

# 裂叶榆叶绿体基因组及 CP-SSR 位点分析

杜久军<sup>1,2</sup>, 左力辉<sup>1,2</sup>, 刘易超<sup>1,3</sup>, 于晓跃<sup>1,2</sup>, 董研<sup>1,2</sup>, 王进茂<sup>1,2</sup>, 杨敏生<sup>1,2</sup>

(<sup>1</sup> 河北农业大学林学院, 保定 071000; <sup>2</sup> 河北省林木种质资源与森林保护重点实验室, 保定 071000;

<sup>3</sup> 河北省林业科学研究院, 石家庄 050061)

**摘要:** 榆树材质优良, 具有良好的耐旱、耐寒和耐盐碱能力, 从温带、暖温带至亚热带都有分布。本研究对榆科植物(裂叶榆)的叶绿体基因组进行 *De novo* 测序, 裂叶榆叶绿体基因组序列全长为 158953 bp, 为典型的四段式结构, 其中 LSC 区长度为 88032 bp, SSC 区长 18846 bp, 两个 IR 区长 26037 bp, GC 含量为 35.57%。裂叶榆叶绿体基因组总共编码 139 个基因, 包括 85 个蛋白编码基因、8 个 rRNA 和 46 个 tRNA 基因。裂叶榆的叶绿体基因组存在 755 个 SSR 位点, SSR 序列长度主要以 6~8 bp 的短序列为主, SSR 共有 49 个重复单元, 以 A/T 和 AT/AT 为主, 占所有 SSR 位点的 66.09%。选取 42 个物种叶绿体基因组的共有蛋白编码基因进行系统进化分析表明, 裂叶榆与大麻科、桑科物种亲缘关系最近, 榆科与大麻科、桑科均属于荨麻目, 与传统分类学相吻合。本研究报道了裂叶榆的叶绿体基因组序列, 对今后榆树的光合作用研究、CP-SSR 引物开发、进化研究及叶绿体转基因工程研究具有重要意义。

**关键词:** 裂叶榆; 叶绿体基因组; CP-SSR

## Chloroplast Genome and CP-SSR Site Analyses of *Ulmus laciniata*

DU Jiu-jun<sup>1,2</sup>, ZUO Li-hui<sup>1,2</sup>, LIU Yi-chao<sup>1,3</sup>, YU Xiao-yue<sup>1,2</sup>,

DONG Yan<sup>1,2</sup>, WANG Jin-mao<sup>1,2</sup>, YANG Min-sheng<sup>1,2</sup>

(<sup>1</sup> Institute of Forest Biotechnology, Forestry College, Agricultural University of Hebei, Baoding 071000;

<sup>2</sup> Hebei Key Laboratory for Tree Genetic Resources and Forest Protection, Baoding 071000;

<sup>3</sup> Hebei Academy of Forestry, Shijiazhuang 050061)

**Abstract:** Elm (*Ulmus*) has a long history of using as high-quality hard wood and could grow in temperate, warm temperate, and subtropical regions for its strong ability to resist biotic stresses (drought, cold and salt). By *de novo* sequencing of the chloroplast genome of *Ulmus laciniata* Mayr, here we unlocked the chloroplast genome with a length of 158953 bp, which contains four distinct regions, including: a pair of inverted repeat regions (IR) of 26037 bp, large single-copy region (LSC) of 88032 bp, and small single-copy region (SSC) of 18846 bp. The chloroplast genome shows a GC content of 35.57% and contains a total of 139 genes, consisting of 85 protein encoding genes, 8 rRNA genes and 46 tRNA genes. A total of 755 SSR loci derived from 49 repeat units were found, and most of them were developed based on 6 to 8 bp short sequence. The large proportion of SSR loci was designed according to A/T and AT/AT (499, 66.09%). By phylogenetic analysis using the protein coding sequences of chloroplast genomes in 42 species, *U. laciniata* was found to be closely related to the plants of Cannabaceae and Moraceae, which belong to the genus Urticales, as revealed in the traditional taxonomy. Thus, this study provided a sequence of chloroplast genome in *Ulmus laciniata* Mayr, and this resource might be useful in understanding photosynthesis, evolution, primer development and chloroplast transgenic engineering.

**Key words:** *Ulmus laciniata*; chloroplast genome; CP-SSR

收稿日期: 2018-04-25 修回日期: 2018-06-02 网络出版日期: 2018-09-10

URL: <http://kns.cnki.net/kcms/detail/11.4996.s.20180907.0934.001.html>

基金项目: 国家自然科学基金(31370664); 河北农业大学科研发展基金(1699025)

第一作者研究方向为林业生物技术与抗逆育种。E-mail: dujiujun0517@163.com

通信作者: 杨敏生, 研究方向为林木遗传育种。E-mail: yangms100@126.com

榆树为榆科(Ulmaceae)榆属(*Ulmus* L.)植物,全世界大约有40余种,多分布于原苏联、中亚、哈萨克斯坦、外贝加尔、西伯利亚、蒙古、朝鲜等北半球国家和地区,在热带至寒温带均有分布<sup>[1-2]</sup>。中国有25种,以长江流域以北居多。榆树材质优良,具有良好的耐旱、耐寒和耐盐碱能力,是优良的造林绿化树种<sup>[3-4]</sup>。由于榆树枝叶繁茂、耐修剪等特性,被大量用于园林植物造型材料。

叶绿体是植物细胞中半自主性的细胞器,与线粒体、细胞核并称为三大遗传系统。叶绿体广泛存在于细胞质基质中,一般为绿色扁平椭球形或球形。叶绿体除参与光合作用外,也与淀粉、脂肪酸、色素和氨基酸的合成有密切联系<sup>[5-6]</sup>。叶绿体基因组包含大量的遗传信息,由于叶绿体的自我复制机制,有着相对独立的进化路线,根据叶绿体的遗传信息可以探究物种的发生、发展和演变,对植物基因组学和生物信息学的发展是不可或缺的。1940年,德国科学家首次公布叶绿体电镜照片,之后随着分子生物学及基因组学的发展,人类对叶绿体基因组的研究逐渐深入<sup>[7-10]</sup>。学者研究发现,尽管叶绿体基因组的总体结构相对保守,但仍然存在许多突变,有些是小范围内结构上的改变,例如倒位、易位,以及插入缺失等。到目前为止,已公布了超过800个完整的叶绿体基因组序列<sup>[11]</sup>,大大加速了叶绿体基因组学的研究进展。作为第三纪后古老树种之一,迄今为止,对榆科的遗传学与基因组学研究背景相当薄弱。榆科有230多种植物,对榆科植物进行叶绿体测序,将有助于榆科植物系统进化、物种鉴定、基因工程等相关工作的开展。本研究选取裂叶榆为试验材料,对叶绿体基因组进行*De novo*测序,通过比较基因组学研究确定榆树的叶绿体特征和系统进化地位,并为榆树的系统进化、标记开发、物种鉴定、遗传多样性分析等提供指导。

## 1 材料与方法

### 1.1 试验材料

本研究选取裂叶榆(*Ulmus laciniata* Mayr)为试验材料,在7月份生长旺盛季节,采集当年枝条上的嫩叶,置于干冰中,低温保存带回实验室,于-80℃保存,用于后续试验。

### 1.2 裂叶榆叶绿体基因组测序

采用史宝胜等<sup>[2]</sup>提出的高盐低pH的方法,将裂叶榆新鲜叶片经匀浆过滤和纯化后,获得叶绿

体,并用显微镜检测叶绿体的完整性,用DNase消化细胞核及线粒体等其他DNA,防止干扰。加入裂解液后使得叶绿体充分裂解,释放出cpDNA,经离心、抽提、纯化后获得叶绿体基因组DNA。经检测合格后,送北京源宜基因科技有限公司进行高通量测序。

**1.2.1 从头组装和间隙填充** 对原始数据进行过滤,去除影响数据质量和后续分析的低质量区域。处理后得到期望的高质量数据。先使用SOAPdenovo进行初步组装,得到contig序列。再用BLAT将组装得到的长序列定位到近缘物种的叶绿体参考基因组上,得到contig序列之间的相对位置;根据其相对位置对contig进行拼接并校正组装错误,得到叶绿体基因组全长框架图,使用GapCloser软件用高质量的短序列对框架图序列上的gap进行填补,然后用一代测序对剩余的gap及可疑区域进行补充和验证,再对LSC、SSC、IR区域连接处进行验证,最终得到一条环形的叶绿体基因组完成图序列。

**1.2.2 基因组注释和系统发育分析** 用CpGAVAS软件对叶绿体基因组序列进行注释。将裂叶榆叶绿体基因组与已发表的烟草、拟南芥、杨树和桑树等叶绿体基因组进行比对分析,探究裂叶榆叶绿体基因组的进化,确定榆树的分类学地位。

**1.2.3 SSR位点和RNA编辑分析** 使用MISA程序(<http://pgrc.ikp-gatersleben.de/misa>)对裂叶榆叶绿体基因组进行SSR位点搜索,搜索标准为:单核苷酸、二核苷酸、三核苷酸、四核苷酸和五核苷酸最少重复次数分别为8、3、3、3和3次,且SSR侧翼序列长度大于等于50bp,不同位点间最小间距为100bp。统计分析裂叶榆叶绿体SSR位点的数量和分布规律<sup>[11-20]</sup>。

## 2 结果与分析

### 2.1 裂叶榆叶绿体的基本特征

如表1所示,裂叶榆叶绿体基因组序列全长为158953bp,为典型的四段式结构,其中LSC区长度为88032bp,SSC区长18846bp,两个IR区长26037bp,GC含量为35.57%,上述结果与其他已经发表的植物叶绿体基因组类似。裂叶榆叶绿体基因组中,共有139个基因,LSC区、SSC区和IR区的基因数量分别为37个(26.62%)、13个(9.35%)和89个(64.03%),蛋白编码基因有85个,占比61.15%,核糖体RNA(rRNA)数量为8个,占比5.76%,转运RNA(tRNA)数量为46个,占比33.09%。

表 1 裂叶榆叶绿体基因组基本特征统计

Table 1 Basic characteristics of chloroplast genome of *U. laciniata*

特征 Feature	数值 Numerical value
GC 含量(%) GC content	35.57
AT 含量(%) AT content	64.43
LSC 区长度(bp) LSC length	88032
SSC 区长度(bp) SSC length	18846
IR 区长度(bp) IR length	26037
基因数量 Gene number	139
LSC 区基因数量 Gene number in LSC regions	37
SSC 区基因数量 Gene number in SSC regions	13
IR 区基因数量 Gene number in IR regions	89
蛋白编码基因数量 Protein-coding gene number	85
蛋白编码基因百分比(%) Protein-coding gene percentage	61.15
rRNA 数量 rRNA gene number	8
rRNA 百分比(%) rRNA percentage	5.76
tRNA 数量 tRNA gene number	46
tRNA 百分比(%) tRNA percentage	33.09
长度(bp) Length	158953

如图 1 所示,在叶绿体基因组中,有 4 个 rRNA 基因(16 S, 23 S, 4.5 S 和 5 S)和 5 个 tRNA 基因(*trnI-CAT*, *trnL-CAA*, *trnN-GTT*, *trnR-ACG*, *trnV-GAC*)和 5 个蛋白质编码基因(*rps7*, *rps12*, *rpl23*, *ycf1*, *ycf2*)均位于 IR 区域,这 14 个基因在叶绿体基因组中均为 2 个拷贝,需要指出的是 *ycf15* 基因在 IR 区存在 2 个拷贝,所以 *ycf15* 基因为 4 个拷贝。

对叶绿体基因组所有基因功能进行分类可知(表 2),裂叶榆叶绿体基因中与光合作用(genes for photosynthesis)和自我复制(self replication)相关的基因占绝大多数。在光合作用中,主要包括 ATP 合酶亚基(subunits of ATP synthase)、NADH-脱氢酶的亚基(subunits of NADH-dehydrogenase)、细胞色素 b/f 复合物的亚基(subunits of cytochrome b/f complex)、光系统 I 的亚基(subunits of photosystem I)、光系统 II 的亚基(subunits of photosystem II)和二磷酸核酮糖羧合酶/羧化酶亚基(subunit of rubisco)6 大类,其中与光系统 II 的亚基(subunits of photosys-

tem II)相关的基因数目最多,共 15 个基因;与二磷酸核酮糖羧合酶/羧化酶亚基(subunit of rubisco)相关的仅有一个基因。在自我复制(self replication)相关基因中,总共包括核糖体大亚基(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)5 大类。其中 tRNA 基因(tRNA genes)家族包含最多的 25 个基因,DNA 依赖性 RNA 聚合酶(DNA dependent RNA polymerase)类基因数目最少,仅有 4 个。

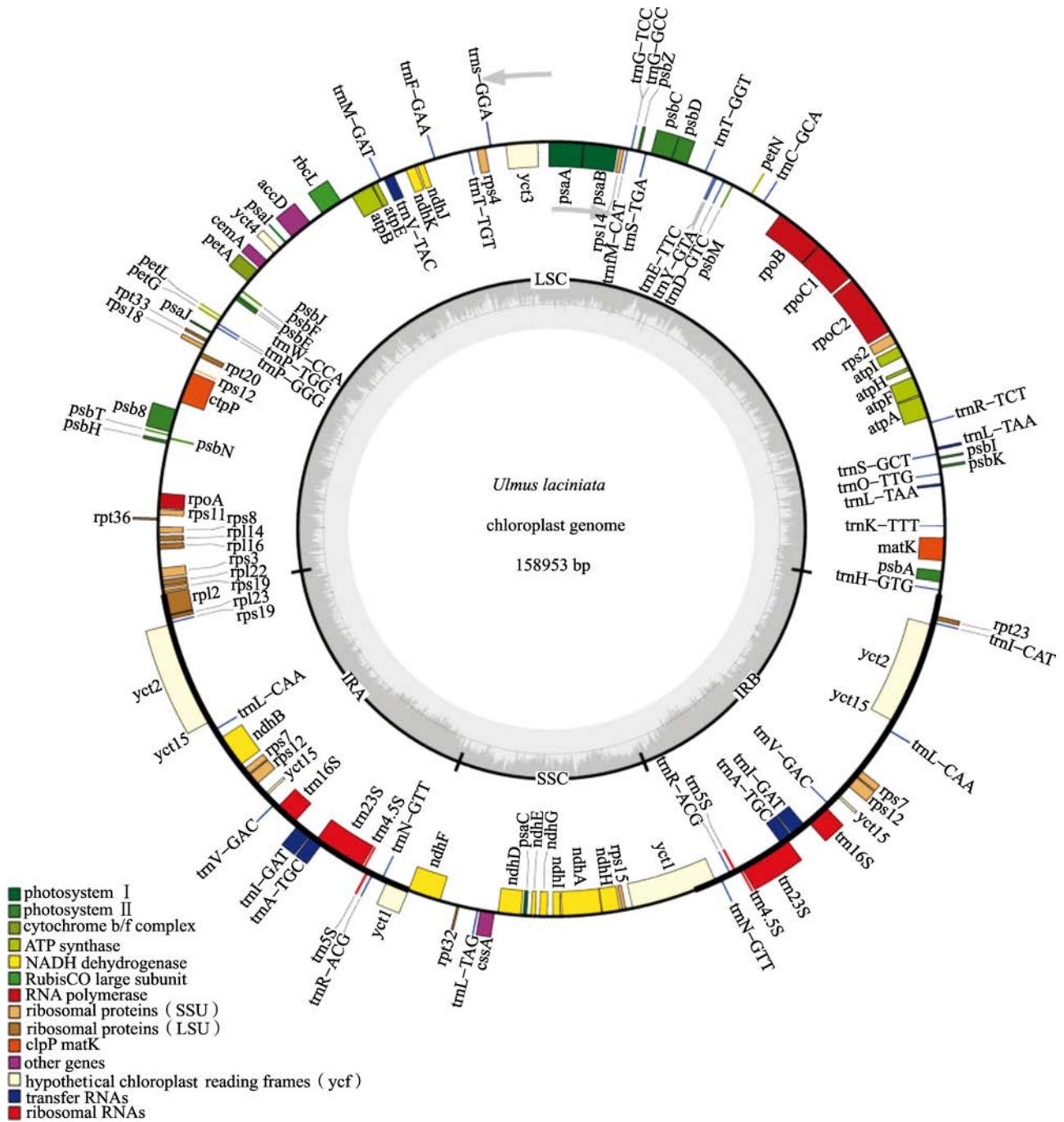
对裂叶榆叶绿体基因进一步分析可知(表 3),裂叶榆的叶绿体基因和其他植物的叶绿体基因类似,大多数基因均不存在内含子。本研究中裂叶榆叶绿体基因组中仅有 14 个基因存在内含子,除 *ycf3* 基因存在 2 个内含子外,其余均含有 1 个内含子。其中 *ycf1* 基因的内含子最长,达到 5675 bp;*trnL-TAA*的内含子最小,仅为 108 bp。

在裂叶榆叶绿体基因组中,有 61.76% 的序列为基因编码区,对所有蛋白编码基因的密码子进行统计分析可知(表 4),有 2554 个(10.47%)密码子编码亮氨酸 Leu,是编码率最高的氨基酸;而只有 314 个(1.11%)密码子编码半胱氨酸 Cys,是编码率最低的氨基酸。在蛋白编码区,密码子第 1 位、第 2 位和第 3 位的 AT 含量分别为 55.47%、62.88% 和 71.63%。这种第 3 位高 AT 含量的密码子编码偏好性在其他高等植物的叶绿体基因组中极其常见。

将裂叶榆叶绿体基因组与近亲物种桑树(*Morus mongolica* Schneid.)、模式植物烟草(*Nicotiana sylvestris*)、拟南芥(*Arabidopsis thaliana* (L.) Heynh.)、杨树(*Populus trichocarpa* Torr. & Gray)的叶绿体基因组进行共线性分析可知(图 2),裂叶榆叶绿体基因组与桑树同源性最高,且与其他模式物种彼此间均存在较多共线性关系,说明叶绿体基因在不同物种间具有较高的同源性。

## 2.2 重复序列分析

对裂叶榆叶绿体基因组的 1~5 核苷酸型 SSR 位点进行查找分析发现,裂叶榆叶绿体基因组总共存在 49 种重复单元,共 755 个 SSR 位点。由图 3 分析可知,不同 SSR 位点的重复单元数量存在较大差异,以二核苷酸和单核苷酸为主,分别占 60.00% (453) 和 28.74% (217),SSR 序列长度以 6~8 bp 的短序列为主,占全部的 64.50% (487)。



在圆内侧绘制的基因顺时针转录,在外侧的基因则是逆时针转录。属于不同功能组的基因是彩色编码的

Genes inside the circle are transcribed clockwise, and those outside are counter clockwise.

Genes with similar functional speculation are coded with same color

图1 裂叶榆叶绿体基因组的基因图谱

Fig.1 Gene map of chloroplast genome of *Ulmus laciniata* Mayr

表 2 裂叶榆叶绿体基因组表达基因

Table 2 List of genes present in chloroplast genome of *Ulmus laciniata* Mayr

功能 Functions	家族 Family name	基因代码 Gene code	基因 Genes
光合作用基因 Genes for photosynthesis	ATP 合酶亚基 Subunits of ATP synthase	atp	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	NADH-脱氢酶的亚基 Subunits of NADH-dehydrogenase	ndh	<i>ndhA, ndhB, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	细胞色素 b/f 复合物的亚基 Subunits of cytochrome b/f complex	pet	<i>petA, petB, petD, petG, petL, petN</i>
	光系统 I 的亚基 Subunits of photosystem I	psa	<i>psaA, psaB, psaC, psaI, psaJ</i>
	光系统 II 的亚基 Subunits of photosystem II	psb	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	二磷酸核酮糖羧合酶/羧化酶亚基 Subunit of rubisco	rbc	<i>rbcL</i>
自我复制 Self replication	核糖体大亚基 Large subunit of ribosome	rpl	<i>rpl14, rpl16, rpl2, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
	DNA 依赖性 RNA 聚合酶 DNA dependent RNA polymerase	rpo	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	核糖体小亚基 Small subunit of ribosome	rps	<i>rps11, rps12, rps14, rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps8</i>
	rRNA 基因 rRNA genes	rrn	<i>rrn16S, rrn23S, rrn4.5S, rrn5S</i>
	tRNA 基因 tRNA genes	trn	<i>trnC-GCA, trnD-GTC, trnE-TTC, trnF-GAA, trnG-GCC, trnH-GTG, trnI-CAT, trnL-CAA, trnL-TAA, trnL-TAG, trnM-CAT, trnN-GTT, trnP-TGG, trnQ-TTG, trnR-ACG, trnR-TCT, trnS-GCT, trnS-GGA, trnS-TGA, trnT-GGT, trnT-TGT, trnV-GAC, trnW-CCA, trnY-GTA, trnY-M-CAT</i>
其他基因 Other genes	乙酰-CoA-羧化酶的亚基 Subunit of Acetyl-CoA-carboxylase	acc	<i>accD</i>
	c 型细胞色素合成基因 c-type cytochrom synthesis gene	ccs	<i>ccsA</i>
	包膜蛋白基因 Envelop membrane protein	cem	<i>cemA</i>
	蛋白酶基因 Protease	clp	<i>clpP</i>
	转译起始因子 Translational initiation factor	inf	<i>infA</i>
	成熟酶基因 Maturase	mat	<i>matK</i>
未知功能基因 Genes of unknwn function	保守开放阅读框 Conserved open reading frames	yef	<i>yef1, yef15, yef2, yef3, yef4</i>

如表 5 所示,对所有 SSR 位点进行分析可知,在各类型 SSR 重复单元中,不同类型的 SSR 重复单元数量差异较大。在单核苷酸重复单元中,主要以 A/T 为主,占 99.54%;在二核苷酸重复单元中,AT/AT 的数量最多,达到 62.61%;而三核苷酸重复至五

核苷酸重复则相对较少。其中 A/T 和 AT/AT 重复单元占有所有 SSR 位点的 65.96%。裂叶榆叶绿体基因组的 SSR 位点富含 AT,与其他学者提出的植物叶绿体 SSR 位点主要以 AT 为主要的说法吻合,多聚 A 或多聚 T 也是叶绿体基因组高 AT 含量的一个重要原因。

表 3 基因中内含子和外显子的长度

Table 3 The length of introns and exons in the genes

基因 Gene	方向 Strand	起始 Start	终止 End	外显子 I(bp) Exon I	内含子 I(bp) Intron I	外显子 II(bp) Exon II	内含子 II(bp) Intron II	外显子 III(bp) Exon III
<i>trnL-TAA</i>	-	7095	7278	37	108	38		
<i>trnL-TAA</i>	+	9455	9643	37	119	32		
<i>rpoCl</i>	-	22177	25069	434	2892	1610		
<i>yef3</i>	-	44898	46985	126	1183	229	1133	152
<i>rpl2</i>	-	87910	89419	393	1509	469		
<i>ndhA</i>	-	124153	126758	555	2605	538		
<i>yef1</i>	-	128729	134404	2400	5675	3232		
<i>rps12</i>	-	101209	102007	231	798	28		
<i>rps12</i>	+	144980	145778	231	539	28		
<i>trnA-TGC</i>	+	106859	107732	36	804	36		
<i>trnA-TGC</i>	-	139256	140129	33	804	36		
<i>trnV-TAC</i>	-	54474	55181	35	635	37		
<i>trnI-GAT</i>	+	105774	106788	40	941	33		
<i>trnI-GAT</i>	-	140200	142214	33	941	40		

表 4 榆树叶绿体基因组中的密码子用法

Table 4 Codon Usage in chloroplast genome of *Ulmus laciniata* Mayr

密码子 Codon	氨基酸 Amino acid	数量 Number									
GCT	A	557	CAT	H	478	CCT	P	374	AGT	S	369
GCG	A	118	CAC	H	123	CCG	P	123	AGC	S	107
GCC	A	180	ATT	I	1049	CCC	P	172	ACT	T	504
GCA	A	351	ATC	I	377	CCA	P	304	ACG	T	124
TGT	C	209	ATA	I	684	CAG	Q	192	ACC	T	194
TGC	C	62	AAG	K	345	CAA	Q	693	ACA	T	404
GAT	D	810	AAA	K	1057	CGT	R	314	GTT	V	473
GAC	D	185	TTC	L	536	CGG	R	104	GTG	V	181
GAG	E	282	TTA	L	823	CGC	R	85	GTC	V	162
GAA	E	999	CTT	L	546	CGA	R	339	GTA	V	475
TTT	F	975	CTG	L	160	AGG	R	164	TGG	W	438
TTC	F	456	CTC	L	172	AGA	R	469	TAT	Y	747
GGT	G	523	CTA	L	317	TCT	S	546	TAC	Y	172
GGG	G	250	ATG	M	554	TCG	S	174	TGA	*	19
GGC	G	153	AAT	N	985	TCC	S	286	TAG	*	16
GGA	G	665	AAC	N	271	TCA	S	371	TAA	*	42

A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、W、Y 和 \* 分别指丙氨酸、半胱氨酸、天冬氨酸、谷氨酸、苯丙氨酸、甘氨酸、组氨酸、异亮氨酸、赖氨酸、亮氨酸、蛋氨酸、天冬酰胺、脯氨酸、谷氨酰胺、精氨酸、丝氨酸、苏氨酸、缬氨酸、色氨酸、酪氨酸、终止密码子

A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y and \* refer to Alanine, Cysteine, Aspartic acid, Glutamic acid, Phenylalanine, Glycine, Histidine, Isoleucine, Lysine, Leucine, Methionine, Asparagine, Proline, Glutamine, Arginine, Serine, Threonine, Valine, Tryptophan, Tyrosine and stop codon

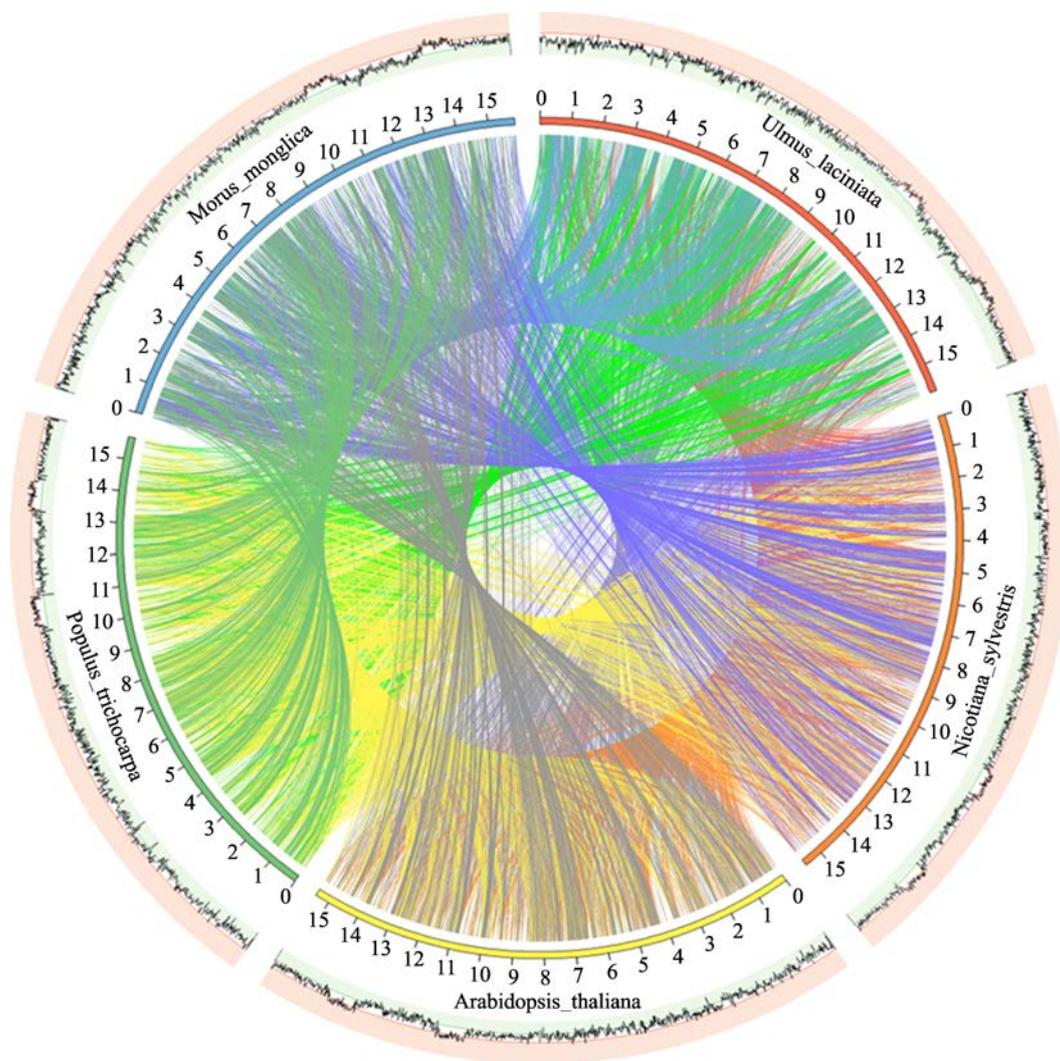


图 2 裂叶榆叶绿体基因组与其他植物叶绿体基因组比较

Fig. 2 Comparison of chloroplast genomes in *Ulmus laciniata* Mayr and other plant species

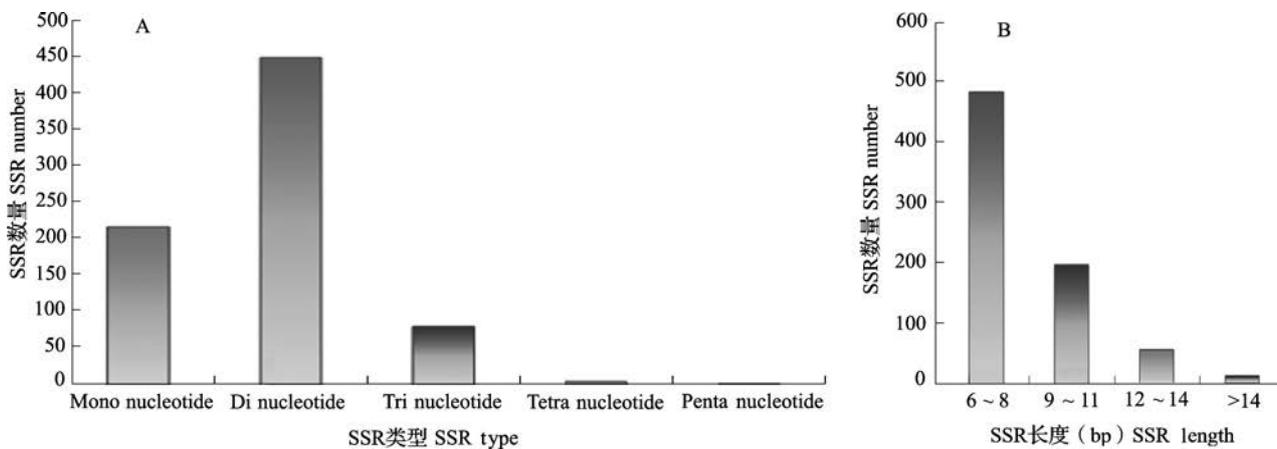


图 3 裂叶榆的叶绿体 SSR 类型和长度

Fig. 3 The type (A) and length (B) of the CP-SSR of *Ulmus laciniata* Mayr

表 5 裂叶榆叶绿体基因组 SSR 序列

Table 5 List of identified simple sequence repeats of the *Ulmus laciniata* Mayr chloroplast genome

重复碱基 SSR repeats	重复次数 Number of repeats									总计 Total
	3	4	5	6	7	8	9	10	>10	
A/T						54	42	48	72	216
C/G						1				1
AC/GT	43									43
AG/CT	107	11	1							119
AT/AT	220	43	11	5	4					283
CG/CG	8									8
AAC/GTT	5									5
AAG/CTT	24									24
AAT/ATT	26	8			1					35
ACT/AGT	1	1								2
AGC/CTG	6									6
AGG/CCT	3									3
ATC/ATG	3									3
AAAG/CTTT		1								1
AAAT/ATTT	5									5
AAAGT/ACTTT	1									1

### 2.3 系统发育分析

为确定裂叶榆在植物中的系统进化位置,参考 NCBI 已经公布的其他 42 种双子叶植物的叶绿体序列,选取 71 个共有蛋白编码基因进行系统进化分析。结果如下,聚类的支持率较高,有 31 个节点的检验分值达到 100%,说明聚类结果的可靠性较高。由图 4 分析可知,所有 43 个植物被分为 3 大类,第 I 类主要包括杨柳科、卫矛科、葡萄科、黄杨科、山茶科、茄科、五加科、十字花科等共 27 种植物;第 II 类主要包括胡桃科、蔷薇科、榆科和桑科共 14 种植物;第 III 类只包括 2 种豆科物种。由聚类图可知,裂叶榆与大麻科和桑科物种亲缘关系最近,其次为蔷薇科植物。榆科与大麻科、桑科均属于荨麻目,上述结果与传统分类学相吻合。

### 3 讨论

叶绿体是植物细胞内半自主的细胞器,是光合作用的主要场所,对植物的生长发育具有重要意义。叶绿体基因组的测序为进一步了解叶绿体基因表达、光合作用、细胞质相互作用等相关研究提供了依据。我们报道了裂叶榆高质量叶绿体基因组的组装和注释。裂叶榆的叶绿体基因组长度为 158953 bp,IR 区大小为 26037 bp,在被子植物叶绿体 120 ~ 180 kb,

IR 20 ~ 30 kb 的区间内<sup>[21]</sup>,具有典型被子植物叶绿体基因组环状四分体结构,包含 4 个不同的区域:一对为 26037 bp 的反向重复区 (IR),88032 bp 的大的单拷贝区 (LSC) 和一个小的单拷贝区 (SSC) 18846bp。叶绿体基因组 GC 含量为 35.57%,由 139 个基因组成。在这些基因中,13 个 (*trnL-TAA* -, *trnL-TAA* +, *rpoC1* -, *rpl2* -, *ndhA* -, *ycf1* -, *rps12* -, *rps12* +, *trnA-TGC* +, *trnA-TGC* -, *trnV-TAC* -, *trnI-GAT* +, *trnI-GAT* -) 含有单个内含子,1 个 (*ycf3* -) 含有 2 个内含子。裂叶榆蛋白编码基因密码子的第 3 个核苷酸具有明显的 AT 偏好,其中第 1 个碱基为 A/T 的密码子占总密码子数的 55.47%,第 2 个为 62.88%、第 3 个为 71.63%,这与其他高等植物类似<sup>[22]</sup>。共发现 755 个 SSR 位点,包含 49 个重复单位,其中,二核苷酸重复最多,共 453 个,占全部 SSR 总数的 60.00%;其次为单核苷酸重复,共 217 个,占全部 SSR 的 28.74%;再次为三核苷酸重复,共 78 个,占全部 SSR 总数的 10.33%。大量以二核苷酸为基序的 SSR 在其他物种中也较为罕见<sup>[21,23-29]</sup>,这一现象可能是筛选标准不同造成的。AG、AT、AAG、AGG 和 AGC 等是主要的重复基序类型,这点是与报道一致的<sup>[30]</sup>,A/T 和 AT/AT 重复单元占有所有 SSR 位点的 66.09%,多聚 A 或多聚 T 是

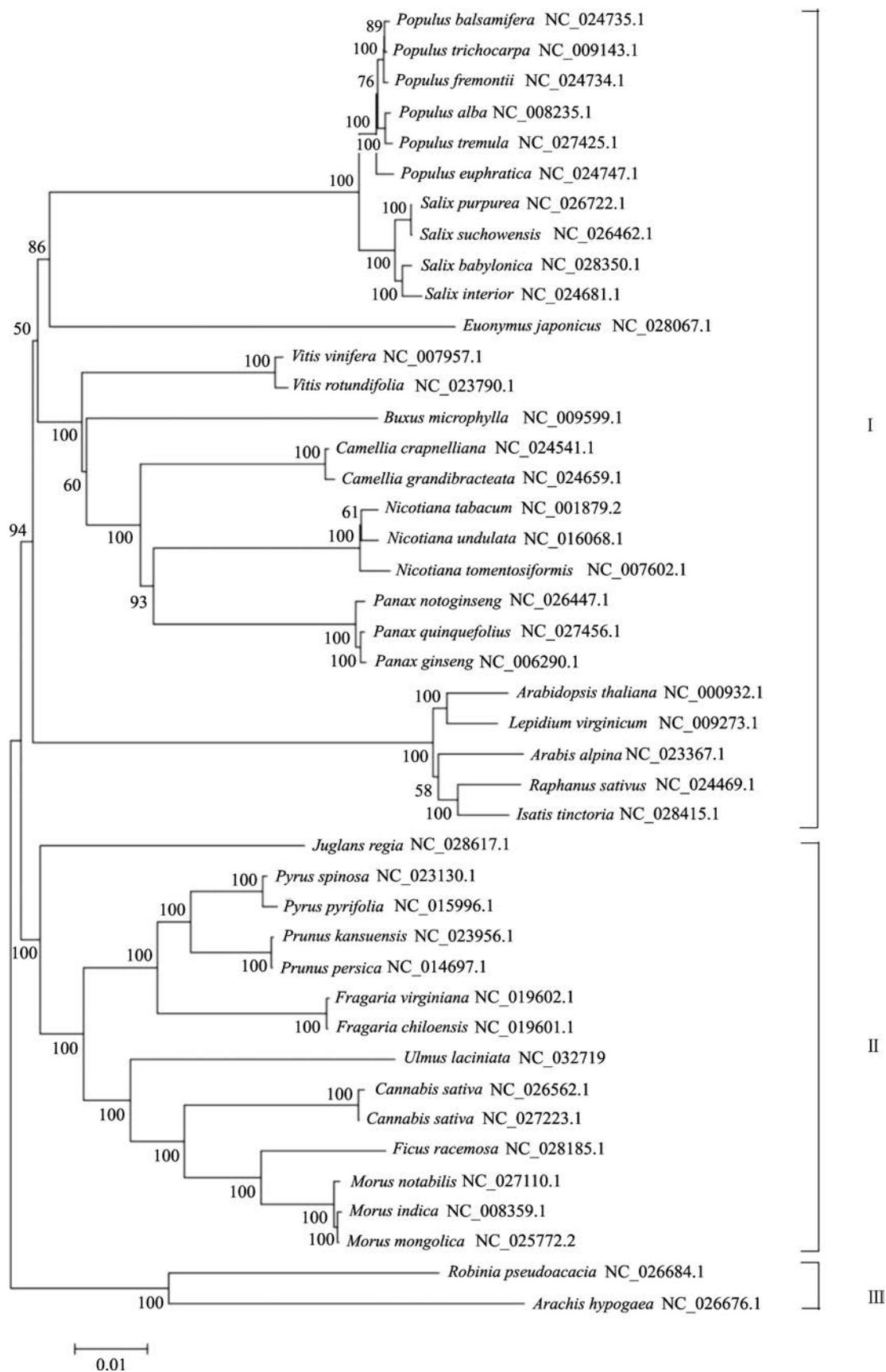


图 4 系统进化树

Fig. 4 The phylogenetic tree using maximum likelihood method

叶绿体基因组高 AT 含量的一个重要原因。大多数完美的 SSR 序列基于长度为 6 ~ 8 bp 的短序列,这与大量基于二核苷酸的 SSR 有关。与其他 42 种已经公布的植物叶绿体的共有蛋白编码基因进行系统进化分析表明,裂叶榆与大麻科和桑科物种亲缘关系最近,榆科、大麻科和桑科同属荨麻目,进化分析结果与传统分类学相吻合。本研究报道榆树叶绿体基因组,对于理解光合作用、进化、引物开发和叶绿体转基因工程具有重要意义。

#### 参考文献

- [1] 续九如,宋婉,邹受益,张卫东. 榆属树种遗传改良研究现状及思考. 北京林业大学学报,2000(6):95-99
- [2] 史宝胜,任子蓓,杨露,张卫东. 金叶榆叶片黄色素的提取及性质研究. 北方园艺,2000(16):133-136
- [3] Conde P, Sousa A, Costa A, Santos C. A protocol for *Ulmus minor* Mill. Micro propagation and acclimatization. Plant Cell Tissue and Organ Culture, 2008, 92:113-119
- [4] Dias M C, Monteiro C, Moutinho-Pereira J, Correia C, Gonçalves B, Santos C. Cadmium toxicity affects photosynthesis and plant growth at different levels. Acta Physiologiae Plantarum, 2013, 35: 1281-1289
- [5] Wolfe K H, Li W H, Sharp P M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences, 1987, 84(24):9054-9058
- [6] Ernes M J, Neuhaus H E. Metabolism and transport in non-photosynthetic plastids. Journal of Experimental Botany, 1997, 48(317): 1995-2005
- [7] Bremer B, Bremer K, Heidari N, Erixon P, Olmstead R G, Anderberg A A, Källersjö M, Barkhordarian E. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. Molecular Phylogenetics and Evolution, 2002, 24(2):274-301
- [8] 田欣,李德铎. DNA 序列在植物系统学研究中的应用. 云南植物研究, 2002(2):170-184
- [9] Leliaert F, Smith D R, Moreau H, Herron M D, Verbruggen H, Delwiche C F, Clerck O D. Phylogeny and molecular evolution of the green algae. Critical Reviews Plant Sciences, 2012, 31:1-46
- [10] Zuo L H, Shang A Q, Zhang S, Yu X Y, Ren Y C, Yang M S, Wang J M. The first complete chloroplast genome sequences of *Ulmus* species by *de novo* sequencing: Genome comparative and taxonomic position analysis. PLoS One, 2017, 12(2):1-19. DOI: 10.1371/journal.pone.0171264
- [11] 程慧,葛春峰,张红,乔玉山. 果树叶绿体基因组测序及系统发育研究进展. 核农学报, 2018, 32(1):58-69
- [12] 鄢秀芹,鲁敏,安华明. 刺梨转录组 SSR 信息分析及其分子标记开发. 园艺学报, 2015, 42(2):341-349
- [13] 李炎林,杨星星,张家银,黄三文,熊兴耀. 南方红豆杉转录组 SSR 挖掘及分子标记的研究. 园艺学报, 2014, 41(4): 735-745
- [14] 李满堂,张仕林,邓鹏,侯喜林,王建军. 洋葱转录组 SSR 信息分析及其多态性研究. 园艺学报, 2015, 42(6):1103-1111
- [15] 何海,郭继云,马毅平,周梦春,王沫,舒少华. 茯苓转录组 SSR 序列特征及其基因功能分析. 中草药, 2015, 46(23): 3558-3563
- [16] 李翠婷,张广辉,马春花,孟珍贵,陈军文,杨生超. 野三七转录组中 SSR 位点信息分析及其多态性研究. 中草药, 2014, 45(10):1468-1472
- [17] 刘越,岳春江,王翊,马加强,孙洪波,罗敏,马鹏举,张琳霞,马徐,陈川川,李华,唐丽. 藏茵陈川西樟叶菜转录组 SSR 信息分析. 中国中药杂志, 2015, 40(11):2068-2076
- [18] 张振,张含国,莫迟,张磊. 红松转录组 SSR 分析及 EST-SSR 标记开发. 林业科学, 2015, 51(8):114-120
- [19] 黄海燕,杜红岩,乌云塔娜,刘攀峰. 基于杜仲转录组序列的 SSR 分子标记的开发. 林业科学, 2013, 49(5):176-181
- [20] 文亚峰,韩文军,周宏,徐刚标. 杉木转录组 SSR 挖掘及 EST-SSR 标记规模化开发. 林业科学, 2015, 51(11):40-49
- [21] 宋菊,龙月红,林丽梅,尹峰,邢朝斌. 五加科植物叶绿体基因组结构与进化分析. 中草药, 2017, 48(24):5070-5075
- [22] 冯琛,张云婷,肖婕,汤浩茹. 黑莓 *RuTT12-1* 基因密码子偏好性分析. 基因组学与应用生物学, 2016, 35(8):2133-2144
- [23] 潘磊,李依,郭瑞,张凤银,曾长立,胡志辉,吴华,陈禅友. 豇豆叶绿体微卫星标记的开发及其在近缘种的通用性研究. 长江蔬菜, 2014(6):9-15
- [24] 李绪英,肖炳光,高玉龙,姬广海. 烟草叶绿体基因组和线粒体基因组 SSR 位点分析. 西北植物学报, 2011, 31(12): 2399-2405
- [25] 李巧丽,延娜,宋琼,郭军战. 鲁桑叶绿体基因组序列及特征分析. 植物学报, 2018, 53(1):94-103
- [26] 陈文超,苏琬涵,刘志高,季梦成. 铁线莲属植物叶绿体微卫星引物开发及其遗传分析. 浙江农业学报, 2017, 29(12): 2023-2031
- [27] 王文斌,于欢,邱相坡. 黄芩叶绿体基因组重复序列及密码子偏好性分析. 分子植物育种, 2018, 16(8):2445-2452
- [28] 高阳,王芙蓉,刘国栋,张景霞,张传云,张军. 基于陆地棉茎尖转录组序列 genic-SSR 标记的开发与评价. 植物遗传资源学报, 2015, 16(4):815-822
- [29] 魏明明,陈钰辉,刘富中,张映,连勇. 基于转录组测序的茄子 SSR 标记开发. 植物遗传资源学报, 2016, 17(6):1082-1091
- [30] 李炎林,杨星星,张家银,黄三文,熊兴耀. 南方红豆杉转录组 SSR 挖掘及分子标记的研究. 园艺学报, 2014, 41(4): 735-745