnE</i> -TTC, <i>trnT</i> -GGT, <i>trnS</i> -TGA, <i>trnG</i> -GCC, <i>trnM</i> -CAT, <i>trnS</i> -GGA, <i>trnT</i> -TGT, <i>trnF</i> -GAA, <i>trnM</i> -CAT, <i>trnW</i> -CCA, <i>trnP</i> -TGG, <i>trnH</i> -GTG, <i>trnI</i> -CAT, <i>trnL</i> -CAA, <i>trnV</i> -GAC, <i>trnR</i> -ACG, <i>trnN</i> -GTT, <i>trnL</i> -TAG, <i>trnN</i> -GTT, <i>trnR</i> -ACG, <i>trnV</i> -GAC, <i>trnL</i> -CAA, <i>trnI</i> -CAT, <i>trnH</i> -GTG	
Other genes	Subunit of acetyl-CoA-carboxylase	<i>acc</i>	<i>accD</i>
	Envelop membrane protein	<i>cem</i>	<i>cemA</i>
	c-type cytochrom synthesis gene	<i>ccs</i>	<i>ccsA</i>
	Protease	<i>clp</i>	<i>clpP</i> (pseudogene)
	Translational initiation factor	<i>inf</i>	<i>infA</i>
	Maturase	<i>mat</i>	<i>matK</i>
Genes of unkown function	Conserved open reading frames	<i>ycf</i>	<i>ycf3</i> , <i>ycf4</i> , <i>ycf2</i> , <i>ycf15</i> , <i>ycf15</i> , <i>ycf2</i>

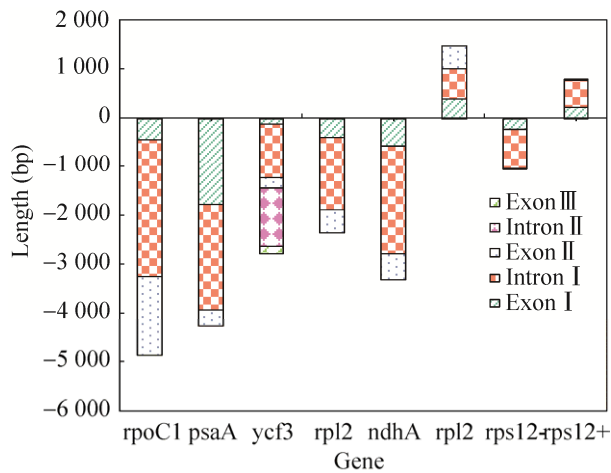


图2 盐肤木叶绿体基因组内含子基因长度

Fig. 2 Intron length in the chloroplast genomes of *Rhus chinensis*.

表3 盐肤木叶绿体基因密码子统计表

Table 3 Codon usage in the *Rhus chinensis* chloroplast genome

Amino acid	Codon	Number	Percent (%)	Amino acid	Codon	Number	Percent (%)
Ala (A)	GCA	345	1.58	Leu (L)	CTC	165	0.75
Ala (A)	GCC	208	0.95	Leu (L)	TTA	726	3.32
Ala (A)	GCG	156	0.71	Leu (L)	TTG	484	2.21
Ala (A)	GCT	596	2.73	Lys (K)	AAG	257	1.18
Arg (R)	AGG	124	0.57	Lys (K)	AAA	765	3.50
Arg (R)	AGA	383	1.75	Met (M)	ATG	493	2.25
Arg (R)	CGA	292	1.34	Phe (F)	TTT	784	3.59
Arg (R)	CGC	101	0.46	Phe (F)	TTC	423	1.93
Arg (R)	CGG	114	0.52	Pro (P)	CCT	337	1.54
Arg (R)	CGT	292	1.34	Pro (P)	CCG	135	0.62
Asn (N)	AAC	220	1.01	Pro (P)	CCA	266	1.22
Asn (N)	AAT	756	3.46	Pro (P)	CCC	193	0.88
Asp (D)	GAT	692	3.16	Ser (S)	AGC	97	0.44
Asp (D)	GAC	187	0.86	Ser (S)	AGT	356	1.63
Cys (C)	TGT	186	0.85	Ser (S)	TCG	148	0.68
Cys (C)	TGC	67	0.31	Ser (S)	TCA	321	1.47
Gln (Q)	CAA	570	2.61	Ser (S)	TCC	290	1.33
Gln (Q)	CAG	196	0.90	Ser (S)	TCT	448	2.05
Glu (E)	GAG	267	1.22	Thr (T)	ACT	435	1.99
Glu (E)	GAA	780	3.57	Thr (T)	ACA	342	1.56
Gly (G)	GGT	555	2.54	Thr (T)	ACG	125	0.57
Gly (G)	GGG	303	1.39	Thr (T)	ACC	215	0.98
Gly (G)	GGA	613	2.80	Trp (W)	TGG	393	1.80
Gly (G)	GGC	161	0.74	Tyr (Y)	TAT	642	2.94
His (H)	CAT	448	2.05	Tyr (Y)	TAC	146	0.67
His (H)	CAC	149	0.68	Val (V)	GTA	468	2.14
Ile (I)	ATC	368	1.68	Val (V)	GTC	156	0.71
Ile (I)	ATA	550	2.52	Val (V)	GTG	162	0.74
Ile (I)	ATT	916	4.19	Val (V)	GTT	451	2.06
Leu (L)	CTT	476	2.18	End	TAG	17	0.08
Leu (L)	CTG	158	0.72	End	TGA	20	0.09
Leu (L)	CTA	336	1.54	End	TAA	43	0.20

2.3 盐肤木叶绿体密码子分析

在盐肤木叶绿体基因组中, 61.97%的序列为基因编码区, 包括 88 个蛋白质编码基因(表 3)。对氨基酸组成进行分析可知, 亮氨酸(2345)编码率最高, 半胱氨酸(253)编码率最低。密码子的核苷酸具有明显的选择性。密码子第 1 位、第 2 位和第 3 位的 A/T 含量分别为 52.60%、60.54%和 69.43%。第 3 位密码子具有明显的偏好, 这在其他高等植物的叶绿体基因组中非常常见。这也可能是不同植物保存叶绿体基因组的主要原因。

生物体内的氨基酸一般有两个或两个以上的密码子, 这些密码子是同义密码子, 同义密码子编码相同的氨基酸, 这就是密码子的简并性。密

码子的简并性能有效地降低有害突变对植物的影响,具有重要的生物学意义。在没有选择性压力或突变偏好的情况下,同义密码子的概率是相同的,每个氨基酸位点的核苷酸突变是随机的。然而,在实践中,同义密码子的使用并不是随机的,但是同一物种的不同物种或不同基因在编码氨基酸时往往使用特定的同义密码子。这种现象被称为同义密码子使用偏好。密码子使用率(RSCU)的相对比较是指特定密码子的使用频率与预期频率的比值,这是衡量密码子偏好程度的有效方法。

本研究中盐肤木叶绿体编码基因共有 64 个密码子。除色氨酸和蛋氨酸外,所有氨基酸同义密码子都是预先确定的,共有 29 个密码子,涉及 18 个氨基酸和 1 个终止密码子(图 3)。高密码子偏好高等植物叶绿体基因中很常见,是叶绿体基因相对保守的主要原因。

2.4 盐肤木基因组 SSR 位点分析

对盐肤木叶绿体基因组的 SSR 位点进行分析可知,其叶绿体的 SSR 基因位点总共有 755 个,由 49 个重复单元组成。不同类型 SSR 位点的重

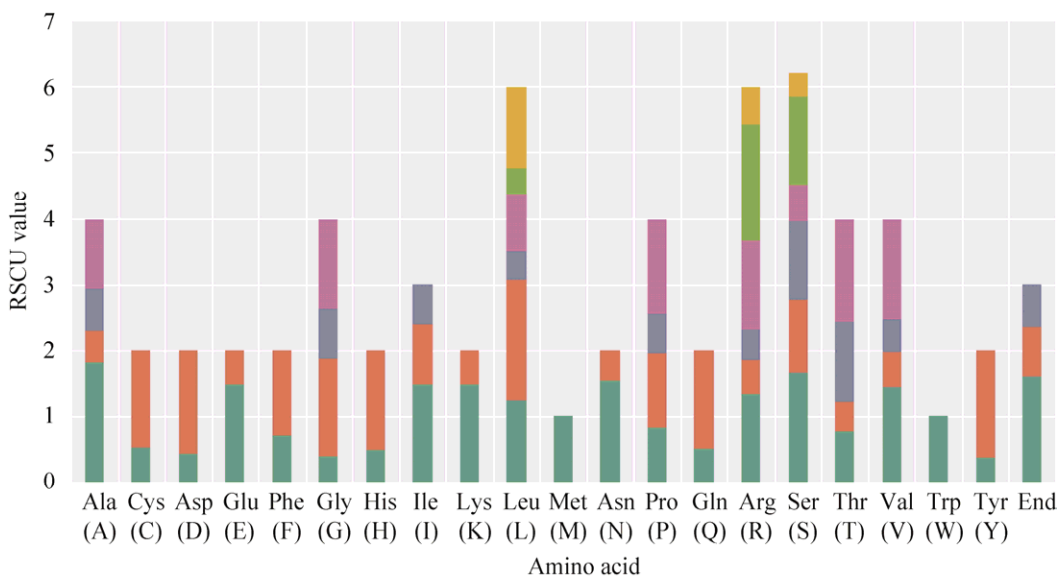


图 3 20 个氨基酸终止密码子含量

Fig. 3 Codon content of 20 amino acid and stop codons in all protein-coding genes.

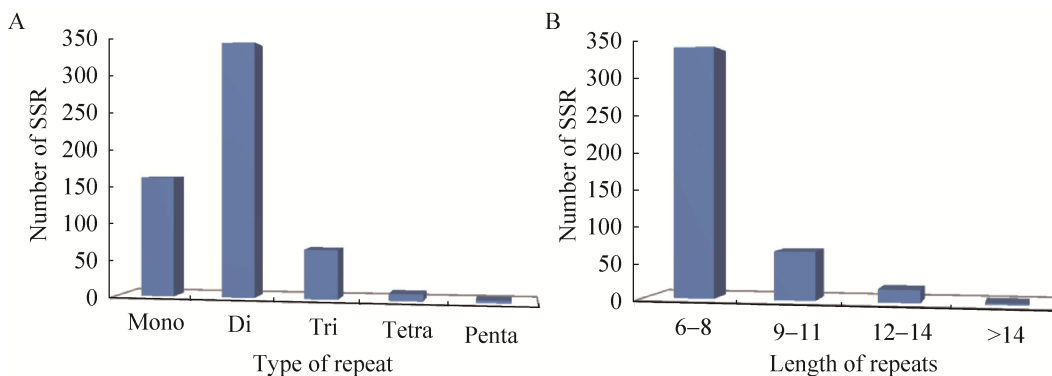


图 4 SSR 位点类型和长度

Fig. 4 Type and length of SSR repeats.

复单元数目不同。SSR 位点主要由二核苷酸和单核苷酸组成,分别占总数的 60% (453) 和 28.74% (217)。SSR 序列的长度主要由 6 bp 到 8 bp 的短序列组成,占 64.50% (487) (图 4)。

对所有 SSR 位点的分析表明,不同类型的 SSR 重复单元中的 SSR 重复单元数目有显著差异 (表 4)。单核苷酸重复单位主要由 A/T 组成,占

表 4 盐肤木叶绿体基因组 SSR 位点

Table 4 List of identified simple sequence repeats of the *Rhus chinensis* chloroplast genome

SSR repeats	Number of repeats									Total
	3	4	5	6	7	8	9	10	>10	
A/T	-	-	-	-	-	53	43	36	23	155
C/G	-	-	-	-	-	3	4	1		8
AC/GT	51	3								54
AG/CT	111	18								129
AT/AT	125	23	2	1	1					152
CG/CG	10	1								11
AAC/GTT	6									6
AAG/CTT	21	2								23
AAT/ATT	17	5								22
ACC/GGT	3									3
ACT/AGT	1									1
AGC/CTG	5									5
AGG/CCT	3									3
ATC/ATG	3									3
AAAG/CTTT	1									1
AAAT/ATTT	2									2
AACT/AGTT	1									1
AATC/ATTG	1									1
AATG/ATTC	2									2
ACAT/ATGT	1									1
AGAT/ATCT	1									1
AAAAT/ATTTT	1									1
AAATT/AATTT	1									1
AATAT/ATATT	1									1

95.10%。在二核苷酸重复序列中,AT/AT 和 AG/CT 最常见,分别占 43.91%和 37.28%。然而,三核苷酸、四核苷酸和五核苷酸重复序列的数量相对较少。

2.5 系统发育分析

为了确定盐肤木在无患子目植物中的系统发育位置,我们利用 NCBI 中发表的其他 10 种无患子目植物的 66 个常见叶绿体蛋白编码基因进行了系统发育分析。聚类的支持率较高,大多数分支节点的检验分值都达到了 100%,表明聚类结果的可靠性很高 (图 5)。通过分析,将所有 11 种植物分为 4 大类。第一大类由 5 个槭树科物种和 1 个无患科物种,表明槭科和无患子科的亲缘关系密切;第二大类由腰果和盐肤木组成,说明盐肤木和腰果的亲缘关系最近;第三大类同样由两个物种组成:小叶黄杨和云南黄柏;而大叶黄杨与其他植物相距甚远,单独分为一类。聚类分析结果表明,漆树科与盐肤木的亲缘关系最近,其次为槭树科和无患子科。本研究为盐肤木的分类提供了分子基础。

3 讨论与结论

叶绿体是植物细胞内半自主的细胞器,在植物细胞光合作用和为植物生长提供营养物质等方面发挥着重要作用^[12-13]。因此,通过基因组学等方法,研究盐肤木的叶绿体基因组的特征及其在盐肤木系统进化地位,对盐肤木的系统进化和遗传多样性分析等具有重要的理论意义。本研究对盐肤木叶绿体基因组进行测序,首次完成了高质量叶绿体基因组的组装和注释并获得了总长度为 159 082 bp 盐肤木叶绿体基因组,其具有典型被子植物叶绿体基因组环状四分体结构,即包含 4 个不同的区域:一对为 27 376 bp 的反向重复区 (IR), 85 394 bp 的大单拷贝区 (LSC) 和一个小单拷贝区 (SSC) 18 663 bp。具有典型的四部分结构,两个单拷贝区被一对反向重复 (IRs) 分隔。在盐

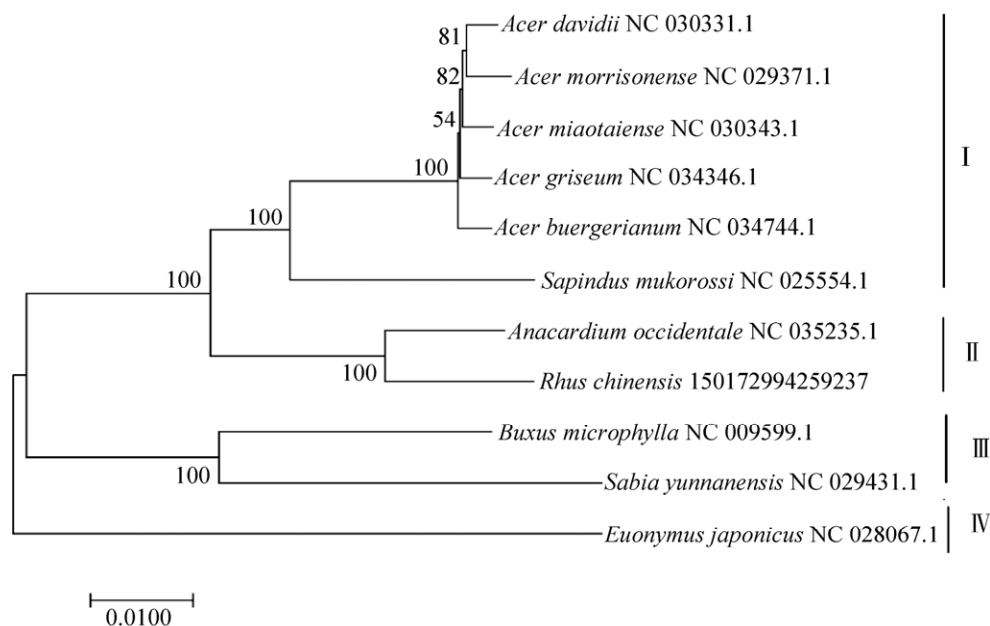


图 5 基于共有蛋白的盐肤木 ML 系统发育树

Fig. 5 The ML phylogenetic tree of the *Rhus chinensis* clade based on same protein-coding genes. Note: Numbers above or below the nodes are bootstrap support values.

肤木叶绿体基因组中, 只有 8 个基因含有内含子, 除 *ycf3* 基因含有 2 个内含子外, 其余均含有 1 个内含子。盐肤木叶绿体基因组包含 755 个 SSR 位点。聚类分析结果表明, 漆树科与盐肤木最为接近, 其次为槭树和漆树科。本研究为盐肤木的分类提供了分子基础, 对了解其光合作用、进化和叶绿体转基因工程具有重要意义。

REFERENCES

- [1] Wu M. Molecular cloning of a phenylalanine ammonia-lyase gene from *Rhus chinensis* and the study on its function and quality[D]. Taiyuan: Shanxi University, 2012 (in Chinese).
武敏. 盐肤木苯丙氨酸解氨酶基因的克隆及其功能性质研究[D]. 太原: 山西大学, 2012.
- [2] Chen XL. Study on the treatment of proteinuria with *Rhus chinensis*. China's Naturop, 2017, 25(4): 92–93 (in Chinese).
陈小蕾. 五倍子治疗蛋白尿的研究. 中国民间疗法, 2017, 25(4): 92–93.
- [3] Liu P, Ren WB, Yang Y, et al. Polymorphism analysis of cytochrome oxidase I (CO I) gene in horned gall aphid *Schlechtendalia chinensis*. Amino Acids Biot Resour, 2018, 40(1): 36–40 (in Chinese).
刘平, 任维宾, 杨瑛, 等. 角倍蚜细胞色素氧化酶 (CO I) 基因的多样性分析. 生物资源, 2018, 40(1): 36–40.
- [4] Zhao J, Cui CB, Cai B, et al. Research progress of *Rhus* plants in China. Pharm J Chin PLA, 2006, 22(1): 48–50 (in Chinese).
赵军, 崔承彬, 蔡兵, 等. 国产盐肤木属植物的研究进展. 解放军药学学报, 2006, 22(1): 48–50.
- [5] Liu CQ. Basic research and utilization of *Rhus chinensis* Mill. Forestry of China, 2008, (23): 53 (in Chinese).
刘春强. 盐肤木基础研究及利用. 中国林业, 2008, (23): 53.
- [6] Wang L, Wang N, Li T, et al. Sumac (*Rhus chinensis* Mill) biomass refinery engineering. Chin J Biotech, 2014, 30(5): 695–706 (in Chinese).
王岚, 王宁, 李坦, 等. 盐肤木生物质炼制工程.

- 生物工程学报, 2014, 30(5): 695–706.
- [7] Hu FM, Tan XF, Liu HM. Culture and Utilization of Chinese Non-Wood Product Forest Trees. Beijing: China Forestry Publishing House, 2006: 719–723 (in Chinese).
胡芳名, 谭晓凤, 刘惠民. 中国主要经济树种栽培与利用. 北京: 中国林业出版社, 2006: 719–723.
- [8] 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.
- [9] Zhang YX, Iaffaldano BJ, Zhuang XF, et al. Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. *BMC Plant Biol*, 2017, 17: 34.
- [10] Katayama H, Tachibana M, Iketani H, et al. Phylogenetic utility of structural alterations found in the chloroplast genome of pear: hypervariable regions in a highly conserved genome. *Tree Genet Genom*, 2012, 8(2): 313–326.
- [11] Haberle RC, Fourcade HM, Boore JL, et al. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol*, 2008, 66(4): 350–361.
- [12] Cui CJ. Studies on chloroplast genetic engineering and plant organelle genomes evolution[D]. Wuhan: Huazhong University of Science and Technology, 2010 (in Chinese).
崔翠菊. 叶绿体基因工程与植物细胞器基因组进化研究[D]. 武汉: 华中科技大学, 2010.
- [13] Gao XP. Function analysis of the *Arabidopsis* WXR1, WXR3 proteins during the starch metabolism[D]. Tai'an: Shandong Agricultural University, 2015 (in Chinese).
高雪萍. 拟南芥 WXR1 和 WXR3 蛋白参与淀粉代谢作用机理的研究[D]. 泰安: 山东农业大学, 2015.

(本文责编 郝丽芳)