

野生和栽培酸枣叶绿体基因组测序与特征分析

杨丁卯¹, 李 婷^{1*}, 宋忠兴¹, 焦红红^{1,2}, 王 梅^{1,3*}

1. 陕西中医药大学, 陕西中药资源产业化省部共建协同创新中心, 陕西 咸阳 712000

2. 道地药材品质保障与资源持续利用全国重点实验室, 北京 100700

3. 中国中医科学院, 中国中医科学院望京医院, 北京 100091

摘要: **目的** 明确野生和栽培酸枣 *Ziziphus jujuba* var. *spinosa* 叶绿体基因组特征与系统发育关系, 为酸枣的分子鉴定提供依据。 **方法** 以 2 种不同品种的酸枣为研究对象, 采用高通量测序技术, 获得酸枣叶绿体基因组序列, 并对其结构区及功能区进行注释, 对密码子偏好性、重复序列、IR 区收缩与扩张、核酸多态性以及分子进化等方面进行研究。 **结果** 野生和栽培酸枣叶绿体基因组均具有典型的四分体结构, 全长序列分别为 161 343、161 601 bp, 总 GC 含量均为 37.0%, 均编码 129 个基因; 野生和栽培酸枣分别检测到 73、71 个散在重复序列, 均为 F 型和 P 型, 简单重复序列 SSR 位点分别为 89、91 个; 野生和栽培酸枣叶绿体基因组密码子偏好性较弱, 偏向于以 A/T 碱基结尾, 野生酸枣较栽培酸枣最优密码子多 1 个, 核苷酸多态性指数 (nucleotide polymorphism index, P_i) 显示野生酸枣位于大单拷贝区 (large single-copy region, LSC) 的 *rpoC2* 基因和 *rpoB* 基因具有高变异性, 栽培酸枣位于 LSC 区的 *atPA* 基因有高变异性; IR 区域收缩与扩张显示 JLB (LSC/IRb) 区域、JSB (IRb/SSC) 区域和 JSA (SSC/Ira) 区域存在不同程度的基因变异及序列收缩与扩张现象; 系统发育树结果表明野生酸枣与参考序列 KU351660.1 聚为一支、栽培酸枣则与参考序列 KX266829.1 (骏枣 *Ziziphus jujube* ‘Junzao’)、KU351660.1 (枣 *Ziziphus jujuba*)、KX266830.1、MF781071.1 (冬枣 *Ziziphus jujuba* ‘Dongzao’) 亲缘关系比较密切。 **结论** 野生和栽培酸枣叶绿体基因组在简单重复序列、IR 区收缩与扩张、最优密码子数目、核酸多态性片段及系统发育树几个方面存在差异, 后续可以利用存在差异的 SSR 位点、高变异片段及 IR 区域基因类型对野生和栽培酸枣有效鉴别, 为野生和栽培酸枣的分子标记提供理论依据。

关键词: 叶绿体基因组; 酸枣; 野生; 栽培; 序列结构; 特征分析

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2025)22-8334-13

DOI: 10.7501/j.issn.0253-2670.2025.22.023

Sequencing and characterization of chloroplast genome of wild and cultivated *Ziziphus jujuba* var. *spinosa*

YANG Dingmao¹, LI Ting¹, SONG Zhongxing¹, JIAO Honghong^{1,2}, WANG Mei^{1,3}

1. Shaanxi Provincial Cooperative Innovation Center for the Industrialization of Traditional Chinese Medicine Resources, Shaanxi University of Chinese Medicine, Xianyang 712000, China

2. State Key Laboratory of Quality Ensurance and Sustainable Use of Dao-di Herbs, Beijing 100700, China

3. Wangjing Hospital, China Academy of Chinese Medical Sciences, Beijing 100091, China

Abstract: Objective To clarify the relationship between chloroplast genome characteristics and phylogeny of wild and cultivated *Ziziphus jujuba* var. *spinosa* (ZS) and provide the basis for molecular identification of ZS. **Methods** In this study, the whole-genome sequence of two different varieties of ZS was obtained using high-throughput sequencing technology. Its structural and functional regions were annotated, and the coding preference, repeat sequence, contraction and expansion of the IR region, nucleic acid polymorphism and molecular evolution were studied. **Results** The chloroplast genomes of both wild

收稿日期: 2025-03-03

基金项目: 陕西省重点研发计划 (2024CY-JJQ-41); 咸阳市重点研发计划 (L2024-ZDYF-ZDYF-SF-0035); 现代农业产业技术体系建设专项 (CARS-21)

作者简介: 杨丁卯, 硕士研究生, 研究方向为中药鉴定学。E-mail: 1913470599@qq.com

*通信作者: 李 婷, 讲师, 从事中药鉴定方向研究。E-mail: Abby1230@163.com

王 梅, 硕士生导师, 从事中药药效物质基础及质量控制研究。E-mail: 2285181192@qq.com

and cultivated ZS had typical tetrad structure, with full-length sequences of 161 343 and 161 601 bp, total GC content of 37.0%, encoding 129 genes. There were 73 and 71 scattered repeats in wild and cultivated ZS, which were F type and P type, respectively, and 89 and 91 simple repeat SSR loci were detected, respectively. The codon preference of the chloroplast genome of wild and cultivated ZS was weak and tended to end in A/T base. Wild ZS had one more optimal codon than cultivated ZS. The nucleotide polymorphism index (P_i) value showed that there were two fragments with high variability in the wild ZS, *rpoC2* gene and *rpoB* gene in the large single-copy region (LSC) region, and one fragment with high variability in cultivated ZS, *atPA* gene in the LSC region. The contraction and expansion of IR region showed that JLB (LSC/IRb) region, JSB (IRb/SSC) region and JSA (SSC/Ira) region had different degrees of gene variation and sequence contraction and expansion; Phylogenetic trees showed that wild ZS was closely related to the reference sequence KU351660.1, and cultivated ZS was closely related to the reference sequences KX266829.1 (*Ziziphus jujuba* 'Junzao'), KU351660.1 (*Ziziphus jujuba*), KX266830.1, MF781071.1 (*Ziziphus jujuba* 'Dongzao'). **Conclusion** The chloroplast genomes of wild and cultivated ZS are different in simple repeats, IR contraction and expansion, the number of optimal codons, nucleic acid polymorphic fragments and phylogenetic trees, which can be used to distinguish wild and cultivated ZS effectively by using different SSR loci, highly variable fragments and IR genotype and provide a theoretical basis for molecular markers of wild and cultivated ZS.

Key words: chloroplast genome; *Ziziphus jujuba* var. *spinosa* (Bunge) Hu ex H. F. Chow.; wild; cultivated; sequence structure; characteristic analysis

叶绿体是一种重要的光合器官,在高等植物中普遍存在着一种独特的遗传材料,被称为叶绿体基因组(chloroplast genome)^[1]。叶绿体基因组的特征是:长度比细胞核小,提取纯化容易,测定费用低廉;其基因组在结构上较为保守;基因组中包含了丰富的遗传信息,包括基因组的数目和结构特点,可以用来进行系统进化分析;母系单性遗传,除反向复制外,其他基因都是单一复制,以确保种间基因的直接同源性^[2]。因此,利用叶绿体基因组所包含的基因信息,对其进行系统发育、鉴别和分类,是当前研究的热点^[3]。通过对叶绿体基因组在植物中的位置进行分析,可以使对植物的进化关系有一个更加全面的认识,并为探讨叶绿体基因组与其他叶绿体基因、核基因的互作关系提供一个新的思路。

酸枣 *Ziziphus jujuba* var. *spinosa* (Bunge) Hu ex H. F. Chow.为鼠李科枣属植物,主要分布于辽宁、内蒙古、河北、山东、山西、河南、陕西、甘肃、宁夏、新疆、江苏、安徽等地^[4]。酸枣的干燥成熟种子酸枣仁是中医治疗失眠的首选用药,在 2002 年被国家卫健委列为药食同源品种。经研究,酸枣富含黄酮、酚类及三萜类化合物^[5],具有明显的镇静催眠活性、抗衰老、抗氧化及增强免疫力等作用^[6-9]。

酸枣作为枣的野生变种,其存在方式主要是野生状态^[10]。由于近年来酸枣的需求量激增,价格不断上涨,野生酸枣资源被肆意掠夺,栽培酸枣成为市场上的主流品种。但是,由于枣树的种植方法多是参照农作物的种植管理方式,存在种植密度大,

施肥量大,农药用量大,甚至使用了大量的植物生长调节剂的现象,导致栽培酸枣在外观上与野生酸枣发生了差异,而表型变异的同时也提醒了内在质量及遗传方面发生变化。野生植物种群是珍贵的植物资源,蕴含着丰富的优良基因,但由于长期的引种驯化,其种群遗传结构相对保守,造成其环境适应性差,病虫害抗性差,产量难以提高等问题,因此对野生与栽培植物的遗传多样性差异探索有利于开发和保护植物群体的可持续发展。目前有关野生和栽培酸枣叶绿体基因组特征分析的研究鲜见报道,本研究通过对野生和栽培酸枣的叶绿体基因组测序、叶绿体基因组基本特征、密码子偏好性、简单重复序列分析、红外(IR)区收缩与扩张分析、核酸多态性分析及系统发育分析,以期为其物种的鉴定和遗传多样性研究提供基础资料。

1 材料

野生酸枣采集于 2023 年 9 月份河北省邢台市邢台县吕家洞村,栽培酸枣采集于 2023 年 10 月份陕西省渭南市白水县,待人工砸出酸枣种子后,挑取完整饱满的种子种植于花盆中,播种条件为每盆装土量为 4 kg,土壤为壤土、砂质土、腐殖质(1:2:1)的混合土,每桶种植 6 株,待种植的酸枣植株长出两对真叶后,由陕西中药资源产业化省部共建协同创新中心宋忠兴主任药师鉴定为鼠李科枣属植物酸枣 *Z. jujuba* var. *spinosa* (Bunge) Hu ex H. F. Chow.的叶。分别摘取不同盆中长势良好的酸枣嫩叶装于不同的取样管中,每管样品约 5 g,重复 3

次, 取样后经液氮速冻保存于 -80°C 备用。

2 方法

2.1 DNA 提取与测序

分别取野生和栽培酸枣叶片适量, 采用植物基因组 DNA 提取试剂盒(天根生化科技有限公司, 北京)对酸枣叶片的 gDNA 进行提取, 并利用 Qubit dsDNA HS 检测试剂盒(Sangon 公司)和 1%琼脂糖凝胶电泳确认 DNA 浓度以及完整性。检测合格的 DNA 由生工生物工程(上海)有限公司进行 DNA 文库构建和下一代序列。首先, 由 Covaris(美国 Woburn)随机片段化 500 ng 定量 DNA。接下来, 使用 Hieff NGS®MaxUp II DNA Library Prep Kit for Illumina®(YEASEN 公司)进行下一步实验操作: 添加 Endprep 酶以修复末端和 3'末端 A 尾连接, 用增强子和 Fast T4 DNA 连接酶连接接头蛋白, 通过 PCR 添加索引引物, 使用 DNA 筛选珠筛选约 400 bp 的扩增产物。通过 Qubit 4.0(Thermo 公司, 美国)和 2%琼脂糖凝胶电泳验证测序文库质量。根据制造商的说明, 将文库合并并用 2×150 bp 配对末端序列试剂盒加载到 Novaseq 6000(Illumina 公司, 美国)/DNBseq-T7(BGI 公司)测序仪上进行测序。为保证分析信息质量, 通过 Fastp 0.36^[11]软件对测序的原始数据(Raw Reads)去除原始序列中的杂质数据, 如接头、低质量数据和 N 等, 获得高质量序列数据(Clean Reads)。

2.2 酸枣叶绿体基因组序列拼接与基因元件注释

利用 SPAdes 3.15^[12]软件将筛选出的 Clean Reads 组装为 contig 序列, 拼接完成后对序列进行评估和质控, 使用 blastn 将 scaffolds 与 NCBI nt 库进行比对以获取序列相似度信息, 提取各 scaffolds 的测序深度和覆盖度信息, 在对上述信息整理并综合考量之后, 人工选取可能的目标 scaffolds。在此基础上, 采用 GapFiller1.11^[13]对拼接后的 contig 进行修复, 并利用 PrInSeS-G 对其进行顺序校正, 以纠正拼接时出现的剪切误差和小片段的丢失。最后, 利用内建的 script 序列, 对拼接结果进行环合评价和加工, 从而获得完整的酸枣叶绿体基因组。

使用 CpGAVAS(<http://47.96.249.172:16019/alyzer/home>)注释酸枣叶绿体基因组序列, 并采用反向 blast 与参考基因组或内建数据库进行比对, 优化并提取拼接结果, 注释其 CDS、tRNA、rRNA 等基因元件。最后使用 tbl 文件(five-column tab-delimited table of feature locations and qualifiers)和

fsa 文件 DNA sequence (in FASTA format))上传至 NCBI 网站(GenBank: PQ841431、PQ841432)。

2.3 酸枣叶绿体基因组特征分析

2.3.1 基因组基本特征分析 利用在线软件 OGDRAW^[14]进行叶绿体基因组物理图谱绘制。上传已获得的 GenBank 格式注释文件, 选择输出文件状态, 点击提交后获得酸枣的叶绿体基因组物理图谱。利用 Chloroplot 在线网站(<https://irscope.shinyapps.io/Chloroplot/>)^[15]计算各区域长度信息及所注释基因数量信息。利用软件 BioEdit^[16]的 Nucleotide Composition 功能计算叶绿体基因组 GC 含量。

2.3.2 密码子偏好性分析 同义密码子使用度(relative synonymous codon usage, RSCU)是指每一个同义密码子所对应的氨基酸所对应的频率。由于不同种属之间对密码子的偏好存在差异, 因此我们可以通过对密码子的偏好来进行判别亲缘关系的远近。当 $\text{RSCU} > 1$ 时, 表示该密码子使用频率较高, 偏好性较强; 当 $\text{RSCU} < 1$ 时, 表示该密码子使用频率较低, 偏好性较弱; 当 $\text{RSCU} = 1$ 时, 表明该密码子无偏好性^[17]。本研究拟采用 Codon W1.4.2 软件, 对酸枣叶绿体基因组中蛋白质编码基因的 RSCU 进行比较, 并以氨基酸密码子数作为纵坐标, 用 ggplot 软件对其进行可视化绘制。为确定野生和栽培酸枣最优密码子, 需对其高频率和高表达密码子进行分析, 首先选取 RSCU 值高于 1 的密码子为高频密码子^[18], 以筛选出的 85 条 CDS 序列的 ENC 值为参考, 将其从小到大进行排序, 然后选取两端 10%的基因, 建立高低基因表达库(ENC 值越低, 表示密码子偏性越高), 取 $\Delta\text{RSCU} \geq 0.08$ 的密码子作为高表达密码子^[19], 将同时满足高频($\text{RSCU} > 1$)和高表达($\Delta\text{RSCU} \geq 0.08$)的密码子定义为酸枣叶绿体基因组的最优密码子。

2.3.3 散在重复序列及微卫星重复序列分析 使用 REPuter 注释细胞器基因组的重复序列, 包括 4 种类型, 正向重复(F)、倒置重复(R)、互补重复(C)、回文重复(P); 使用在线程序 misa(Misa-web-IPK Gatersleben)对酸枣叶绿体基因组进行 SSR 分子标记。设定各参数: 单碱基重复次数 10 次及以上, 双碱基重复次数 5 次及以上, 3 碱基重复次数 4 次及以上, 4~6 碱基重复次数 3 次及以上。

2.3.4 IR 边界的收缩与扩张 IR 区域边界的收缩和扩张被认为是反复发生的过程, 其区域的收缩与

扩张是导致叶绿体基因组变化的主要原因，为评估酸枣叶绿体基因组各分区边界的差异，本研究采用在线软件 IRscope (<https://irscope.shinyapps.io/irapp/>) 对野生和栽培酸枣及多个参考基因组的 IR 区域以及附近的基因进行比较分析。

2.3.5 核酸多态性分析 单核苷酸多态性 (single nucleotide polymorphism, SNP) 是一种发生在全基因组范围内由单核苷酸突变所导致的 DNA 序列多态性^[20]。在此基础上，利用 DnaSP^[21] v6.12.03 对野生和栽培酸枣叶绿体基因组进行测序，对多个基因组间的核苷酸多态性进行分析。

2.3.6 系统发育树分析 从 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中下载 30 个酸枣的近源物种基因组，使用 Muscle^[22] 进行多重序列比对，并利用 RAxML 8.1.5 (<https://sco.h-its.org/exelixis/web/software/raxml/index.html>)^[23] 使用最大似然法

(maximum likelihood, ML) 得到合适的进化树后导入到 ggtree (<https://github.com/YuLab-SMU/ggtree>) 中进行画图。

3 结果与分析

3.1 酸枣叶绿体基因组测序质量

对野生酸枣进行叶绿体基因组测序，获得 58 169 588 条 raw reads，筛选出 58 131 860 条 clean reads，总有效率 99.94%，GC 占 36.48%；对栽培酸枣叶绿体基因组进行测序，获得 53 362 066 条读数，筛选出 53 313 404 条 clean reads，总有效率 99.91%，GC 占 35.29%。测序深度为 203.3 X，大于 100 X 的平均测序深度，表明该测序结果较好。

3.2 酸枣叶绿体基因组结构特征

本研究通过对酸枣种质材料进行叶绿体基因组拼接注释，基因组各部分长度信息比对结果统计见表 1。酸枣叶绿体基因组谱图见图 1。

表 1 酸枣叶绿体基因组比较

Table 1 Comparison of chloroplast genome of *Z. jujuba* var. *spinosa*

类型	全长/bp	LSC长度/bp	SSC长度/bp	IR长度/bp	GC/%
野生 (HL)	161 343	89 009	19 358	52 976	37
栽培 (SL)	161 601	89 287	19 356	52 958	37

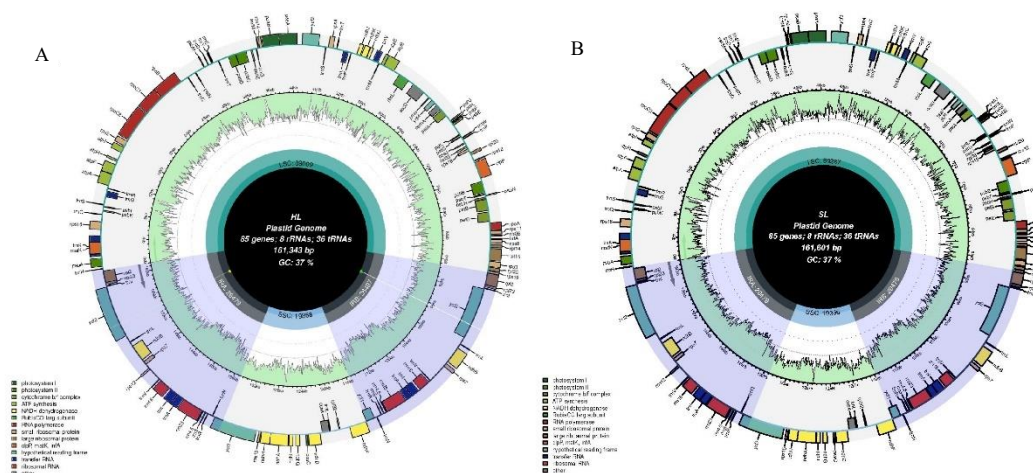


图 1 野生 (A) 与栽培 (B) 酸枣叶绿体基因组谱图

Fig. 1 Chloroplast genome profiles of wild (A) and cultivated (B) *Z. jujuba* var. *spinosa*

野生和栽培酸枣的叶绿体基因组均由大单拷贝区 (large single-copy region, LSC)、小单拷贝区 (small single-copy region, SSC) 和反向重复区 (inverted repeat region, IRa、IRb) 组成。研究发现，2 种酸枣叶绿体基因组的长度基本一致，在 1 611 100 bp 左右。

野生和栽培酸枣叶绿体基因组均编码 129 个基

因，归属为 3 大类，包括 85 个蛋白编码基因 (protein coding gene, PCG)、8 个核糖体 RNA (ribosomal RNA, rRNA) 基因和 36 个转运 RNA (transfer RNA, tRNA) 基因。其中与基因表达相关的基因 81 个，与光合作用相关基因 26 个，与生物合成相关的基因 7 个，未知功能基因 4 个 (表 2)。

表 2 酸枣叶绿体基因组的基因组成

Table 2 Gene composition of chloroplast genome of *Ziziphus jujuba* var. *spinosa*

基因功能	家族名称	基因名称
基因表达	核糖体小亚基	<i>rps2</i> 、 <i>rps3</i> 、 <i>rps7</i> 、 <i>rps7</i> 、 <i>rps8</i> 、 <i>rps11</i> 、 <i>rps12</i> 、 <i>rps12</i> 、 <i>rps14</i> 、 <i>rps15</i> 、 <i>rps16*</i> 、 <i>rps18</i> 、 <i>rps19</i>
	rRNA	<i>rrn4.5</i> 、 <i>rrn5</i> 、 <i>rrn16</i> 、 <i>rrn23</i> 、 <i>rrn23</i> 、 <i>rrn16</i> 、 <i>rrn5</i> 、 <i>rrn4.5</i>
	核糖体大亚基	<i>rpl33</i> 、 <i>rpl20</i> 、 <i>rpl36</i> 、 <i>rpl14</i> 、 <i>rpl16</i> 、 <i>rpl22</i> 、 <i>rpl2</i> 、 <i>rpl2</i> 、 <i>rpl23</i> 、 <i>rpl23</i> 、 <i>rpl32</i>
	DNA依赖性RNA聚合酶	<i>rpoC2</i> 、 <i>rpoC1</i> 、 <i>rpoB</i> 、 <i>rpoA</i>
	tRNA	<i>trnH-GUG</i> 、 <i>trnK-UUU</i> 、 <i>trnQ-UUG</i> 、 <i>trnS-GCU</i> 、 <i>trnG-GCC*</i> 、 <i>trnR-UCU</i> 、 <i>trnC-GCA</i> 、 <i>trnD-GUC</i> 、 <i>trnY-GUA</i> 、 <i>trnE-UUC</i> 、 <i>trnT-GGU</i> 、 <i>trnG-UCC</i> 、 <i>trnfM-CAU</i> 、 <i>trnS-GGA</i> 、 <i>trnT-UGU</i> 、 <i>trnL-UAA</i> 、 <i>trnF-GAA</i> 、 <i>trnV-UAC</i> 、 <i>trnM-CAU</i> 、 <i>trnW-CCA</i> 、 <i>trnP-UGG</i> 、 <i>trnI-CAU</i> 、 <i>trnL-CAA</i> 、 <i>trnV-GAC</i> 、 <i>trnI-GAU</i> 、 <i>trnA-UGC</i> 、 <i>trnR-ACG</i> 、 <i>trnN-GUU</i> 、 <i>trnL-UAG</i>
光合作用	ATP合酶亚基	<i>atpA</i> 、 <i>atpF**</i> 、 <i>atpH</i> 、 <i>atpI</i> 、 <i>atpE</i> 、 <i>atpB</i>
	NADH-脱氢酶的亚基	<i>ndhJ</i> 、 <i>ndhK</i> 、 <i>ndhC</i> 、 <i>ndhB</i> 、 <i>ndhF</i> 、 <i>ndhD</i> 、 <i>ndhG</i> 、 <i>ndhI</i> 、 <i>ndhA*</i> 、 <i>ndhH</i>
	细胞色素b/f复合物的亚基	<i>petN</i> 、 <i>petA</i> 、 <i>petL</i> 、 <i>petG</i> 、 <i>petB*</i> 、 <i>petD</i>
	光系统I的亚基	<i>psaB</i> 、 <i>psaA</i> 、 <i>psaI</i> 、 <i>psaJ</i> 、 <i>psaC</i>
	光系统II的亚基	<i>psbA</i> 、 <i>psbK</i> 、 <i>psbI</i> 、 <i>psbM</i> 、 <i>psbD</i> 、 <i>psbC</i> 、 <i>psbZ</i> 、 <i>psbJ</i> 、 <i>psbF</i> 、 <i>psbL</i> 、 <i>psbE</i> 、 <i>psbB</i> 、 <i>psbT</i> 、 <i>psbN</i> 、 <i>psbH</i>
生物合成	二磷酸核酮糖氧合酶/羧化酶亚基	<i>rbcl</i>
	乙酰-CoA-羧化酶的亚基	<i>accD</i>
	包膜蛋白基因	<i>cemA</i>
	蛋白酶基因	<i>ClpP*</i>
	翻译起始因子	<i>infA</i>
	c型细胞色素合成基因	<i>ccsA</i>
	成熟酶基因	<i>matK</i>
未知功能	保守开放阅读框	<i>ycf1</i> 、 <i>ycf2</i> 、 <i>ycf3</i> 、 <i>ycf4</i>

*代表有 1 个内含子，**表示有 2 个内含子。

*indicates one intron, ** indicates two introns.

LSC 区域的基因数量最多，有 82 个基因，包括 61 个蛋白质编码的基因和 21 个 tRNA 基因；SSC 区域的基因数量最少，只有 12 个，包括蛋白质编码的 11 个基因和 1 个 tRNA 基因；IR 两区含有 17 个基因，6 个是蛋白质编码基因，7 个是 tRNA 基因，4 个是 rRNA 基因。

3.3 密码子偏好性分析

计算野生和栽培酸枣叶绿体基因组 CDS 的密码子时发现（表 3），RSCU 最小值为 0.389，与编码丝氨酸 Ser 的 AGC 密码子相对应，RSCU 最高为 1.869，对应编码亮氨酸 Leu 的 TTA；在 61 个密码子中，对编码色氨酸 Trp 和 Met 的应用没有明显的偏好（RSCU=1）；在所有密码子中，编码精氨酸 Arg、亮氨酸 Leu 及丝氨酸 Ser 的密码子数量最多，均有 6 个。野生和栽培酸枣叶绿体基因组中，均有

61 种密码子，共同编码 20 种氨基酸（图 2）。

根据表 3 的结果，野生和栽培酸枣 RSCU 值 > 1 的高频密码子均为 29 个，28 个以 A/T 结尾，1 个以 G 结尾。根据表 4 和表 5 的结果，野生和栽培酸枣筛选出 $\Delta RSCU \geq 0.08$ 的高表达密码子各 25 个，野生酸枣中 9 个以 T 结尾、7 个以 A 结尾、6 个以 C 结尾、3 个以 G 结尾；栽培酸枣中 9 个以 T 结尾、6 个以 A 结尾、7 个以 C 结尾、3 个以 G 结尾。将满足 $RSCU > 1$ 和 $\Delta RSCU \geq 0.08$ 两者条件的密码子确定为酸枣叶绿体基因组的最优密码子，野生酸枣筛选出 GCA、GCT、TGT、GAA、TTT、GGT、CAT、ATT、AAA、CTT、TTA、TTG、CCA、CAA、CGA、TCT、GTT 17 个最优密码子，其中 7 个以 A 结尾、9 个以 T 结尾、1 个以 G 结尾；栽培酸枣筛选出 GCA、GCT、GAA、TTT、GGT、CAT、ATT、AAA、CTT、

表 3 酸枣叶绿体基因组同义密码子相对使用度分析

Table 3 Analysis of relative use of synonymous codons in chloroplast genome of *Z. jujuba* var. *spinosa*

氨基酸	密码子	相对同义密码子	氨基酸	密码子	相对同义密码子
Phe	TTT*	1.305	Thr	ACT*	1.623
Leu	TTC	0.695	Thr	ACC	0.748
Leu	TTA*	1.868	Thr	ACA*	1.172
Leu	TTG*	1.200	Thr	ACG	0.457
Ser	TCT*	1.702	Asn	AAT*	1.510
Ser	TCC	0.902	Asn	AAC	0.490
Ser	TCA*	1.207	Lys	AAA*	1.526
Ser	TCG	0.617	Lys	AAG	0.474
Tyr	TAT*	1.593	Ser	AGT*	1.183
Tyr	TAC	0.407	Ser	AGC	0.389
Cys	TGT*	1.495	Arg	AGA*	1.812
Cys	TGC	0.505	Arg	AGG	0.635
Trp	TGG	1.000	Val	GTT*	1.410
Leu	CTT*	1.264	Val	GTC	0.504
Leu	CTC	0.414	Val	GTA*	1.510
Leu	CTA	0.850	Val	GTG	0.576
Leu	CTG	0.406	Ala	GCT*	1.829
Pro	CCT*	1.501	Ala	GCC	0.658
Pro	CCC	0.758	Ala	GCA*	1.048
Pro	CCA*	1.148	Ala	GCG	0.465
Pro	CCG	0.594	Asp	GAT*	1.595
HIS	CAT*	1.520	Asp	GAC	0.405
HIS	CAC	0.480	Glu	GAA*	1.505
Arg	CGT*	1.274	Glu	GAG	0.495
Arg	CGC	0.406	Gly	GGT*	1.259
Arg	CGA*	1.414	Gly	GGC	0.449
Arg	CGG	0.459	Gly	GGA*	1.612
Ile	ATT*	1.447	Gly	GGG	0.680
Ile	ATC	0.585	Gln	CAA*	1.535
Ile	ATA	0.968	Gln	CAG	0.465
Met	ATG	1.000			

*表示 RSCU>1 的密码子。

* indicates codons with RSCU > 1.

TTG、CCA、CAA、CGA、CGT、TCT、GTT 16 个最优密码子，6 个以 A 结尾，9 个以 T 结尾，1 个以 G 结尾，由此可见，在酸枣叶绿体基因组中，偏向于以 A/T 结尾。对比分析野生和栽培酸枣最优密码子，有 15 个密码子相同，TGT、TTA 为野生酸枣特有最优密码子、CGT 为栽培酸枣特有最优密码子。

3.4 散在重复序列与 SSR 分析

在野生和栽培酸枣叶绿体基因组中，共检测到 144 个散在重复序列，其中野生酸枣 73 个，栽培酸枣 71 个；共包含 103 个 F 型重复序列和 77 个 P 型重复序列，并未发现 R 型和 C 型重复序列；此外，共检测到 180 个 SSR 位点，野生酸枣 89 个，栽培酸枣 91 个。其中单碱基重复数量最多，其次是双碱

基重复序列，五碱基重复数量最少（图 3-A）。2 种酸枣 A/T 的单碱基重复分别有 64、66 个，AT/TA/TC 双碱基重复各有 13 个，AAT/ATT/TTA 的 3 碱基重复各有 4 个，ATAA/ATTG/CTTT/GAAA/TCAA/TTTA/TTTC 4 碱基重复各有 7 个，TAAAT 5 碱基重复各有 1 个。经统计绝大多数 SSR 位于基因组基因间隔区，其次是内含子区域，基因区域内分布的 SSR 数量最低（图 3-B）。

3.5 IR 区域收缩、扩张分析

通过对野生和栽培酸枣与 NCBI 中枣（KU351660.1）、冬枣（MF781071.1）、酸枣（KX266830.1）、鼠李科毛果枳椇（MT225404.1）及滇刺枣（NC-037151.1）的叶绿体基因组序列 IR 边

表 4 野生酸枣基因编码区最优密码子分析

Table 4 Optimal codon analysis of coding region of wild *Z. jujuba* var. *spinosa* genes

氨基酸	密码子	高表达基因		低表达基因		Δ RSCU	氨基酸	密码子	高表达基因		低表达基因		Δ RSCU
		数目	RSCU	数目	RSCU				数目	RSCU	数目	RSCU	
Ala	GCA	6	1.143	14	0.789	0.354 2	Pro	CCC	4	1.067	7	0.609 0	0.458 0
Ala	GCC	2	0.381	11	0.620	-0.238 7	Pro	CCG	2	0.533	5	0.435 0	0.098 5
Ala	GCG	0	0	11	0.620	-0.619 7	Pro	CCT	3	0.800	18	1.565 0	-0.765 2
Ala	GCT	13	2.476	35	1.972	0.504 4	Gln	CAA	6	1.714	25	1.471 0	0.243 7
Cys	TGC	1	0.250	3	0.353	-0.102 9	Gln	CAG	1	0.286	9	0.529 0	-0.243 7
Cys	TGT	7	1.750	14	1.647	0.102 9	Arg	AGA	7	1.313	16	1.655 0	-0.342 7
Asp	GAC	5	0.769	10	0.500	0.269 2	Arg	AGG	2	0.375	6	0.621 0	-0.245 7
Asp	GAT	8	1.231	30	1.500	-0.269 2	Arg	CGA	16	3.000	23	2.379 0	0.620 7
Glu	GAA	22	1.760	40	1.429	0.331 4	Arg	CGC	1	0.188	3	0.310 0	-0.122 8
Glu	GAG	3	0.240	16	0.571	-0.331 4	Arg	CGG	3	0.563	5	0.517 0	0.045 3
Phe	TTC	8	0.533	14	0.683	-0.149 6	Arg	CGT	3	0.563	5	0.517 0	0.045 3
Phe	TTT	22	1.467	27	1.317	0.149 6	Ser	AGC	1	0.200	2	0.194 0	0.006 5
Gly	GGA	11	1.630	24	1.920	-0.290 4	Ser	AGT	8	1.600	16	1.548 0	0.051 6
Gly	GGC	4	0.593	8	0.640	-0.047 4	Ser	TCA	3	0.600	16	1.548 0	-0.948 4
Gly	GGG	4	0.593	8	0.640	-0.047 4	Ser	TCC	6	1.200	7	0.677 0	0.522 6
Gly	GGT	8	1.185	10	0.800	0.385 2	Ser	TCG	3	0.600	7	0.677 0	-0.077 4
His	CAC	1	0.333	6	0.923	-0.589 8	Ser	TCT	9	1.800	14	1.355 0	0.445 2
His	CAT	5	1.667	7	1.077	0.589 8	Thr	ACA	4	0.800	23	1.508 0	-0.708 2
Ile	AUA	13	0.867	27	1.039	-0.171 8	Thr	ACC	7	1.400	13	0.853 0	0.547 5
Ile	AUC	6	0.400	17	0.654	-0.253 8	Thr	ACG	4	0.800	7	0.459 0	0.341 0
Ile	ATT	26	1.733	34	1.308	0.425 6	Thr	ACT	5	1.000	18	1.180 0	-0.180 3
Lys	AAA	26	1.576	26	1.405	0.170 4	Val	GTA	13	1.368	20	1.356 0	0.012 5
Lys	AAG	7	0.424	11	0.595	-0.170 4	Val	GTC	1	0.105	8	0.542 0	-0.437 1
Leu	CTA	5	0.556	9	0.701	-0.145 7	Val	GTG	2	0.211	9	0.610 0	-0.399 7
Leu	CTC	5	0.556	8	0.623	-0.067 8	Val	GTT	22	2.316	22	1.492 0	0.824 3
Leu	CTG	2	0.222	9	0.701	-0.479 1	Trp	TGG	10	1.000	13	1.000 0	0
Leu	CTT	13	1.444	15	1.169	0.275 6	Tyr	TAC	3	0.375	5	0.278 0	0.097 2
Leu	TTA	19	2.111	26	2.026	0.085 1	Tyr	TAT	13	1.625	31	1.722 0	-0.097 2
Leu	TTG	10	1.111	10	0.779	0.331 9	Asn	AAC	9	0.900	11	0.578 9	0.321 1
Met	AUG	21	1.000	29	1.000	0.000 0	Asn	AAT	11	1.100	27	0.421 1	-0.321 1
Pro	CCA	6	1.600	16	1.391	0.208 7							

加粗的密码子为高表达密码子。
The bolded codons are the highly expressed codons.

表 5 栽培酸枣基因编码区最优密码子分析

Table 5 Optimal codon analysis of coding region of cultivated *Z. jujuba* var. *spinosa* genes

氨基酸	密码子	高表达基因		低表达基因		Δ RSCU	氨基酸	密码子	高表达基因		低表达基因		Δ RSCU
		数目	RSCU	数目	RSCU				数目	RSCU	数目	RSCU	
Ala	GCA	6	1.142 9	14	0.788 7	0.354 2	Asn	AAC	10	0.909 1	11	0.578 9	0.330 2
Ala	GCC	2	0.381 0	11	0.619 7	-0.238 7	Asn	AAT	12	1.090 9	27	1.421 1	-0.330 2
Ala	GCG	0	0	11	0.619 7	-0.619 7	Pro	CCA	6	1.500 0	16	1.391 3	0.108 7
Ala	GCT	13	2.476 2	35	1.971 8	0.504 4	Pro	CCC	4	1.000 0	7	0.608 7	0.391 3
Cys	TGC	1	0.285 7	3	0.352 9	-0.067 2	Pro	CCG	2	0.500 0	5	0.434 8	0.065 2
Cys	TGT	6	1.714 3	14	1.647 1	0.067 2	Pro	CCT	4	1.000 0	18	1.565 2	-0.565 2
Asp	GAC	6	0.923 1	10	0.500 0	0.423 1	Gln	CAA	4	1.600 0	25	1.470 6	0.129 4
Asp	GAT	7	1.076 9	30	1.500 0	-0.423 1	Gln	CAG	1	0.400 0	9	0.529 4	-0.129 4
GIT	GAA	21	1.909 1	40	1.428 6	0.480 5	Arg	AGA	6	1.241 4	16	1.655 2	-0.413 8

表 5 (续)

氨基酸	密码子	高表达基因		低表达基因		Δ RSCU	氨基酸	密码子	高表达基因		低表达基因		Δ RSCU
		数目	RSCU	数目	RSCU				数目	RSCU	数目	RSCU	
Glt	GAG	1	0.090 9	16	0.571 4	-0.480 5	Arg	CGA	13	2.689 7	23	2.379 3	0.310 4
Phe	TTC	9	0.562 5	14	0.682 9	-0.120 4	Arg	CGC	1	0.206 9	3	0.310 3	-0.103 4
Phe	TTT	23	1.4375	27	1.317 1	0.120 4	Arg	CGG	3	0.620 7	5	0.517 2	0.103 5
Gly	GGA	12	1.714 3	24	1.920 0	-0.205 7	Arg	CGT	4	0.827 6	5	0.517 2	0.310 4
Gly	GGC	4	0.571 4	8	0.640 0	-0.068 6	Ser	AGC	1	0.193 5	2	0.193 5	0.000 0
Gly	GGG	5	0.714 3	8	0.640 0	0.074 3	Ser	AGT	8	1.548 4	16	1.548 4	0.000 0
Gly	GGT	7	1.000 0	10	0.800 0	0.200 0	Ser	TCA	4	0.774 2	16	1.548 4	-0.774 2
His	CAC	1	0.333 3	6	0.923 1	-0.589 8	Ser	TCC	5	0.967 7	7	0.677 4	0.290 3
His	CAT	5	1.666 7	7	1.076 9	0.589 8	Ser	TCG	3	0.580 6	7	0.677 4	-0.096 8
Ile	ATA	14	0.933 3	27	1.038 5	-0.105 2	Ser	TCT	10	1.935 5	14	1.354 8	0.580 7
Ile	ATC	5	0.333 3	17	0.653 8	-0.320 5	Thr	ACA	4	0.761 9	23	1.508 2	-0.746 3
Ile	ATT	26	1.733 3	34	1.307 7	0.425 6	Thr	ACC	6	1.142 9	13	0.852 5	0.290 4
Lys	AAA	25	1.724 1	26	1.405 4	0.318 7	Thr	ACG	5	0.952 4	7	0.459 0	0.493 4
Lys	AAG	4	0.275 9	11	0.594 6	-0.318 7	Thr	ACT	6	1.142 9	18	1.180 3	-0.037 4
LeT	CTA	7	0.763 6	9	0.701 3	0.062 3	Val	GTA	14	1.400 0	20	1.355 9	0.044 1
LeT	CTC	7	0.763 6	8	0.623 4	0.140 2	Val	GTC	1	0.100 0	8	0.542 4	-0.442 4
LeT	CTG	2	0.218 2	9	0.701 3	-0.483 1	Val	GTG	3	0.300 0	9	0.610 2	-0.310 2
LeT	CTT	13	1.418 2	15	1.168 8	0.249 4	Val	GTT	22	2.200 0	22	1.491 5	0.708 5
LeT	TTA	18	1.963 6	26	2.026 0	-0.062 4	Trp	TGG	9	1.000 0	13	1.000 0	0.000 0
LeT	TTG	8	0.872 7	10	0.779 2	0.093 5	Tyr	TAC	4	0.470 6	5	0.277 8	0.192 8
Met	ATG	19	1.000 0	29	1.000 0	0.000 0	Tyr	TAT	13	1.529 4	31	1.722 2	-0.192 8
Arg	AGG	2	0.413 8	6	0.620 7	-0.206 9							

加粗的密码子为高表达密码子。
 The bolded codons are the highly expressed codons.

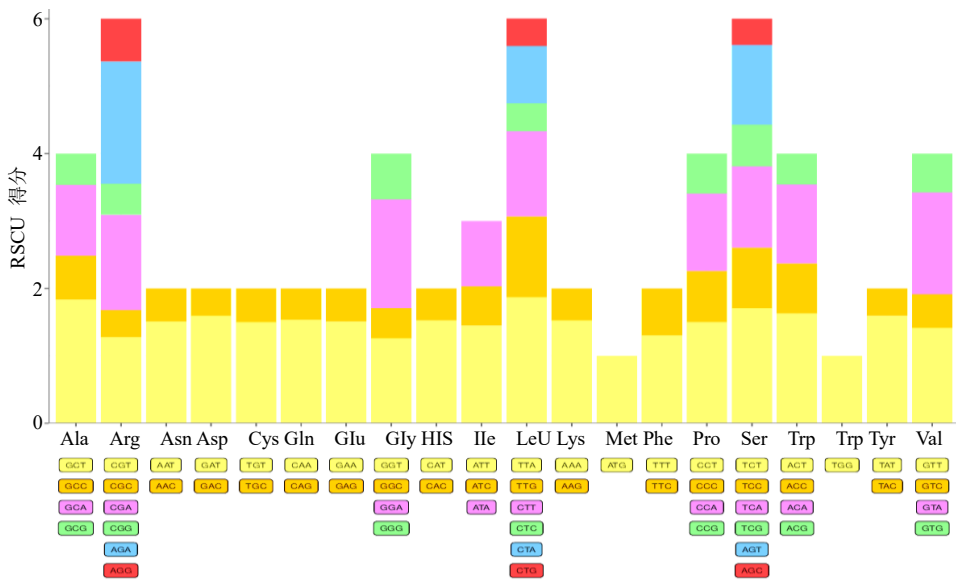
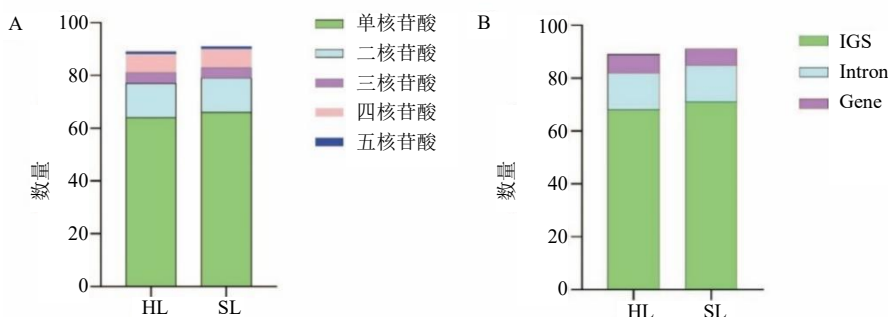


图 2 各编码氨基酸密码子 RSCU 值分布
 Fig. 2 RSCU value distribution of each encoded amino acid codon



A-SSR 类型及数量的关系; B-SSR 位置分布数量关系。

A-relationship between SSR type and number; B-location distribution and quantity relationship of SSR.

图 3 酸枣种质资源 SSR 情况

Fig. 3 SSR situation of *Z. jujuba* var. *spinosa* germplasm resources

界的特征比较分析 (图 4), 发现在 JLB (LSC/IRb) 区域、JSB (IRb/SSC) 区域和 JSA (SSC/Ira) 区域存在差异。在 JLB 区域中, 除了参考序列 MT225404.1、NC-037151.1 的 JLB 位于 *rpl22* 与 *rpl2* 基因之间外, 野生酸枣和栽培酸枣与其他参考序列的 JLB 均是被 *rps19* 基因横跨, *rpl2* 基因完全在 IRb 区, 且两种酸枣 *rps19* 基因均向 IRb 区扩张 1 bp; JSB 区域中, 栽培酸枣与参考序列 KX266830.1 JSB 区域一致, 被 *ycf1* 基因横跨, 且只有 8 bp 位于 SSC 区, 其余 1 132 bp 位于 IRb 区, 其他几对参考序列 JSB 则位于 *trnN* 和 *ndhF* 基因之间, 相同点是 *ndhF* 基因完全位于 SSC 区, 而野生酸枣则完全不同, 其 JSB 区被 *ycf2* 基因横跨, 位于 SSC 区域的是 *trnL* 基因; 在 JSA 区域中, 野生酸枣和栽培酸枣 JSA 与其他参考序列相比, 发生较大差异, 野生酸枣 JSA 被 *ycf2* 基因横跨, 5 129 bp 位于 IRa 区, 1 777 bp 位于 SSC 区, *trnL* 基因完全位于 SSC 区; 栽培酸枣 JSA 同参考序列 KU351660.1、KX266830.1、MT225404.1/NC-037151.1 则被 *ycf1* 基因横跨, 4 592 bp 位于 SSC 区, 1 132 bp 位于 IRa 区, *trnN* 基因完全位于 IRa 区; 酸枣参考序列 MF781071.1 JSA 区域则位于 *rps15* 基因和 *trnN* 基因之间。JLA 区均位于 *rpl2* 基因和 *trnH* 基因之间, 无明显差异。

3.6 核酸多态性分析

通过对酸枣的核酸多态性分析可知 (图 5), 野生酸枣 P_i 值在 0~0.006, 平均值是 0.000 208, 总共获得 2 个高变异片段, 即位于 LSC 区的 *rpoC2* 基因和 *rpoB* 基因, P_i 值均大于 0.004; 栽培酸枣 P_i 值在 0~0.003, 平均值是 0.000 057, 总共获得 1 个高变异片段, 即位于 LSC 区的 *atpA* 基因, P_i 值大于

0.002 5, 这 3 个基因片段可以成为酸枣物种鉴定的候选片段。

3.7 系统发育树分析

从 NCBI 网站上下载 30 条酸枣近缘物种基因组序列, 使用最大似然法 ML 构建系统进化树 (图 6)。结果表明, 野生酸枣与酸枣参考序列 KU351660.1 聚为一支; 而栽培酸枣则与参考序列 KX266829.1 (骏枣 *Ziziphus jujuba* Mill. cv. 'Junzao')、KU351660.1 (枣 *Ziziphus jujuba* Mill.)、KX266830.1 (酸枣 *Ziziphus jujuba* var. *spinosa*)、MF781071.1 (冬枣 *Ziziphus jujuba* Mill. cv. 'Dongzao') 等多个枣属植物聚为一支, 表明栽培酸枣居群间存在频繁调种和换种, 基因交流比较频繁。

4 讨论

植物表型是指在遗传和环境的交互作用下, 表现出的一系列物理、生理、生化特性, 是植物基因型与环境因素相互作用的结果。随着中药材需求量的增加, 野生中药材资源不断减少, 中药材栽培产业不断发展, 野生和栽培中药材资源在外观性状上发生了显著变化。周红涛等^[24]对野生和栽培的芍药遗传变异进行研究, 发现芍药的野生群体具有较丰富的遗传多样性, 而栽培群体的多态性比率较低; 贺宗毅等^[25]利用 SNP 分子标记技术对野生和栽培灵芝遗传多样性研究, 发现野生灵芝菌株遗传背景更复杂, 具有丰富的遗传多态性; Yuan 等^[26]对黄芩野生群体及 22 个栽培群体进行了 3 个叶绿体基因多态性分析, 结果表明, 栽培群体间的遗传分化系数显著降低, 并丧失了许多野生群体的珍稀等位基因。研究结果表明, 野生药用植物在人工栽培和驯化过程中, 其基因型发生了变化, 一些珍稀基因在

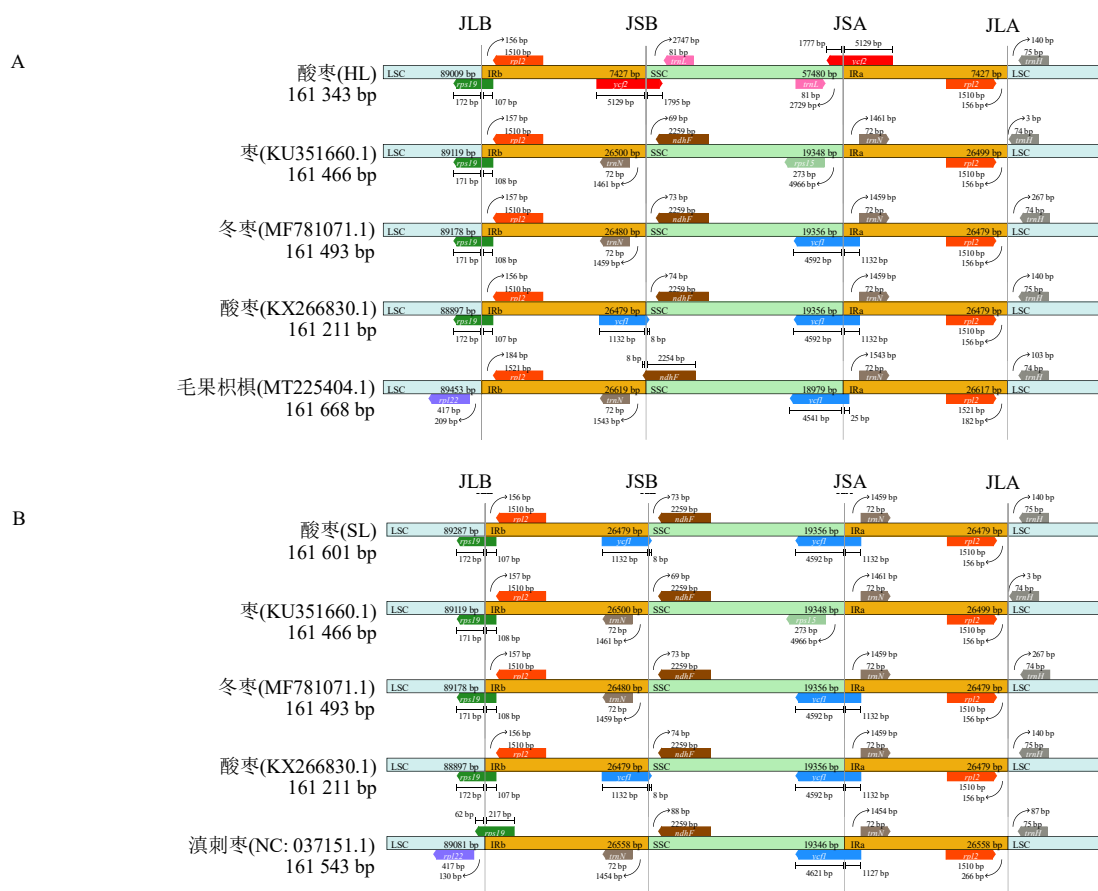


图 4 野生酸枣 (A) 和栽培酸枣 (B) 叶绿体基因组 IR/SC 边界比对图

Fig. 4 IR/SC boundary comparison diagrams of chloroplast genome of wild *Z. jujuba* var. *spinosa* (A) and cultivated *Z. jujuba* var. *spinosa* (B)

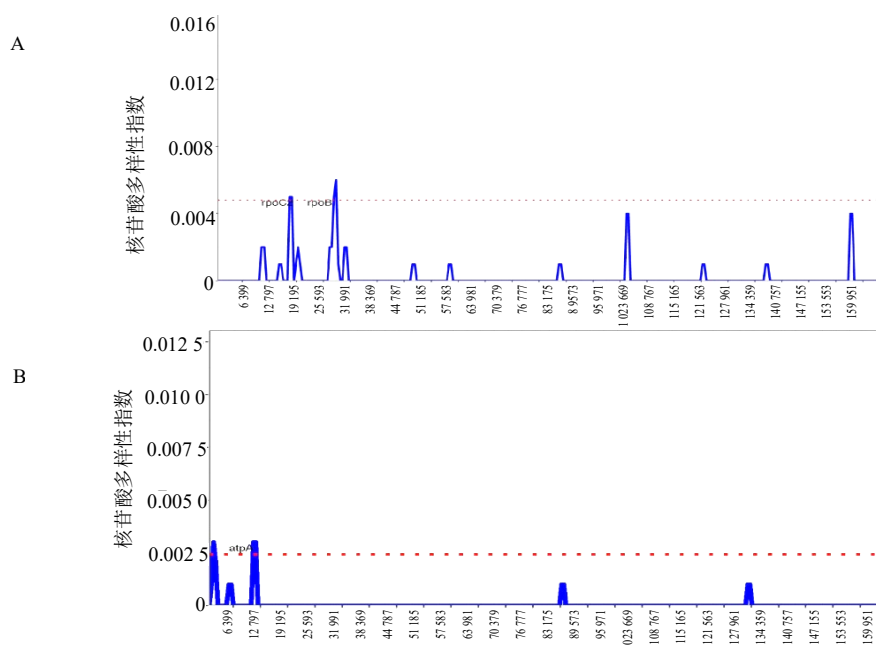
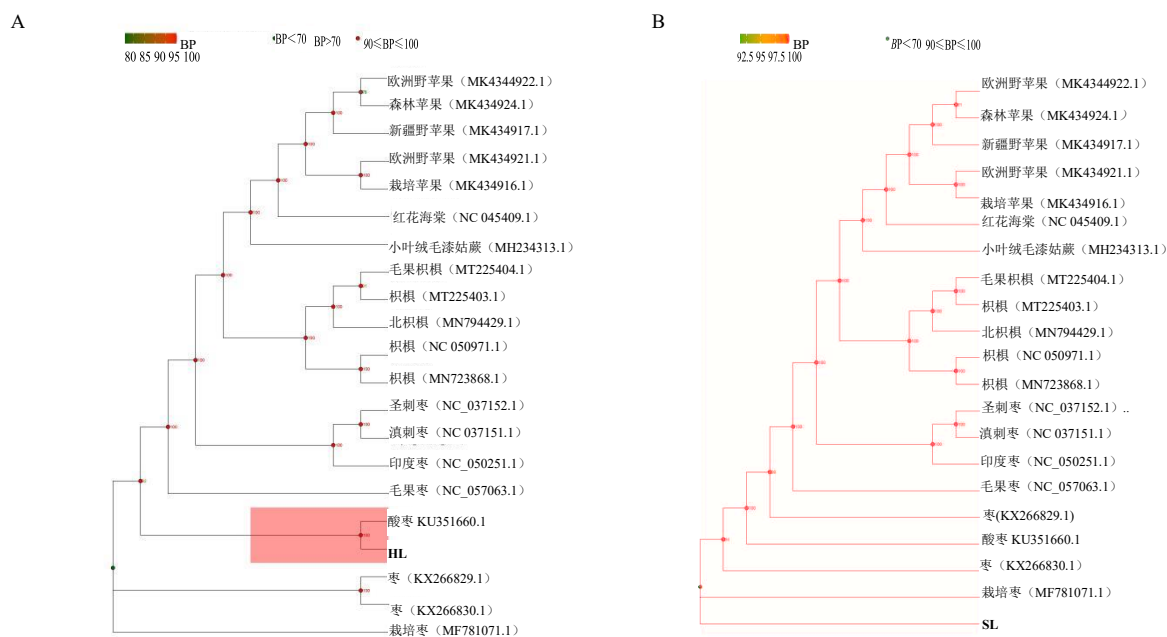


图 5 野生酸枣 (A) 和栽培酸枣 (B) 核酸多态性分析图

Fig. 5 Nucleic acid polymorphism analysis diagrams of wild *Z. jujuba* var. *spinosa* (A) and cultivated *Z. jujuba* var. *spinosa* (B)



A-野生酸枣; B-栽培酸枣。
A-wild *Z. jujuba* var. *spinosa*; B-cultivated *Z. jujuba* var. *spinosa*.

图 6 基于 30 条叶绿体基因组序列 NJ 法构建系统发育树

Fig. 6 Construction of phylogenetic trees based on NJ method of 30 chloroplast genome sequences

人工驯化过程中流失, 导致其遗传多样性下降。因此, 开展野生和栽培酸枣的基因型差异研究具有重要意义。

本研究发现,野生和栽培酸枣叶绿体基因组大小分别为 161 343、161 601 bp,与 Kim 等^[27]对 *Ziziphus jujuba* cv. Bokjo 叶绿体基因组测序结果基本一致,符合大多数高等植物的叶绿体基因组高度保守的四分体结构,包括大单拷贝区,小单拷贝区和 2 个反向重复区^[28]。这体现了叶绿体基因组的保守性和遗传稳定性。酸枣叶绿体基因组与其他鼠李科枣属植物如“玉铃铛”枣(161 720 bp)^[29]、骏枣(161 215 bp)、叙利亚枣(161 615 bp)^[30]相比,大小相差不大,GC 含量均为 37%,在已报道的大多数双子叶植物叶绿体 31%~38%的 GC 含量区间之内。重复序列是指在整个基因组中以多个拷贝出现的单个相同或者对称的核酸序列^[31]。包括串联重复序列和散在重复序列,串联重复序列是在基因组中较短且呈簇状排列,分为 SSR 和小卫星等;而散在重复序列在基因组中较短且散布在整个基因组中。植物叶绿体基因组中含有大量的 SSR 位点,它们提供了丰富的遗传信息,是研究物种遗传进化的重要分子标记^[32]。本研究中共获得 180 个 SSR 位点,以

单核苷酸 A/T 重复类型最多, 这与叶绿体中的 SSR 通常由 A 或 T 重复序列组成, G 或 C 序列较少的说法相一致^[33]。检测到的 SSR 位点可用于后续酸枣的种间种内群体结构、亲缘关系分析和进化关系的研究。

密码子在遗传信息传递过程中具有重要作用, 同一氨基酸可以由 2 个或 2 个以上的密码子共同编码, 这些密码子称为同义密码子, 它们在编码时使用频率不同, 称为密码子使用偏好性^[34]。在野生和栽培酸枣叶绿体基因组中, 发现超过 90% ($\text{RSCU} \geq 1$) 的密码子以 A/T 结尾, 进一步验证了高等植物叶绿体中普遍存在的 A/T 密码子偏好^[35]。野生酸枣中最优密码子较栽培酸枣中更多, 表明野生酸枣受到更多的正向选择, 基因交流更为广泛, 遗传多样性更高, 相对栽培酸枣而言, 适应能力更强, 并且可能具有更灵活的基因表达调控能力以及潜在功能多样性, 这同样意味着在长期的驯化栽培过程中, 栽培酸枣保持了人类所需要的有益或稳定的性状。

IR 边界的扩展与收缩是导致叶绿体基因组的长度和结构存在差异的重要原因, 这些边界变化对于维持叶绿体基因组结构稳定及促进物种进化

至关重要^[36]。本研究中 IR 边界与序列比对的结果表明(图 4),栽培酸枣较野生酸枣叶绿体基因组边界在基因类型方面相对更加保守,而野生酸枣基因发生较多变异,其原因可能是野生酸枣 SSC 区明显比栽培酸枣及其他枣属植物长,导致叶绿体基因组基因变异的概率较高,因而其基因组组成发生明显变异。

叶绿体基因组含丰富的突变位点,这些位点能够引起基因序列结构的改变以及核苷酸多态性的变异,其进化速度极快,因此可以作为物种鉴定的 DNA 条形码。本研究对野生和栽培酸枣叶绿体基因组序列进行比对分析,结果显示,野生酸枣 LSC 区的 *rpoC2*、*rpoB* 基因和栽培酸枣的 *atpA* 基因(图 5)能够提供丰富的变异信息,有助于酸枣物种分化、群体遗传学和系统地理学研究,由此也可以看出野生酸枣相对栽培酸枣而言具有更丰富的基因突变位点。为明确野生酸枣和栽培酸枣在系统发育中的位置,利用 30 个鼠李科物种的完整叶绿体基因组创建了 ML 树(图 6),发现野生酸枣能与枣属其他植物明显区分开,而栽培酸枣栽培过程中,可能由于人为以及环境等因素存在野生居群引种、异地种质交换、产地频繁换种、种质杂交等行为,种质的不稳定导致栽培居群的遗传多样性变化范围较大,使栽培酸枣的基因与枣属其他植物均有一定的交流,在系统发育树中未能完全将其区分开,而 Zhang 等^[37]的研究中显示野生和栽培酸枣系统发育树结果表明两者比较相似,与本研究结论存在出入,这可能与不同种源酸枣生长环境不同、基因交流频率存在差异以及人工选择等多个方面有关。

研究通过高通量测序技术获得了野生和栽培酸枣完整的叶绿体基因组序列,进一步丰富了酸枣叶绿体基因组资料,并对两者基因组基本特征、密码子偏好性、重复序列、IR 区收缩扩张以及与其他鼠李科枣属植物亲缘进行比较和分析,发现两者在简单重复序列、IR 区收缩与扩张、最优密码子数目及组成、核酸多态性片段及系统发育树方面存在差异,后续可以利用存在差异的 SSR 位点、高变异片段及 IR 区域基因类型对野生和栽培酸枣有效鉴别。上述结果对进一步探讨野生和栽培酸枣的叶绿体基因组结果和系统发育提供了一定的理论依据。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] Tanaka H, Sato M, Ogasawara Y, *et al.* Chloroplast

aggregation during the cold-positioning response in the liverwort *Marchantia polymorpha* [J]. *J Plant Res*, 2017, 130(6): 1061-1070.

[2] Kirchhoff H. Chloroplast ultrastructure in plants [J]. *New Phytol*, 2019, 223(2): 565-574.

[3] Gügel I L, Soll J. Chloroplast differentiation in the growing leaves of *Arabidopsis thaliana* [J]. *Protoplasma*, 2017, 254(5): 1857-1866.

[4] 王永惠, 曲泽州. 《中国果树志·枣卷》[M]. 北京: 中国林业出版社, 1993: 56.

[5] 杜军霞, 刘华梁, 王僧虎, 等. 酸枣及其有效活性成分的药理作用综述 [J]. *安徽农业科学*, 2022, 50(19): 7-10.

[6] 李旭, 和建政, 陈彻, 等. 酸枣仁镇静催眠活性成分及药理作用研究进展 [J]. *中华中医药学刊*, 2022, 40(2): 23-31.

[7] 唐雪燕, 秦令祥, 曹源, 等. 酸枣仁总皂苷提取工艺优化及其改善睡眠作用机制研究 [J]. *食品安全质量检测学报*, 2023, 14(17): 147-155.

[8] 张兵, 朱亚辉, 王孝玉, 等. 酸枣仁皂苷 A 促进大鼠股骨骨折愈合的机制研究 [J]. *浙江医学*, 2021, 43(18): 1953-1957.

[9] 周文文, 刘晖, 徐志佳, 等. 酸枣仁不同极性部位成分含量与抗氧化活性及其相关性分析 [J]. *食品工业科技*, 2023, 44(16): 288-296.

[10] 段开行, 王晓玲, 毛永民, 等. 酸枣种质资源果实数量性状遗传多样性分析 [J]. *园艺学报*, 2023, 50(12): 2568-2576.

[11] Chen S F, Zhou Y Q, Chen Y R, *et al.* Fastp: An ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, 34(17): i884-i890.

[12] 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.

[13] Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller [J]. *Genome Biol*, 2012, 13(6): R56.

[14] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes [J]. *Nucleic Acids Res*, 2019, 47(W1): W59-W64.

[15] Zheng S Y, Pocai P, Hyvönen J, *et al.* Chloroplast: An online program for the versatile plotting of organelle genomes [J]. *Front Genet*, 2020, 11: 576124.

[16] Hall T A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT[J]. *Nucleic Acids Symp. Ser.*, 1999, 41(41):95-98.

[17] 陆奇丰, 骆文华, 黄至欢. 两种梧桐叶绿体基因组密码

- 子使用偏性分析 [J]. 广西植物, 2020, 40(2): 173-183.
- [18] 吴宪明, 吴松峰, 任大明, 等. 密码子偏性的分析方法及相关研究进展 [J]. 遗传, 2007, 29(4): 420-426.
- [19] 唐晓芬, 陈莉, 马玉韬. 密码子使用偏性量化方法研究综述 [J]. 基因组学与应用生物学, 2013, 32(5): 660-666.
- [20] 刘心雨, 姜丹, 杨楚楚, 等. 基于 ITS 条形码的滇鸡血藤及混伪品分子鉴定 [J]. 中国现代中药, 2024, 26(1): 10-17.
- [21] Rozas J, Ferrer-Mata A, Sánchez-DelBarrio J C, *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets [J]. *Mol Biol Evol*, 2017, 34(12): 3299-3302.
- [22] Rø E. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping [J]. *Cold Spring Harbor Laboratory*, 2021.
- [23] Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies [J]. *Bioinformatics*, 2014, 30(9): 1312-1313.
- [24] 周红涛, 胡世林, 郭宝林, 等. 芍药野生与栽培群体的遗传变异研究 [J]. 药学学报, 2002, 37(5): 383-388.
- [25] 贺宗毅, 王伯初, 马传贵, 等. 野生与不同栽培类型的灵芝种质遗传多态性及群体特征 [J]. 菌物学报, 2024, 43(11): 105-119.
- [26] Yuan Q J, Zhang Z Y, Hu J, *et al.* Impacts of recent cultivation on genetic diversity pattern of a medicinal plant, *Scutellaria baicalensis* (Lamiaceae) [J]. *BMC Genet*, 2010, 11: 29.
- [27] Kim M, Park J H, Gil J, *et al.* The complete chloroplast genome of *Ziziphus jujuba* cv. Bokjo (Rhamnaceae) [J]. *Mitochondrial DNA B Resour*, 2022, 7(10): 1805-1806.
- [28] Jansen R K, Raubeson L A, Boore J L, *et al.* Methods for obtaining and analyzing whole chloroplast genome sequences [J]. *Methods Enzymol*, 2005, 395: 348-384.
- [29] 郭晓雨, 张晓慧, 马宗新, 等. ‘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, 2023, 21(8): 2488-2498.
- [30] Huang J, Yang X T, Zhang C M, *et al.* Development of chloroplast microsatellite markers and analysis of chloroplast diversity in Chinese jujube (*Ziziphus jujuba* Mill.) and wild jujube (*Ziziphus acidojujuba* Mill.) [J]. *PLoS One*, 2015, 10(9): e0134519.
- [31] 张瑞, 张天留, 宋美华, 等. 重复序列在牛亚科中的研究进展 [J]. 中国畜牧杂志, 2021, 57(4): 33-38.
- [32] Xue Z P, Feng W H, Cao J K, *et al.* Antioxidant activity and total phenolic contents in peel and pulp of Chinese jujube (*Ziziphus jujuba* Mill) fruits [J]. *J Food Biochem*, 2009, 33(5): 613-629.
- [33] Shen X F, Wu M L, Liao B S, *et al.* Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua* [J]. *Molecules*, 2017, 22(8): 1330.
- [34] 张晶晶, 袁庆, 魏瑶, 等. 忍冬属囊管组植物叶绿体基因组进化分析 [J]. 中草药, 2024, 55(9): 3085-3097.
- [35] Wu C S, Chaw S M. Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion [J]. *Genome Biol Evol*, 2015, 7(7): 2000-2009.
- [36] Li L, Hu Y F, He M, *et al.* Comparative chloroplast genomes: Insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia* [J]. *BMC Genomics*, 2021, 22(1): 138.
- [37] Zhang Y P, Hu G L, Mao W T, *et al.* Chloroplast genome sequence of the wild *Ziziphus jujuba* Mill. var. *spinosa* from north China [J]. *Mitochondrial DNA B Resour*, 2021, 6(2): 666-667.

[责任编辑 时圣明]