堇花槐与苦豆子叶绿体基因组特征及其序列比较分析

鲁仪增

(山东省林草种质资源中心 暖温带林草种质资源保存与利用国家林业和草原局重点实验室 山东济南 250102)

摘 要:【目的】基于叶绿体基因组序列比较分析加快槐属植物特异种质资源的开发利用。【方法】以槐属花色特异种质堇花槐和苦豆子为研究对象，利用 Illumina Hiseq 高通量测序，然后进行叶绿体基因组组装、结构特征与基因组注释。【结果】堇花槐和苦豆子的叶绿体基因组均为典型的由四部分组成的环状结构，其中堇花槐的 cpDNA 长度较长，为 158837bp，其反向重复序列 IRa 和 IRb 长为 25398bp，将 88983bp 大单拷贝区（LSC）和 19058bp 小单拷贝区（SSC）分开。苦豆子的 cpDNA 长度相对较短，为 154108bp，其反向重复序列 IRa 和 IRb 长为 25874bp，将 84221bp 大单拷贝区（LSC）和 18139bp 小单拷贝区（SSC）分开。堇花槐和苦豆子二者发生碱基替换的数量较多（17464bp），插入缺失相对较少（1688bp）。堇花槐的总 GC 含量 36.13%，比苦豆子的 GC 总含量低（36.57%）。其中，在反向重复区中，堇花槐的 GC 含量为 43.17%，高于苦豆子的 GC 含量（42.79%）；在大单拷贝区中，堇花槐的 GC 含量为 33.51%，低于苦豆子的 GC 含量（34.17%）；在小单拷贝区中，堇花槐的 GC 含量 29.61%，低于苦豆子的 GC 含量（30%）。AT 总含量与 GC 总含量不同，其均以小单拷贝区的含量高，大单拷贝区的含量中等，反向重复区的含量最低。其中，堇花槐的总 AT 含量 63.87%，比苦豆子的 AT 总含量高（63.43%）。在小单拷贝区中，堇花槐的 AT 含量 70.39%，高于苦豆子的 AT 含量（70%）。在大单拷贝区中，堇花槐的 AT 含量 66.49%，高于苦豆子的 AT 含量（65.83%）。在反向重复区中，堇花槐的 AT 含量 56.83%，高于苦豆子的 AT 含量（57.21%）。堇花槐和苦豆子的叶绿体基因组共编码基因的数量分别为 130 和 129 个，在 IR 及 IRa、IRb 区中，其叶绿体基因组共编码基因的数量分别为 38（19、19）和 35（18、17）个；在 LSC 区中，其叶绿体基因组共编码基因的数量分别为 80 个、82 个；SSC 区中，二者叶绿体基因组共编码基因的数量均为 12 个。堇花槐较苦豆子多一个自我复制的功能的 *rps19* 基因、1 个保守阅读框架基因 *ycf1* 基因和 1 个转运 RNA 基因 *trnfM-CAU* 基因；苦豆子独有 1 个转运 RNA 基因 *trnT-CGU* 基因。堇花槐使用的 65 个密码子中，不同氨基酸中所使用的密码子的数量及其所占比例均不相同。其中，Ala（A）中 GCT 使用 634 个密码子，占 1.85%；超过 3%以上的密码子也有 8 个，分别为 ATT（4.22%）、AAA（4.21%）、GAA（3.89%）、TTT（3.79%）、AAT（3.82%）、GAT（3.32%）、TTA（3.18%）和 TAT（3.02%），除 ATT 和 TTA 密码子较苦豆子使用较少外，其余 6 个密码子均较苦豆子多。使用最少的密码子为 TAG 和 TGA，均占 0.07%，不同于苦豆子。在堇花槐和苦豆子 cpDNA 中，单核苷酸重复基序 SSR 的数量分别为 110 和 79 个，分别占 SSR 总量的 64.33%和 69.91%；二核苷酸重复基序 SSR 的数量分别为 43 和 14 个，分别占 SSR 总量的 25.15%和 12.39%；单核苷酸重复基序与二核苷酸重复基序之和分别占所有叶绿体核基因组 SSR 的 89.47%（153 个）和 82.30%（93 个）。三核苷酸、四核苷酸及五核苷酸重复基序的 SSR 在堇花槐较少，均为 1 个；三核苷酸、四核苷酸重复基序的 SSR 在苦豆子中却相对较多，三核苷酸重复基序的 SSR 占总 SSR 数量的 6.19%（7 个），四核苷酸重复基序的 SSR 占总 SSR 数量的 11.50%（13 个），缺少五核苷酸重复基序的 SSR。【结论】探明了槐属植物堇花槐和苦豆子的叶绿体基因组同源性及其二者叶绿体基因组在序列特征、编码基因分布、密码子使用频率及 SSR 分布等方面的差异，为其叶绿体基因挖掘及开发利用奠定了基础。

关键词: 堇花槐；苦豆子；叶绿体基因组；密码子；SSR

Comparative analysis of chloroplast genome between *Sophora japonica* var. *violacea* and *S. alopecuroides*

Lu Yizeng
(Shandong Provincial Center of Forest and Grass Germplasm Resources, Key Laboratory of National Forestry and Grassland Administration on Conservation and Utilization of Warm Temperate Zone Forest and Grass Germplasm Resources, Jinan 250102, China)

Abstract:【Objective】Based on the comparative analysis of chloroplast genome sequences, to accelerate the development and utilization of *Sophora* specific germplasm resources. **【Method】** *Sophora japonica* var. *violacea* and *S. alopecuroides*, which are flower color specific Germplasms of *Sophora* genus, were selected as the research objects. The chloroplast genome assembly, structural characteristics and genome annotation were carried out by using Illumina hiseq high-throughput sequencing. **【Result】**the chloroplast genomes of *S.japonica* var. *violacea* and *S. alopecuroides* were typical circular structures composed of four parts. The cpDNA length of *S.japonica* var. *violacea* was 158837 bp, and the length of its reverse repeat IRA and IRB was 25398 bp. The 88983 bp large single copy region (LSC) and 19058 bp small single copy region (SSC) were separated. The cpDNA length of *S. alopecuroides* is relatively short, which is 154108 bp. The length of its reverse repeat IRA and IRB is 25874 bp, which separates 84221 bp large single copy region (LSC) and 18139 bp small single copy region (SSC). Both *Sophora japonica* var. *violacea* and *S. alopecuroides* has more base substitutions (17464 bp) and less insertion deletions (1688 bp). The total GC content of *S.japonica* var. *violacea* was 36.13%, which was lower than that of *S. alopecuroides* (36.57%). In the reverse repeat region, the GC content of *S. japonica* var. *violacea* was 43.17%, which was higher than that of *S. alopecuroides* (42.79%); In the large single copy region, the GC content of *S. japonica* var. *violacea* was 33.51%, which was lower than that of *S. alopecuroides* (34.17%); In the small single copy region, the GC content of *S. japonica* var. *violacea* was 29.61%, which was lower than that of *S. alopecuroides* (30%). The total content of AT was different from that of GC. The content of small single copy region was higher, the content of large single copy region was medium, and the content of reverse repeat region was the lowest. The total AT content of *S. japonica* var. *violacea* was 63.87%, which was higher than that of *S. alopecuroides* (63.43%). In the small single copy region, the AT content of *S. japonica* var. *violacea* was 70.39%, which was higher than that of *S. alopecuroides* (70%). In the large single copy region, the AT content of *S. japonica* var. *violacea* was 66.49%, which was higher than that of *S. alopecuroides* (65.83%). In the reverse repeat region, the AT content of *S. japonica* var. *violacea* was 56.83%, which was higher than that of *S. alopecuroides* (57.21%). The number of co-coding genes in the chloroplast genome of *S. japonica* var. *violacea* and *S. alopecuroides* were 130 and 129, respectively. In the IR, IRA and IRB regions, the number of co-coding genes in the chloroplast genome was 38 (19,19) and 35 (18,17), respectively; In the LSC region, the number of chloroplast genome co-coding genes was 80 and 82, respectively; In SSc region, the number of co-coding genes in the chloroplast genome of the two genes was 12. Compared with *S. alopecuroides*, *S. japonica* var. *violacea* has a self-replicating *rps19* gene, a conserved reading frame gene *ycf1* gene and a *trnfm -CAU* gene; *S. alopecuroides* has a unique *trnt-CGU* gene. Among the 65 codons used by *S. japonica* var. *violacea*, the number and proportion of codons used in different amino acids were different. Among them, the GCT in ALa (a) uses 634 codons, accounting for 1.85%; There are also 8 codons with more than 3%, which are ATT (4.22%), AAA (4.21%), GAA (3.89%), TTT (3.79%), AAT (3.82%), GAT (3.32%), TTA (3.18%) and TAT (3.02%), respectively. Except that ATT and TTA codons are less used than *S. alopecuroides*, the other 6 codons are more than *S. alopecuroides*. The lowest codons used were tag and TGA, accounting for 0.07%, which was different from *S. alopecuroides*. In the cpDNA of *S. japonica* var. *violacea* and *S. alopecuroides*, the number of SSRs with single nucleotide repeat motifs was 110 and 79 respectively, accounting for 64.33% and 69.91% of the total SSRs, respectively; The number of dinucleotide repeat motif SSRs was 43 and 14, accounting for 25.15% and 12.39% of the total SSRs, respectively; The sum of single nucleotide repeat motifs and dinucleotide repeat motifs accounted

for 89.47% (153) and 82.30% (93) of all chloroplast nuclear genomic SSRs, respectively. The SSRs of trinucleotide, tetranucleotide and pentanucleotide repeat motifs were less and they were all 1; The SSRs with trinucleotide and tetranucleotide repeat motifs were relatively more in *S. alopecuroides*. The SSRs with trinucleotide repeat motifs accounted for 6.19% (7) of the total SSRs, and those with tetranucleotide repeat motifs accounted for 11.50% (13) of the total SSRs, while those without pentanucleotide repeat motifs. **【Conclusion】** The homology of chloroplast genomes of *S. japonica* var. *violacea* and *S. alopecuroides* and the differences in sequence characteristics, coding gene distribution, codon usage frequency and SSR distribution between the two chloroplast genomes were identified, which laid a foundation for their chloroplast gene mining and development and utilization.

Key words: *Sophora japonica* var. *violacea*; *Sophora alopecuroides*; Chloroplast genome; Codon; SSR