

· 农业生物技术 ·

红锥叶绿体基因组特征及系统发育分析

薛光宇^{1,2}, 邓智文³, 朱雪萍^{1,2}, 吴俊多^{1,2}, 董诗涛^{1,2}, 谢先金^{1,2}, 曾冀^{1,2*}

1 中国林业科学研究院热带林业实验中心, 广西 凭祥 532600

2 广西友谊关森林生态系统国家定位观测研究站, 广西 凭祥 532600

3 福建农林大学林学院, 福建 福州 350002

薛光宇, 邓智文, 朱雪萍, 吴俊多, 董诗涛, 谢先金, 曾冀. 红锥叶绿体基因组特征及系统发育分析[J]. 生物工程学报, 2023, 39(2): 670-684.

XUE Guangyu, DENG Zhiwen, ZHU Xueping, WU Junduo, DONG Shitao, XIE Xianjin, ZENG Ji. Chloroplast genomic characterization and phylogenetic analysis of *Castanopsis hystrix*[J]. Chinese Journal of Biotechnology, 2023, 39(2): 670-684.

摘 要: 为确定红锥(*Castanopsis hystrix*)叶绿体基因组的结构组成情况, 判定其在锥属中的进化位置及与同锥属叶绿体基因组的区别, 为锥属物种鉴定、遗传多样性分析和资源保护提供相关依据。使用 Illumina HiSeq 2500 测序平台对红锥叶绿体基因组进行测序, 通过生物信息学分析方法进行序列组装、注释和特征分析, 并利用 R、Python、MISA、CodonW 和 MEGA 6 等生物信息学软件对其基因组结构和数目、密码子偏好性、序列重复、简单重复序列(simple sequence repeat, SSR)位点和系统发育进行分析。结果表明红锥叶绿体基因组大小为 153 754 bp, 呈现四分体结构; 共拥有 130 个基因, 包含 85 个编码基因、37 个 tRNA 基因和 8 个 rRNA 基因; 通过密码子偏好性分析, 平均有效密码子数为 55.5, 说明其密码子随机性强、偏好性低; 通过 SSR 及长重复片段分析, 检测到 45 个重复序列及 111 个 SSR 位点; 与近缘种比较, 发现其叶绿体基因组序列高度保守, 尤其蛋白质编码序列相似度极高; 此外, 系统发育分析发现红锥与海南锥聚为一支, 关系密切。本研究得到了红锥的叶绿体基因组基本情况与系统发育位置, 为红锥的物种辨别、天然种群遗传多样性与功能基因组学提供前期研究铺垫。

关键词: 红锥; 叶绿体基因组; 密码子偏好性; 系统发育

资助项目: 国家自然科学基金(31971655); 中国林业科学研究院热带林业实验中心科学计划项目(RL2020-05)

This work was supported by the National Natural Science Foundation of China (31971655) and the Scientific Program Project of Experimental Centre of Tropical Forestry, Chinese Academy of Forestry (RL2020-05).

*Corresponding author. E-mail: richzeng@caf.ac.cn

Received: 2022-08-04; Accepted: 2022-10-10; Published online: 2022-10-12

Chloroplast genomic characterization and phylogenetic analysis of *Castanopsis hystrix*

XUE Guangyu^{1,2}, DENG Zhiwen³, ZHU Xueping^{1,2}, WU Junduo^{1,2}, DONG Shitao^{1,2},
XIE Xianjin^{1,2}, ZENG Ji^{1,2*}

1 Experimental Centre of Tropical Forestry, Chinese Academy of Forestry, Pingxiang 532600, Guangxi, China

2 Guangxi Youyiguan Forest Ecosystem Research Station, Pingxiang 532600, Guangxi, China

3 College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

Abstract: The structure and size of the chloroplast genome of *Castanopsis hystrix* was determined by Illumina HiSeq 2500 sequencing platform to understand the difference between *C. hystrix* and the chloroplast genome of the same genus, and the evolutionary position of *C. hystrix* in the genus, so as to facilitate species identification, genetic diversity analysis and resource conservation of the genus. Bioinformatics analysis was used to perform sequence assembly, annotation and characteristic analysis. R, Python, MISA, CodonW and MEGA 6 bioinformatics software were used to analyze the genome structure and number, codon bias, sequence repeats, simple sequence repeat (SSR) loci and phylogeny. The genome size of *C. hystrix* chloroplast was 153 754 bp, showing tetrad structure. A total of 130 genes were identified, including 85 coding genes, 37 tRNA genes and 8 rRNA genes. According to codon bias analysis, the average number of effective codons was 55.5, indicating that the codons were highly random and low in bias. Forty-five repeats and 111 SSR loci were detected by SSR and long repeat fragment analysis. Compared with the related species, chloroplast genome sequences were highly conserved, especially the protein coding sequences. Phylogenetic analysis showed that *C. hystrix* is closely related to the Hainanese cone. In summary, we obtained the basic information and phylogenetic position of the chloroplast genome of red cone, which will provide a preliminary basis for species identification, genetic diversity of natural populations and functional genomics research of *C. hystrix*.

Keywords: *Castanopsis hystrix*; chloroplast genome; codon bias; phylogeny

红锥(*Castanopsis hystrix*)属于壳斗科(Fagaceae)锥属(*Castanopsis*), 别名刺栲, 是常绿硬木类珍贵用材树种^[1], 属于地带性亚热带常绿阔叶林的主要建种群及顶级群落种, 树体高大, 干形笔直, 在天然林中心位置分布的红锥树体胸径 150 cm 以上, 高 30 m 以上, 具有常绿、速生、寿命长等特点^[2]。红锥是我国热带亚热带地区生长较快的国家Ⅱ级保护珍稀阔叶树种, 具

有重要的经济价值和生态价值, 主要分布于我国福建、湖南、广东、海南、广西、贵州、西藏地区 30–1 600 m 的坡地和山地常绿阔叶林中^[3], 其木材坚硬耐腐, 心材颜色淡红, 不但美观且容易加工, 是深受产区群众喜爱的珍贵木材, 在家装家饰、木材艺术、船舶制造等行业具有较大的发展前景^[4]。因此开展红锥叶绿体基因组特征分析, 说明红锥在壳斗科植物中

的系统发育位置及与锥属其他植物的遗传关系,对锥属物种鉴定、遗传多样性分析及资源保护具有重要意义。

叶绿体通过光合作用把太阳能转换为化学能,是植物体内活跃的代谢枢纽,对于生态系统的平衡和稳定作用显著^[5];光合作用严格受遗传控制,具备独立于核基因组的完整叶绿体基因组,主要用于编码与光合作用紧密相关的蛋白质和核糖体蛋白,且被子植物的叶绿体基因组是母系遗传^[6]。陆生植物叶绿体基因组以环状形式存在于叶绿体中,大小为107–218 kb,包括120–130个基因^[7];其结构具有高度保守性,由2个反向重复区域(inverted repeat, IR)、1个小单拷贝区域(small single copy, SSC)、1个大单拷贝区域(large single copy, LSC)组成,叶绿体基因组在揭示物种起源、生物进化及不同物种之间的亲缘关系等方面作用重要^[8]。随着叶绿体基因组学快速发展,多种植物的叶绿体基因组被测序,且广泛应用于类群间的物种天然种群遗传多样性与功能基因组学、分类学研究等方面^[9]。本研究通过Illumina HiSeq 2500平台测序红锥叶绿体,并对红锥叶绿体基因组进行密码子偏好性分析、简单重复序列(simple sequence repeat, SSR)位点分析、叶绿体基因组比较分析和系统发育分析,为红锥的物种辨别、功能基因组学研究、多角度开发与利用提供前期铺垫。

1 材料与方法

1.1 材料

样品的新鲜叶片采自中国林业科学研究院热带林业实验中心苗圃(22°7'22.0"N, 106°44'38.7"E),将健康叶片装入取样袋后带回实验室,用无菌水冲洗数次,擦干后置于干冰中保存备用。

1.2 方法

1.2.1 基因组 DNA 提取及测序

用70%的乙醇处理约5 g新鲜红锥叶片后,用TruSeq DNA sample Preparation Kit提取红锥基因组DNA。采用Illumina HiSeq 2500平台开展测序, QIAquick Gel Extraction Kit电泳和Invitrogen Qubit Spectrophotometer显微分光光度计检测DNA相关指标^[10]。

1.2.2 叶绿体基因组组装和注释

利用SOAPdenovo2.04软件,以帽斗栎(*Quercus guyavifolia*) (NC_050963)叶绿体基因组序列作为参考序列进行组装,再对组装结果进行进一步检查分析,多频次整理后获得红锥完整叶绿体基因组,随后采用DOGMA对红锥叶绿体基因组进行注释,结合蛋白质直系同源组群(cluster of orthologous group, COG)等数据库对基因组进行注释^[11]。通过NCBI把注释好的序列在线提交,经审核后获得序列登录号OM321038。

1.2.3 密码子偏好性

采用python CAI包中的RSCU模块计算红锥叶绿体基因组中相对同义密码子使用频率(relatively synonymous codon usage, RSCU)和蛋白编码基因的密码子使用频率(大于200 bp),利用Codon W软件对红锥叶绿体基因组蛋白质编码区的密码子数量、密码子使用频率和同义密码子相对使用度进行分析^[12]。

1.2.4 重复序列和 SSR 位点分析

使用在线软件REPuter对红锥叶绿体基因组进行长重复序列分析^[13]。参数设置如下: maximum computed repeats 设置为1 000, minimal repeat size 设置为30, Hamming distance 设置为3, 其他设置默认参数, SSR位点分析通过采用在线程序MISA进行^[14], 其中单、2、3、3、5和6核苷酸的重复数分别设置为10、5、4、3、3、3。

1.2.5 叶绿体基因组比较分析

利用 IRscope 可视化工具, 比较红锥、锥栗、米楮、长果锥、华南锥和海南锥叶绿体基因组四分体边界的差异。并采用 mVISTA 对其进行全基因组比对分析^[15]。

1.2.6 系统发育分析

从 NCBI 下载壳斗科锥属的米楮(*Castanopsis carlesii*, MK840978.1)、海南锥(*C. hainanensis*, OL581718.1)、华南锥(*C. concinna*, KT793041.1)、长果锥(*C. sieboldii*, MZ028444.1)、湄公锥(*C. mekongensis*, NC 053865.1)、短刺锥(*C. echinocarpa*, KJ001129.1)、栗属的日本栗(*Castanea crenata*, MN402457.1)、欧洲栗(*C. sativa*, MW327507.1)、锥栗(*C. henryi*, KX954615.1)、栗(*C. mollissima*, KY951992.1)、茅栗(*C. seguinii*, NC039749.1)共 11 个物种完整的叶绿体基因组序列, 同时以壳斗科栎属植物星毛栎(*Quercus stellata*, MK105467.1)为外类群, 进行基因组序列的多重比对, 采用最大似然法(maximum likelihood, ML)对系统进化关系进行分析。使用 MEGA 6 软件生成 ML 系统发育树, 使用适合模型“GTR+G+I”, 自展值 bootstrap 值设为 1 000^[16]。

2 结果与分析

2.1 红锥叶绿体基因组结构

由图 1 可以看出, 红锥叶绿体基因组是一个环状分子基因组, 具有典型的四分区域结构, 包含了 1 个大单拷贝区(large single-copy, LSC)、1 个小单拷贝区(small single-copy region, SSC)和 1 对反向互补重复区(inverted repeats, IR)^[6], 大小分别为 90 318、18 957、25 699 bp。此外, LSC、IR 和 SSC 区域的 GC 值存在差别, 碱基组成结果见表 1。由表 1 可看出, 红锥叶绿体

基因组总长 160 673 bp, G/C 含量为 36.8%, A/T 含量为 63.2%, AT、GC 偏斜分别是-0.012 7、-0.016 3; 其中 LSC 区长度为 90 318 bp, A+T 含量为 65.3%, AT、GC 偏斜分别是-0.023 0、-0.023 1; SSC 区长度为 18 957 bp, A+T 含量为 69.1%, AT、GC 偏斜分别是-0.001 4、-0.055 0; IR 长度为 26 057 bp, A+T 含量为 57%; 从整体上分析, LSC 与 SSC 区的 A+T 含量明显高于 IR 区; 从 AT、GC 偏斜值来看, 全基因组双链上碱基组成的不对称性较低。

2.2 红锥叶绿体基因组功能及分类

红锥叶绿体基因组共注释出 130 个基因, 编码蛋白质基因、tRNA 基因、rRNA 基因分别是 85、37 和 8 个, 如表 2 所示将这些基因根据功能分为 4 类^[17]。由表 2 可以看出, 在注释的 130 个基因中, 共有 19 个多拷贝基因, 其中包括 1 个 NADH 脱氢酶基因(*ndhB*)、2 个核糖体小亚基基因(*rps7* 和 *rps12*)、3 个核糖体大亚基基因(*rpl2*、*rpl12*、*rpl23*)、4 个 rRNA 基因(*rrn4.5*、*rrn5*、*rrn16*、*rrn23*)、1 个未知功能基因(*ycf2*)和 8 个 tRNA 基因(*trnS-GCU*、*trnI-CAU*、*trnL-CAA*、*trnV-GAC*、*trnI-GAU*、*trnA-UGC*、*trnR-ACG*、*trnN-GUU*)。

2.3 密码子偏好性分析

红锥叶绿体中基因组的密码子中 GC 的含量比例是 37.9%, 其中密码子第 3 个碱基是 GC 的比例是 35.4%, 有效密码子数(effective number of codon, ENC)为 55.504, 密码子中 GC 的比例是 37.9%, 其中密码子第 3 个碱基是 GC 的比例为 35.4%, 低于 50%; 此外, 还检测到 31 个同义密码子相对使用度(relative synonymous codon usage, RSCU)大于 1 的密码子, 除 UUG 外其余密码子都以 A/U 结尾, RSCU 值最高的是编码亮氨酸(Leu)的密码子 TTA (1.895) (表 3 和图 2)。

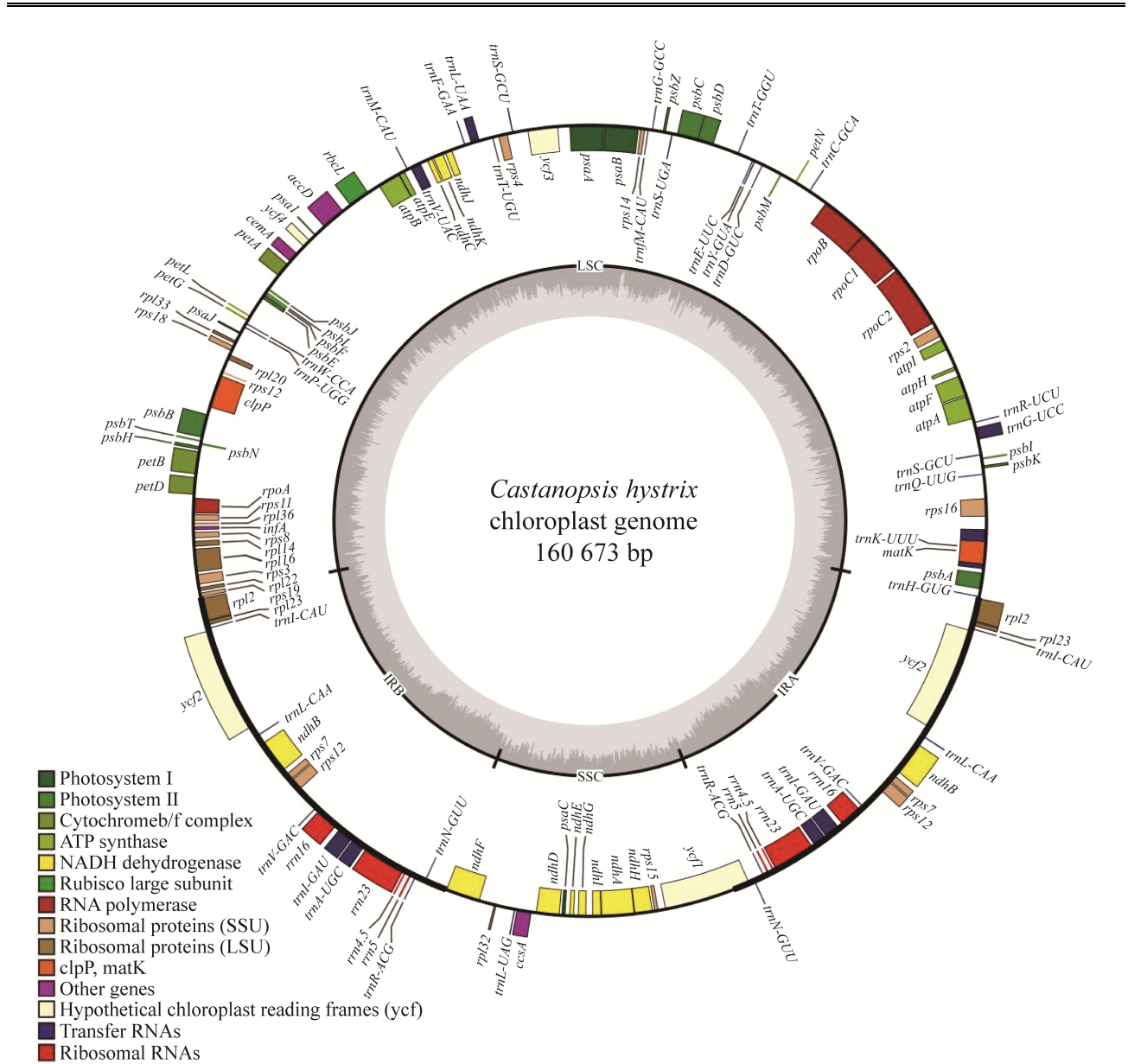


图 1 红锥叶绿体全基因组图谱
Figure 1 Whole genome map of *Castanopsis hystrix* chloroplast.

表 1 红锥叶绿体全基因组核苷酸组成
Table 1 Nucleotide composition of the whole genome of *Castanopsis hystrix* chloroplast

Area	T (%)	C (%)	A (%)	G (%)	GC (%)	Base length (bp)	AT skew	GC skew
LSC	33.4	17.7	31.9	16.9	34.6	90 318	-0.023 0	-0.023 1
SSC	34.6	16.3	34.5	14.6	30.9	18 957	-0.001 4	-0.055 0
IRa	28.8	22.2	28.5	20.6	42.8	25 699	-0.005 2	-0.037 4
IRb	28.5	20.6	28.8	22.2	42.8	25 699	0.005 2	0.037 4
Total	32.0	18.7	31.2	18.1	36.8	160 673	-0.012 7	-0.016 3

AT skew=(A-T)/(A+T); GC skew=(G-C)/(G+C).

表 2 红锥叶绿体基因组所注释到的基因

Table 2 Genes annotated in the *Castanopsis hystrix* chloroplast genome

Gene function	Gene type	Gene name
Genes involved in photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Subunits of photosystem II	<i>psbA, psbK, psbI, psbD, psbC, psbJ, psbL, psbF, psbE, psbB, psbT, psbN, psbH, psbM, psbZ</i>
	Subunits of ATP synthase	<i>atpA, atpF, atpH, atpI, atpE, atpB</i>
	Subunits of NADH dehydrogenase	<i>ndhA, ndhB, ndhC, ndhD, ndhJ, ndhK, ndhF, ndhE, ndhG, ndhI, ndhH</i>
	Subunits of cytochrome b/f complex	<i>petN, petA, petL, petG, petB, petD</i>
	Subunit of Rubisco	<i>rbcL</i>
Genes involved in protein expression	Small subunit of ribosome	<i>rps12, rps16, rps2, rps14, rps4, rps18, rps12*, rps11, rps8, rps3, rps19, rps7*, rps15*</i>
	Large subunit of ribosome	<i>rpl33, rpl20, rpl36, rpl14, rpl16, rpl22, rpl2, rpl23, rpl32, rpl23, rpl2</i>
	rRNA	<i>rrn16*, rrn23*, rrn4.5*, rrn5*</i>
	DNA dependent RNA polymerase	<i>rpoC1, rpoC2, rpoB, rpoA</i>
	tRNA	<i>trnH-GUG, trnK-UUU, trnQ-UUG, trnS-GCU, trnG-UCC, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-GCU, trnT-UGU, trnL-UAA, trnF-GAA, trnS-UGA, trnG-GCC, trnM-CAU, trnV-UAC, trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU, trnL-CAA, trnV-GAC, trnI-GAU, trnA-UGC, trnR-ACG, trnN-GUU, trnL-UAG, trnN-GUU, trnR-ACG, trnA-UGC, trnI-GAU, trnV-GAC, trnL-CAA, trnI-CAU</i>
	Subunit of acetyl-CoA carboxylase	<i>accD</i>
Other genes	Envelop membrane protein	<i>cemA</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
	Protease	<i>clpP</i>
	Translational initiation factor	<i>infA</i>
Unknown genes	Putative open reading frames	<i>matK</i>
		<i>ycf3, ycf4, ycf2*, ycf1*</i>

*: Two gene copies.

表 3 红锥各氨基酸相对同义密码子使用度

Table 3 Relative synonymous codon usage of each amino acid in *Castanopsis hystrix*

Amino acid	Total	AA frequency (%)	Codon	Count	Codon frequency (%)	RSCU
L	2 780	10.6	CTA	366	13.2	0.790
			CTC	197	7.1	0.425
			CTG	194	7.0	0.419
			CTT	578	20.8	1.247
			TTA	878	31.6	1.895
			TTG	567	20.4	1.224
I	2 310	8.8	ATA	746	32.3	0.969
			ATC	453	19.6	0.588
			ATT	1 111	48.1	1.443
S	1 992	7.6	AGC	123	6.2	0.370
			AGT	388	19.5	1.169
			TCA	397	19.9	1.196
			TCC	349	17.5	1.051
			TCG	189	9.5	0.569
			TCT	546	27.4	1.645

(待续)

(续表 3)

Amino acid	Total	AA frequency (%)	Codon	Count	Codon frequency (%)	RSCU
G	1 800	6.9	GGA	696	38.7	1.547
			GGC	207	11.5	0.460
			GGG	324	18.0	0.720
			GGT	573	31.8	1.273
R	1 587	6.1	AGA	496	31.3	1.875
			AGG	176	11.1	0.665
			CGA	353	22.2	1.335
			CGC	112	7.1	0.423
			CGG	119	7.5	0.450
F	1 496	5.7	CGT	331	20.9	1.251
			TTC	529	35.4	0.707
K	1 421	5.4	TTT	967	64.6	1.293
			AAA	1 052	74.0	1.481
V	1 418	5.4	AAG	369	26.0	0.519
			GTA	541	38.2	1.526
A	1 398	5.3	GTC	175	12.3	0.494
			GTG	202	14.2	0.570
			GTT	500	35.3	1.410
			GCA	383	27.4	1.096
			GCC	222	15.9	0.635
E	1 390	5.3	GCG	164	11.7	0.469
			GCT	629	45.0	1.800
			GAA	1 040	74.8	1.496
T	1 321	5.0	GAG	350	25.2	0.504
			ACA	397	30.1	1.202
			ACC	243	18.4	0.736
			ACG	155	11.7	0.469
			ACT	526	39.8	1.593
N	1 287	4.9	AAC	301	23.4	0.468
			AAT	986	76.6	1.532
P	1 095	4.2	CCA	311	28.4	1.136
			CCC	224	20.5	0.818
			CCG	154	14.1	0.563
			CCT	406	37.1	1.483
D	1 069	4.1	GAC	208	19.5	0.389
			GAT	861	80.5	1.611
Y	985	3.8	TAC	206	20.9	0.418
			TAT	779	79.1	1.582
Q	935	3.6	CAA	725	77.5	1.551
			CAG	210	22.5	0.449
H	625	2.4	CAC	137	21.9	0.438
			CAT	488	78.1	1.562
M	523	2.0	ATG	523	100.0	1.000
W	454	1.7	TGG	454	100.0	1.000
C	303	1.2	TGC	86	28.4	0.568
			TGT	217	71.6	1.432

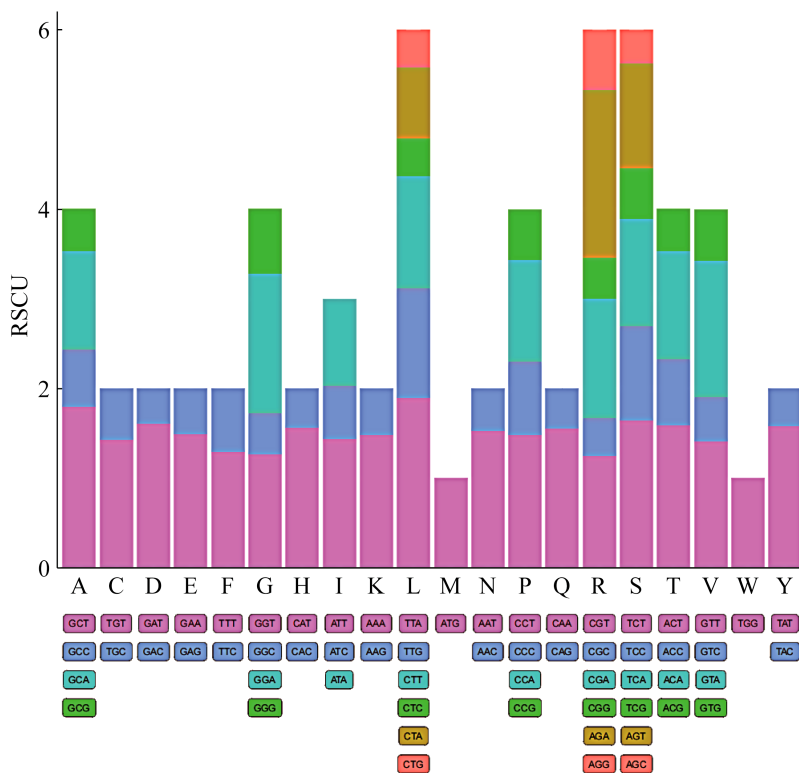


图2 红锥叶绿体基因组密码子同义密码子相对使用度分布情况

Figure 2 Distribution of codon RSCU in *Castanopsis hystrix* chloroplast genome.

2.4 重复序列和 SSR 位点分析

通过检测红锥叶绿体基因组, 获得了 45 个长重复序列(长度范围在 30–77 bp), 其中包含正向重复、回文重复、反向重复、互补重复分别是 27、17、3 和 1 个(图 3)。这些重复序列很大部分位于基因间区(intergenic spacer, IGS)和内含子区域, 定位于 LSC 的长序列重复占 53.33%, 定位于 IR 的长序列重复占 11.11%, 定位于 SSC 的长序列重复占 35.56% (图 3)。红锥叶绿体基因组中检测出 111 个 SSR 位点, 共 13 种类型(表 4)。单、2、3、4、5、6 核苷酸 SSR 位点数目分别占比为 64.86%、15.32%、6.36%、9.09%、3.64%和 0.90% (表 3)。AG/CT 和 AT/AT 数目分别是 3、14 个, AAT/ATT 数目为 6 个, AAAT/ATTT 数目为 7 个。位于 LSC 的 SSR 位点占 79.31%, 位于 SSC 的 SSR 位点

占 16.09%, 只有 4.60%的 SSR 位点位于 IR。红锥叶绿体基因组 SSR 序列偏好 A/T 碱基。

2.5 叶绿体基因组比较分析

通过对锥属 6 种植物的叶绿体基因组 IR 与 SSC 边界区域收缩与扩张研究(图 4), 6 种锥属叶绿体基因组的长度大小无明显差别, 但有较为明显差别的是边界基因的位置和类型。*rps19*、*rps19*、*rpl22* 基因分布在边界 LSC/IRb 中, 其中锥栗、华南锥的边界虽均为 *rps19* 基因, 但是分布在 LSC 和 IRb 区域中的大小不同, LSC (256 bp 和 295 bp), IRb (20 bp 和 41 bp); *ycf1*、*ndhF* 基因分布在边界 IRb/SSC 中, 其中锥栗和红锥的边界基因为 *ndhF*, 但是分布在 IRb 和 SSC 区域中的大小不同, IRb (32 bp 和 2 bp), SSC (2 248 bp 和 2 261 bp), *ycf1* 基因在米槠、长果锥、华南锥和海南锥中跨越 IRb/SSC 边界

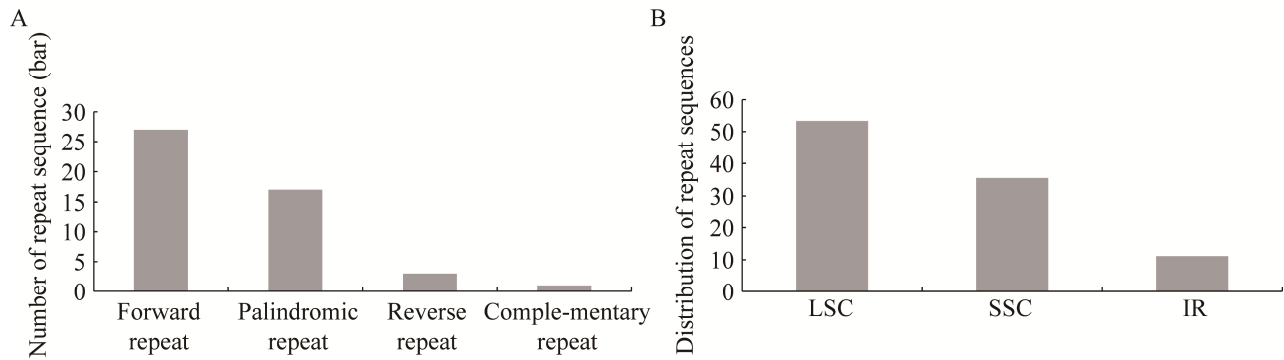


图3 红锥叶绿体基因组长序列重复类型及分布结果

Figure 3 Types and distribution of long sequence repeats in the *Castanopsis hystrix* chloroplast genome. A: Number of repeat sequence. B: Distribution of repeat sequences.

表4 红锥叶绿体全基因组单核苷酸 SSRs 类型及分布

Table 4 SSR type and distribution in the complete genome of *Castanopsis hystrix*

SSR repeats	Repeat times														Total
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
A/T								30	25	4	7	1	1		68
C/G											2	1		1	4
AG/CT			3												3
AT/AT			9	4	1										14
AAG/CTT		1													1
AAT/ATT		6													6
AAAT/ATTT	5	2													7
AATG/ATTC	1														1
AATT/AATT	2														2
AAAAT/ATTTT	1														1
AATAG/ATTCT	1														1
AATGC/ATTGC	2														2
AATATC/ATATTG	1														1

且 *ndhF* 均分布于 SSC 区域; *ycf1* 基因分布在边界 SSC/IRa 中, 大小不同且横跨该区域, 6 种锥属植物均有分布; *rpl2*、*trnH*、*psbA* 3 个基因分布在边界 IRa/LSC 中, 且无基因跨越该区域。

2.6 系统发育分析

为了揭示红锥与壳斗科其他成员的系统发

育位置, 本研究对壳斗科 11 个完整的叶绿体基因组和一个外类群进行了系统发育分析, 相关数据均从 NCBI GenBank 下载。系统发育树结果显示, 红锥与海南锥亲缘关系密切, 所有分支节点的后验概率(posterior probability, PP)证明了树的高可信度, 且这个小分支与华南锥组成的小支位于同一分支上, 支持度同样是 100% (图 5)。

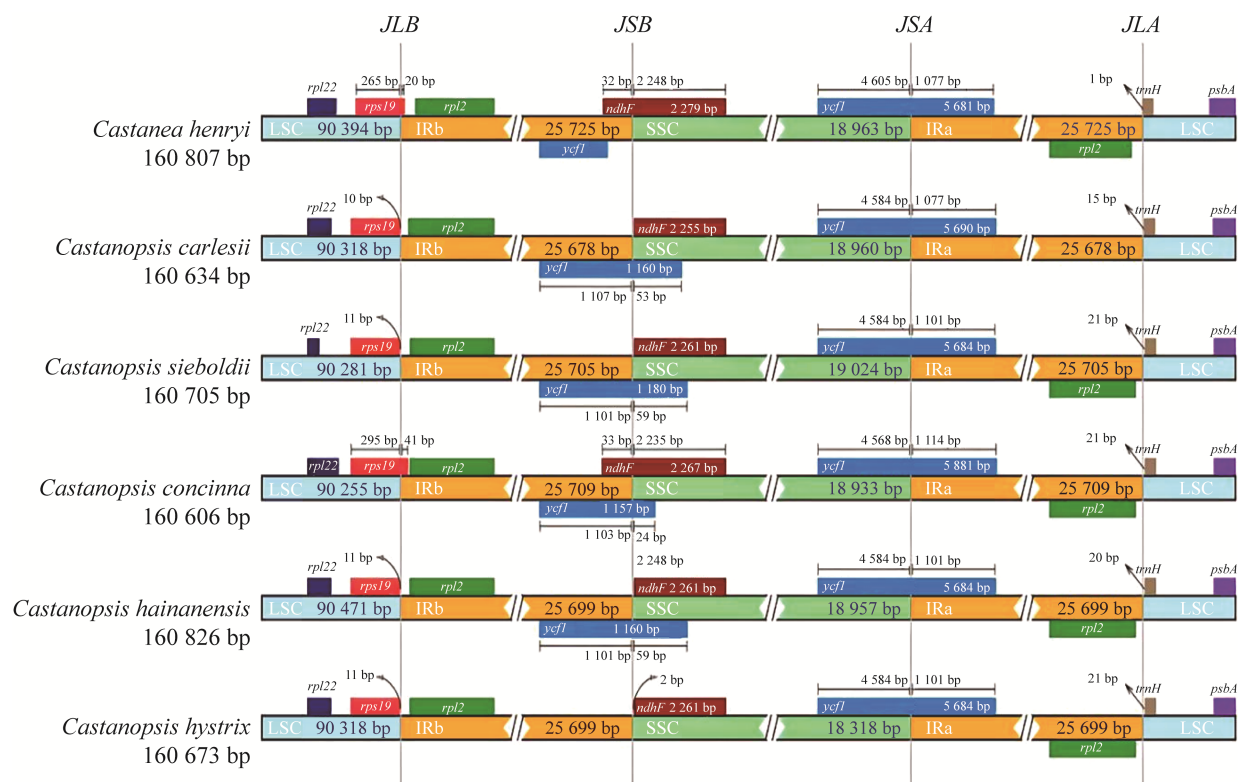


图 4 锥属 6 种植物叶绿体基因组的 IR 边界分析

Figure 4 IR boundary analysis of the chloroplast genomes of six species of *Castanopsis*.

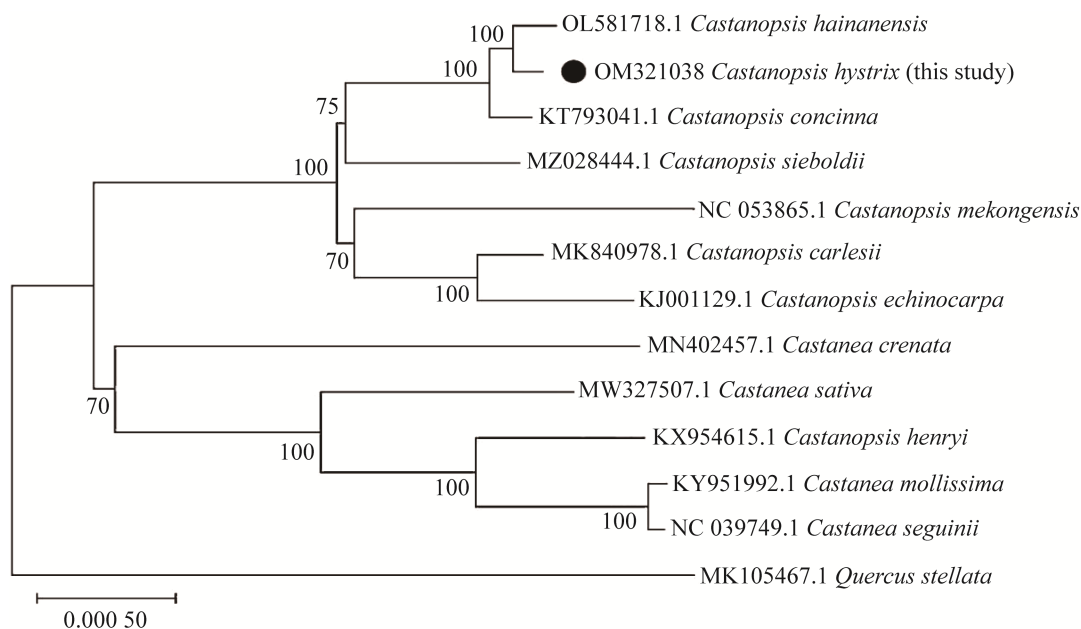


图 5 基于壳斗科 13 个物种叶绿体全基因组构建的系统进化树 支上的数字为 1 000 次自展支持率(%)

Figure 5 Phylogenetic tree based on the whole chloroplast genomes of 13 species of Fagaceae. The 1 000 bootstrap value (%) was shown above the branch.

3 讨论

植物的整个叶绿体基因组包含约 63–209 个基因, 其中大部分集中在 110–130 个基因, 其组成和排列高度保守, 主要包含了光合作用、叶绿体基因组转录和表达的相关基因等^[18]。通过 NCBI 检索叶绿体基因组和比较红锥与已报道的壳斗科锥属物种, 确认锥属叶绿体基因组长度在 160 205–160 826 bp 之间, 本研究组装的红锥叶绿体基因组全序列长度是 160 673 bp, 未发现基因重排或基因缺失现象^[19–21], 表明锥属物种叶绿体基因组结构相对保守。红锥叶绿体基因组的 GC 含量为 36.8%、AT 含量为 65.3%, AT 含量大于 GC 含量, 是因为 GC 含量越高, DNA 密度就越高, 序列容易保守且不易变性^[22]。因此推测, 红锥的叶绿体基因组存在变异性, 这也许就是使红锥叶绿体基因组相对于其他物种更易发生突变的原因。

通过对红锥叶绿体基因组的密码子分析, ENC 值大于 45, 表示红锥叶绿体基因组的密码子偏好性不强。RSCU 值>1 的密码子共有 31 个, 大多数植物在蛋白质合成过程中更喜欢使用同义密码子, 此类现象是不同因素导致的, 而密码子的碱基组成是较为常见的原因^[23–24]。因为密码子第 3 个碱基的变动一般不会导致编码氨基酸的变动, 所以第 3 个碱基的选择相对稳定, 能够成为分析密码子使用偏差的重要参数^[25], 与许多同方向研究结果类似, 红锥各基因密码子的第 3 位碱基 A 和 T 高于 G 和 C 的使用频次, 存在使用偏好性, 此外, T 高于 A, C 高于 G, 这与大花君子兰(*Clivia miniata*)^[26]、西南桦(*Betula alnoides*)^[27]等植物叶绿体基因的第 3 位密码子偏好性一致, 但不同于“玉铃铛”枣(‘Yulingdang’ jujuba)^[28]、蒺藜苜蓿(*Medicago truncatula*)^[29]和银白杨(*Populus alba* Linn.)^[30]等植物的分析报

告结果。这表示在不同物种中其叶绿体基因组的碱基组成具备个体差异性, 密码子使用偏好性存在一定的差异。

被子植物的叶绿体基因组大多是母系遗传的, SSR 位点鉴定是群体遗传进化和研究的有效方法, 被相关研究广泛应用^[31–35]。本研究从红锥叶绿体基因组中鉴定到长序列重复与 SSR 位点分别为 45 对与 111 个, 其中单核苷酸 SSR 占比最高, 红锥重复序列特征与锥栗、米槠、板栗等其他锥属植物类似, 结果表明红锥系统发育信息主要由单核苷酸重复提供。红锥叶绿体基因组中, 53.33% 的长序列重复分布在 LSC, 95.40% 的 SSR 位点分布在单拷贝区(LSC 和 SSC), 与锥属^[36]和木姜子属(*Litsea*)^[37]结果相似。红锥基因组中 AG/CT、AT/AT、AAT/ATT 和 AAAT/ATTT 比例较高, A/T 重复类 SSR 的丰富证明了此类区域具备较高的遗传多样性, 本研究结果与海甘蓝属(*Crambe*)^[38]和吊兰属(*Chlorophytum*)^[39]结果相同。本研究判定的重复序列能够为分子标记提供有效参考, 且用于红锥的物种鉴定和群体遗传学研究。

整个叶绿体基因组的大小变化主要是由 IR 和 SSC 边界区域的扩张和收缩造成的^[40], 这种扩张和收缩在演化过程中起到决定性作用^[41–43]。本研究中红锥的 *rps19* 基因无跨越 LSC/IRb 边界处现象(距离边界处 11 bp), 这显示了红锥的 IR 区对比其他锥属存在显著的收缩^[44], 这种收缩一定程度上影响了红锥叶绿体基因组的长度与其他锥属物种, 但是在 IR 边界的收缩扩张变化会较为显著, 如 *ndhF* 基因在红锥中跨过 IRB/SSC 边界, 向 IRB 扩张了 2 bp, 其他 3 个壳斗科物种则仍在 SSC 区, 而米槠、海南锥、长果锥与边界齐平, 这与已报道的壳斗科 *ndhF* 从距离 IRB/SSC 边界最长 32 bp 至越过边界 1 023 bp 的结果一致^[45]。SSC/IRa 边界常位于

ycf1 编码序列中, 本研究中红锥向 IRa 越过边界 1 101 bp, 而其他壳斗科物种大于等于 1 101 bp, 表明红锥的 *ycf1* 与其他壳斗科的 *ycf1* 差异较大。综上所述, 本研究中 IR 边界分析结果均与前人已有研究趋势相似, 但基因在边界处的扩张与收缩程度仍存在差异性, 证明壳斗科叶绿体基因组中基因与其排序的保守性和 IR 边界收缩扩张的多样性并存。

将相关物种的叶绿体基因组与红锥叶绿体基因组使用最大似然法测序进行聚类分析, 聚类结果表明红锥与海南锥聚为一支, 与前人研究相似^[46]。该研究结果为壳斗科各属植物的起源及亲缘关系的进一步确定提供了相关证据, 有助于后续的壳斗科系统发育研究和物种鉴定工作。

4 结论

通过 Illumina HiSeq 2500 测序平台对红锥的完整叶绿体基因组进行组装, 红锥叶绿体基因组为典型的四分体结构, 全长为 160 673 bp, 全部注释到 130 个基因(85 个蛋白编码基因、37 个 tRNA 基因和 8 个 rRNA 基因); 密码子偏好性分析表明红锥偏爱使用 A/U 结尾的密码子; 共检测到 111 个 SSR 位点, 以单碱基重复 A/T 为主, 获得的 SSR 位点可用于物种鉴定和群体遗传学研究; 对红锥叶绿体基因组进行生物信息学等系统化分析, 能够充分了解锥属种间的系统发育关系, 有利于对锥属植物资源的开发利用, 为以后红锥亲缘关系解析、遗传多样性分析、物种鉴定、基因工程等的相关研究提供了分子理论依据。

REFERENCES

- [1] 国家林业和草原局. 中国主要栽培珍贵树种参考名录(2017 年版)[EB/OL]. (2017-11-09) [2021-01-02]. <http://www.forestry.gov.cn/main/3954/content-1045203.html>.
- [2] 郭文福, 蔡道雄. 红锥大径材近自然培育适用性分析[J]. 湖南林业科技, 2015, 42(1): 79-82.
GUO WF, CAI DX. The applicability analysis on growing for large diameter valuable wood of *Castanopsis hystrix* by close-to-nature forest management[J]. Hunan Forestry Science & Technology, 2015, 42(1): 79-82 (in Chinese).
- [3] 廖宇杰, 洪维, 陈富强, 张干荣, 李仕裕, 李志宏, 庄晓纯, 邢福武, 王发国. 广州红锥-马占相思林物种组成与多样性研究[J]. 热带亚热带植物学报, 2021, 29(5): 494-502.
LIAO YJ, HONG W, CHEN FQ, ZHANG GR, LI SY, LI ZH, ZHUANG XC, XING FW, WANG FG. Studies on species composition and diversity of *Castanopsis hystrix*-*Acacia mangium* mixed forest in Guangzhou[J]. Journal of Tropical and Subtropical Botany, 2021, 29(5): 494-502 (in Chinese).
- [4] 魏国余, 亢亚超, 廖曦, 王家妍, 申礼凤, 方小荣, 王凌晖, 滕维超. 配方施肥对红锥幼林生理及叶片养分含量的影响[J]. 西北林学院学报, 2020, 35(4): 32-36, 83.
WEI GY, KANG YC, LIAO X, WANG JY, SHEN LF, FANG XR, WANG LH, TENG WC. Effects of formula fertilization on the physiology characteristics and leaf nutrition contents of young *Castanopsis hystrix* stands[J]. Journal of Northwest Forestry University, 2020, 35(4): 32-36, 83 (in Chinese).
- [5] 余涛, 蒲芬, 管芹, 范敏. 南欧大戟叶绿体基因组特征及其系统发育分析[J]. 分子植物育种, 2022, 20(6): 1828-1837.
YU T, PU F, GUAN Q, FAN M. Chloroplast genome of *Euphorbia peplus* L. characteristic and phylogenetic analysis[J]. Molecular Plant Breeding, 2022, 20(6): 1828-1837 (in Chinese).
- [6] 李倩, 郭其强, 高超, 李慧娥. 贵州威宁红花油茶叶绿体基因组特征分析[J]. 园艺学报, 2020, 47(4): 779-787.
LI Q, GUO QQ, GAO C, LI HE. Characterization of complete chloroplast genome of *Camellia weiningensis* in Weining, Guizhou Province[J]. Acta Horticulturae Sinica, 2020, 47(4): 779-787 (in Chinese).

- [7] DANIELL H, LIN CS, YU M, CHANG WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering[J]. *Genome Biology*, 2016, 17(1): 134.
- [8] ZHAO Z, WANG X, YU Y, YUAN S, JIANG D, ZHANG Y, ZHANG T, ZHONG W, YUAN Q, HUANG L. Complete chloroplast genome sequences of *Dioscorea*: characterization, genomic resources, and phylogenetic analyses[J]. *PeerJ*, 2018, 6: e6032.
- [9] 姜汶君, 郭梦月, 庞晓慧. 叶绿体基因组在药用植物鉴定及系统进化研究中的应用[J]. *世界中医药*, 2020, 15(5): 702-708, 716.
- JIANG WJ, GUO MY, PANG XH. Application of chloroplast genome in identification and phylogenetic analysis of medicinal plants[J]. *World Chinese Medicine*, 2020, 15(5): 702-708, 716 (in Chinese).
- [10] 马孟莉, 孟衡玲, 张薇, 雷恩, 卢丙越. 阳春砂叶绿体全基因组解析及系统发育研究[J]. *西北植物学报*, 2020, 40(6): 978-986.
- MA ML, MENG HL, ZHANG W, LEI E, LU BY. Complete chloroplast genome and phylogenetic analysis of *Amomum villosum*[J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2020, 40(6): 978-986 (in Chinese).
- [11] 吴娟子, 钱晨, 刘智微, 潘玉梅, 钟小仙. 基于转录组测序分析象草木质素合成的研究[J]. *草业学报*, 2019, 28(1): 150-161.
- WU JZ, QIAN C, LIU ZW, PAN YM, ZHONG XX. *De novo* transcriptomic analysis for lignin synthesis in *Cenchrus purpureus* using RNA-Seq[J]. *Acta Prataculturae Sinica*, 2019, 28(1): 150-161 (in Chinese).
- [12] 金刚, 覃旭, 龙凌云, 王丽萍, 覃剑峰, 危丹妮, 陈涛, 蔡中全. 剑麻叶绿体基因组编码序列密码子的使用特征[J]. *福建农林大学学报(自然科学版)*, 2018, 47(6): 705-710.
- JIN G, QIN X, LONG LY, WANG LP, QIN JF, WEI DN, CHEN T, CAI ZQ. Characteristics of codon usage in the chloroplast protein-coding genes of *Agave* hybrid No. 11648[J]. *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, 2018, 47(6): 705-710 (in Chinese).
- [13] KURTZ S, CHOUDHURI JV, OHLEBUSCH E, SCHLEIERMACHER C, STOYE J, GIEGERICH R. REPuter: the manifold applications of repeat analysis on a genomic scale[J]. *Nucleic Acids Research*, 2001, 29(22): 4633-4642.
- [14] 许阳. 不同叶习性栎属物种间的叶绿体基因组比较研究[D]. 北京: 北京林业大学硕士学位论文, 2021.
- XU Y. Comparison of chloroplast genomes of different leaf habit oak species[D]. Beijing: Master's Thesis of Beijing Forestry University, 2021 (in Chinese).
- [15] 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.
- [16] NAKAMURA T, YAMADA KD, TOMII K, KATOH K. Parallelization of MAFFT for large-scale multiple sequence alignments[J]. *Bioinformatics*, 2018, 34(14): 2490-2492.
- [17] 李佳文, 刘强, 张静文, 杨勇, 张世杰, 张颖. 濒危红树植物莲叶桐叶绿体基因组及其系统进化[J]. *西北林学院学报*, 2020, 35(5): 54-61.
- LI JW, LIU Q, ZHANG JW, YANG Y, ZHANG SJ, ZHANG Y. Chloroplast genome of endangered mangrove plants *Hernandia nymphiifolia* and its phylogenetic evolution[J]. *Journal of Northwest Forestry University*, 2020, 35(5): 54-61 (in Chinese).
- [18] LI YF, SYLVESTER SP, LI M, ZHANG C, LI X, DUAN YF, WANG XR. The complete plastid genome of *Magnolia zenii* and genetic comparison to Magnoliaceae species[J]. *Molecules*, 2019, 24(2): 261.
- [19] YANG JB, YANG SX, LI HT, YANG J, LI DZ. Comparative chloroplast genomes of camellia species[J]. *PLoS One*, 2013, 8(8): e73053.
- [20] DENG Z, HUANG L, ZHANG M, TARIN MWK, HE T, CHEN L, RONG J, CHEN L, ZHENG Y. The complete chloroplast genome sequence of *Bambusa vulgaris* cv. Wamin[J]. *Mitochondrial DNA Part B*, 2021, 6(6): 1704-1705.
- [21] CHEN LG, ZHANG YY, LIU S, HU WF, HAN YZ, UL HAQ A, ZHENG YS. The complete chloroplast genome sequence of *Bambusa lapidea* (Bambusodae)[J]. *Mitochondrial DNA Part B*, 2021, 6(7): 2046-2047.
- [22] 李连星, 彭劲谕, 王大玮, 段安安. 长爪栎[木衣]叶绿体基因组特征系统发育及密码子偏好性分析[J]. *生物工程学报*, 2022, 38(1): 328-342.
- LI LX, PENG JY, WANG DW, DUAN AA. Chloroplast genome phylogeny and codon preference of *Docynia longiunguis*[J]. *Chinese Journal of Biotechnology*, 2022, 38(1): 328-342 (in Chinese).
- [23] ROMERO H, ZAVALA A, MUSTO H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces[J]. *Nucleic Acids Research*, 2000, 28(10): 3030-3031.

- 2084-2090.
- [24] CARLINI DB, CHEN Y, STEPHAN W. The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes *adh* and *adhr*[J]. *Genetics*, 2001, 159(2): 623-633.
- [25] 丁锐, 胡兵, 宗小雁, 韩辰阳, 张丽杰, 陈旭辉. 杓兰叶绿体基因组密码子偏好性分析[J]. *林业科学研究*, 2021(5): 177-185.
- DING R, HU B, ZONG XY, HAN CY, ZHANG LJ, CHEN XH. Analysis of codon usage in the chloroplast genome of *Cypripedium calceolus*[J]. *Forest Research*, 2021(5): 177-185 (in Chinese).
- [26] 郑伟, 张卉, 王钦美, 高悦, 张志宏, 孙玉新. 大花君子兰叶绿体基因组及其特征[J]. *园艺学报*, 2020, 47(12): 2439-2450.
- ZHENG Y, ZHANG H, WANG QM, GAO Y, ZHANG ZH, SUN YX. Complete chloroplast genome sequence of *Clivia miniata* and its characteristics[J]. *Acta Horticulturae Sinica*, 2020, 47(12): 2439-2450 (in Chinese).
- [27] 原晓龙, 李云琴, 王毅, 张劲峰. 西南桦叶绿体基因组密码子偏好性分析[J]. *基因组学与应用生物学*, 2020, 39(12): 5758-5764.
- YUAN XL, LI YQ, WANG Y, ZHANG JF. Analysis of codon usage in the chloroplast genome of *Betula alnoides*[J]. *Genomics and Applied Biology*, 2020, 39(12): 5758-5764 (in Chinese).
- [28] 郭晓雨, 张晓慧, 马宗新, 李文峰, 李慧慧. ‘玉铃铛’枣叶绿体基因组特征及密码子偏性[J]. *分子植物育种*, 2021, 19(37): 1-15.
- GUO XY, ZHANG XH, MA ZX, LI WF, LI HH. Analysis of chloroplast genome characteristics and codon usage bias of ‘yulingdang’ jujuba[J/OL]. *Molecular Plant Breeding*, 2021: 1-15 (in Chinese).
- [29] 杨国锋, 苏昆龙, 赵怡然, 宋智斌, 孙娟. 蒺藜苜蓿叶绿体密码子偏好性分析[J]. *草业学报*, 2015, 24(12): 171-179.
- YANG GF, SU KL, ZHAO YR, SONG ZB, SUN J. Analysis of codon usage in the chloroplast genome of *Medicago truncatula*[J]. *Acta Prataculturae Sinica*, 2015, 24(12): 171-179 (in Chinese).
- [30] 周猛, 龙伟, 李霞. 银白杨叶绿体基因组密码子用法分析(英文)[J]. *Journal of Forestry Research*, 2008(4): 293-297, 350.
- ZHOU M, LONG W, LI X. Analysis of synonymous codon usage in chloroplast genome of *Populus alba*[J]. *Journal of Forestry Research*, 2008(4): 293-297, 350.
- [31] 张扬勇, 方智远, 王庆彪, 刘玉梅, 杨丽梅, 庄木, 孙培田. 拟南芥叶绿体 SSR 引物在甘蓝上的应用[J]. *园艺学报*, 2011, 38(3): 549-555.
- ZHANG YY, FANG ZY, WANG QB, LIU YM, YANG LM, ZHUANG M, SUN PT. Utility of *Arabidopsis* chloroplast simple sequence repeat (cpSSR) primers in cabbage (*Brassica oleracea* L. var. *capitata* L.)[J]. *Acta Horticulturae Sinica*, 2011, 38(3): 549-555 (in Chinese).
- [32] TAKAHASHI D, SAKAGUCHI S, ISAGI Y, SETOGUCHI H. Comparative chloroplast genomics of series *Sakawanum* in genus *Asarum* (Aristolochiaceae) to develop single nucleotide polymorphisms (SNPs) and simple sequence repeat (SSR) markers[J]. *Journal of Forest Research*, 2018, 23(6): 387-392.
- [33] 魏潇, 章秋平, 刘威生, 刘宁, 张玉萍, 徐铭, 刘硕, 张玉君, 马小雪. 基于叶绿体 SSR 单倍型分析普通杏演化关系[J]. *植物遗传资源学报*, 2018, 19(4): 705-712.
- WEI X, ZHANG QP, LIU WS, LIU N, ZHANG YP, XU M, LIU S, ZHANG YJ, MA XX. Phylogenetic relationship analysis of common apricot (*Prunus armeniaca* L.) revealed by chloroplast SSR haplotypes[J]. *Journal of Plant Genetic Resources*, 2018, 19(4): 705-712 (in Chinese).
- [34] YAMANE K, KAWAHARA T. Size homoplasy and mutational behavior of chloroplast simple sequence repeats (cpSSRs) inferred from intra- and interspecific variations in four chloroplast regions of diploid and polyploid *Triticum* and *Aegilops* species[J]. *Genetic Resources and Crop Evolution*, 2018, 65(3): 727-743.
- [35] 韦壮敏, 李蒙蒙, 杨晓明, 袁晓慧, 叶俊丽, 谢宗周, 邓秀新, 柴利军. 无核酸橙资源‘抛橘’的遗传鉴定[J]. *园艺学报*, 2020, 47(2): 345-354.
- WEI ZM, LI MM, YANG XM, YUAN XH, YE JL, XIE ZZ, DENG XX, CHAI LJ. Genetic background analysis of seedless sour orange ‘paoju’[J]. *Acta Horticulturae Sinica*, 2020, 47(2): 345-354 (in Chinese).
- [36] YE X, HU D, GUO Y, SUN R. Complete chloroplast genome of *Castanopsis sclerophylla* (Lindl.) Schott: genome structure and comparative and phylogenetic analysis[J]. *PLoS One*, 2019, 14(7): e0212325.
- [37] ZHANG YY, TIAN YJ, TNG DYP, ZHOU JB, ZHANG YT, WANG ZW, LI PF, WANG ZS. Comparative chloroplast genomics of *Litsea* lam. (Lauraceae) and its phylogenetic implications[J]. *Forests*, 2021, 12(6): 744.
- [38] 钱方, 高作敏, 胡利娟, 王洪程. 海甘蓝(*Crambe*

- abyssinica*)叶绿体基因组特征及其系统发育研究[J]. 生物技术通报, 2022, 38(6): 174-186.
- QIAN F, GAO ZM, HU LJ, WANG HC. Characteristics of *Crambe abyssinica* chloroplast genome and its phylogenetic relationship in Brassicaceae[J]. Biotechnology Bulletin, 2022, 38(6): 174-186 (in Chinese).
- [39] MUNYAO JN, DONG X, YANG JX, MBANDI EM, WANGA VO, OULO MA, SAINA JK, MUSILI PM, HU GW. Complete chloroplast genomes of *Chlorophytum comosum* and *Chlorophytum gallabatense*: genome structures, comparative and phylogenetic analysis[J]. Plants: Basel, Switzerland, 2020, 9(3): E296.
- [40] DUGAS DV, HERNANDEZ D, KOENEN EJM, SCHWARZ E, STRAUB S, HUGHES CE, JANSEN RK, NAGESWARA-RAO M, STAATS M, TRUJILLO JT, HAJRAH NH, ALHARBI NS, AL-MALKI AL, SABIR JSM, BAILEY CD. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions and accelerated rate of evolution in *clpP*[J]. Scientific Reports, 2015, 5(1): 16958.
- [41] RAUBESON LA, PEERY R, CHUMLEY TW, DZIUBEK C, FOURCADE HM, BOORE JL, JANSEN RK. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*[J]. BMC Genomics, 2007, 8: 174.
- [42] PARK S, AN B, PARK S. Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion, and intraspecific variation[J]. Scientific Reports, 2018, 8: 13568.
- [43] LI X, LI Y, ZANG M, LI M, FANG Y. Complete chloroplast genome sequence and phylogenetic analysis of *Quercus acutissima*[J]. International Journal of Molecular Sciences, 2018, 19(8): E2443.
- [44] CHENG H, LI J, ZHANG H, CAI B, GAO Z, QIAO Y, MI L. The complete chloroplast genome sequence of strawberry (*Fragaria×ananassa* Duch.) and comparison with related species of Rosaceae[J]. PeerJ, 2017, 5: e3919.
- [45] 张妍彤, 黄剑, 宋菊, 林丽梅, 冯若宣, 邢朝斌. 壳斗科植物叶绿体基因组结构及变异分析[J]. 植物研究, 2018, 38(5): 757-765.
- ZHANG YT, HUANG J, SONG J, LIN LM, FENG RX, XING ZB. Structure and variation analysis of chloroplast genomes in fagaceae[J]. Bulletin of Botanical Research, 2018, 38(5): 757-765 (in Chinese).
- [46] 程丽莉, 胡广隆, 苏淑钗, 黄武刚. 板栗及其近缘种叶绿体 SSR 遗传多样性分析[J]. 华北农学报, 2015, 30(2): 145-149.
- CHENG LL, HU GL, SU SC, HUANG WG. Diversity of cultivated chestnuts and its close relative species using chloroplast SSRs[J]. Acta Agriculturae Boreali-Sinica, 2015, 30(2): 145-149 (in Chinese).

(本文责编 陈宏宇)