

## 研究论文

## Research Article

# 高山特有植物丽江棱子芹的叶绿体基因组结构和系统发育关系

张婕妤<sup>1,2</sup> 俞树良<sup>1,2</sup> 李志敏<sup>1,2\*</sup> 孙文光<sup>1,2\*</sup>

1 云南师范大学生命科学学院, 昆明, 650500; 2 云南师范大学生物能源持续开发利用教育部工程研究中心, 昆明, 650500

\* 共同通信作者, lizhimin\_vip@163.com; sunwenguang@vip.163.com

**摘要** 本文基于高通量技术对丽江棱子芹(*Pleurospermum foetens*)的叶绿体基因组进行测序, 并对丽江棱子芹的叶绿体基因组结构特征及其系统进化关系进行了分析。结果表明, 丽江棱子芹的叶绿体基因组具备典型的四分体结构, 基因组全长 155 057 bp, 包含 87 个蛋白编码基因、8 个核糖体 RNA 基因和 39 个转运 RNA 基因。生物信息学分析结果显示, 叶绿体基因组共编码 26 075 个密码子, 编码异亮氨酸 Ile 的 ATT 数量最多, 除终止子外编码半胱氨酸 Cys 的 TGC 数量最少, 相对同义密码子使用度最高的为 UUA, 最少的是 AGC; SSR 位点共预测到了 34 个, 包括单核苷酸重复序列 27 个, 双核苷酸重复序列 4 个, 以及复合型重复序列 3 个。丽江棱子芹与棱子芹属(*Pleurospermum*)其他植物相比较, 其基因组大小、基因类型及 GC 含量相近。利用 IQ-TREE 软件构建最大似然树, 结果表明丽江棱子芹与线裂棱子芹(*P. linearilobum*)的亲缘关系最近, 棱子芹属为多系类群。该研究丰富了丽江棱子芹叶绿体基因组的遗传信息, 为丽江棱子芹的分子标记的开发、遗传多样性和谱系地理研究等奠定了理论基础, 同时也为构建伞形科(Apiaceae)植物的系统进化提供了支持。

**关键词** 丽江棱子芹; 叶绿体基因组; 序列结构; 系统发育分析

## Chloroplast Genome Structure and Phylogenetic Relationships in the Alpine Endemic Plant *Pleurospermum foetens*

ZHANG Jieyu<sup>1,2</sup> YU Shuliang<sup>1,2</sup> LI Zhimin<sup>1,2\*</sup> SUN Wenguang<sup>1,2\*</sup>

1 School of Life Science of Yunnan Normal University, Kunming, 650500; 2 Engineering Research Center of Sustainable Development and Utilization of Biomass Energy, Ministry of Education in Yunnan Normal University, Kunming, 650500

\* Co-corresponding authors, lizhimin\_vip@163.com; sunwenguang@vip.163.com

DOI: 10.13417/jgab042000903

**Abstract** This study sequenced the chloroplast genome of *Pleurospermum foetens* by high throughput technique, and analyzed its structural characteristics and phylogenetic relationship. The results showed that the chloroplast genome was a typical tetrad structure. The total length of the genome is 155 057 bp, with 87 protein coding genes, 8 ribosomal RNA genes, and 39 transfer RNA genes. Bioinformatics analysis revealed that the chloroplast genome encodes a total of 26 075 codons, with the highest number of ATT coding for Ile and the lowest number of TGC coding for Cys, apart from the terminator. The highest relative synonymous codon usage was UUA, and the lowest was AGC. Additionally, 34 SSR loci were identified, including 27 single nucleotide repeats sequences, 4 double nucleotide re-

基金项目: 本研究由第二次青藏高原综合科学考察研究(2019QZKK05020)、国家自然科学基金项目(31960046)、国家基金-云南省联合基金重点项目(U1802232)及云南省科技厅基础研究专项——青年项目(202201AU070057)共同资助。

引用格式: 张婕妤, 俞树良, 李志敏, 等, 2023. 高山特有植物丽江棱子芹的叶绿体基因组结构和系统发育关系. 基因组学与应用生物学, 42(9): 903-918. [ZHANG J Y, YU S L, LI Z M, et al., 2023. Chloroplast genome structure and phylogenetic relationships in the alpine endemic plant *Pleurospermum foetens*. Genomics and Applied Biology, 42(9): 903-918.]

peats sequences, and 3 complex repeats sequences. The genome size, gene type and GC content of *P. foetens* were similar with other *Pleurospermum* species. A maximum likelihood tree was constructed using the IQ-TREE software, the results showed that *P. foetens* is closely related to *P. linearilobum* and the *Pleurospermum* branch is a paraphyletic group. This study enriched the genetic information of chloroplast genome of *P. foetens*, and laid a theoretical foundation for the development of molecular markers, genetic diversity and phylogeography of *P. foetens*. It also provides support for the systematic evolution of Apiaceae plants.

**Keywords** *Pleurospermum foetens*; Chloroplast genome; Sequence structure; Phylogenetic analysis

叶绿体是绿色植物中可以进行光合作用的质体,其基因组一般为环状双链,典型的四分体结构,包含大单拷贝区( large single copy region, LSC)、小单拷贝区( small single copy region, SSC)和两段序列相同、方向相反的反向重复区( inverted repeat region, IRa 和 IRb) ( Jansen et al., 2005)。在被子植物中叶绿体基因组大多为母系遗传,与核基因组和核糖体基因组相比,叶绿体基因组较小,在高等植物中有 120~220 kb。由于其具有基因组所含的功能基因数目少和突变率较低的特点,在植物系统进化与谱系地理学中常作为优良的遗传变异标记( Palmer, 1985; Wolfe et al., 1987; 汪小全和洪德元, 1997)。基于物种叶绿体基因组的结构及组成等分析,还可以开展更深入的生物信息学、群体遗传学等研究( Cai et al., 2008; Guisinger et al., 2011; Liang et al., 2019; 余涛等, 2022)。

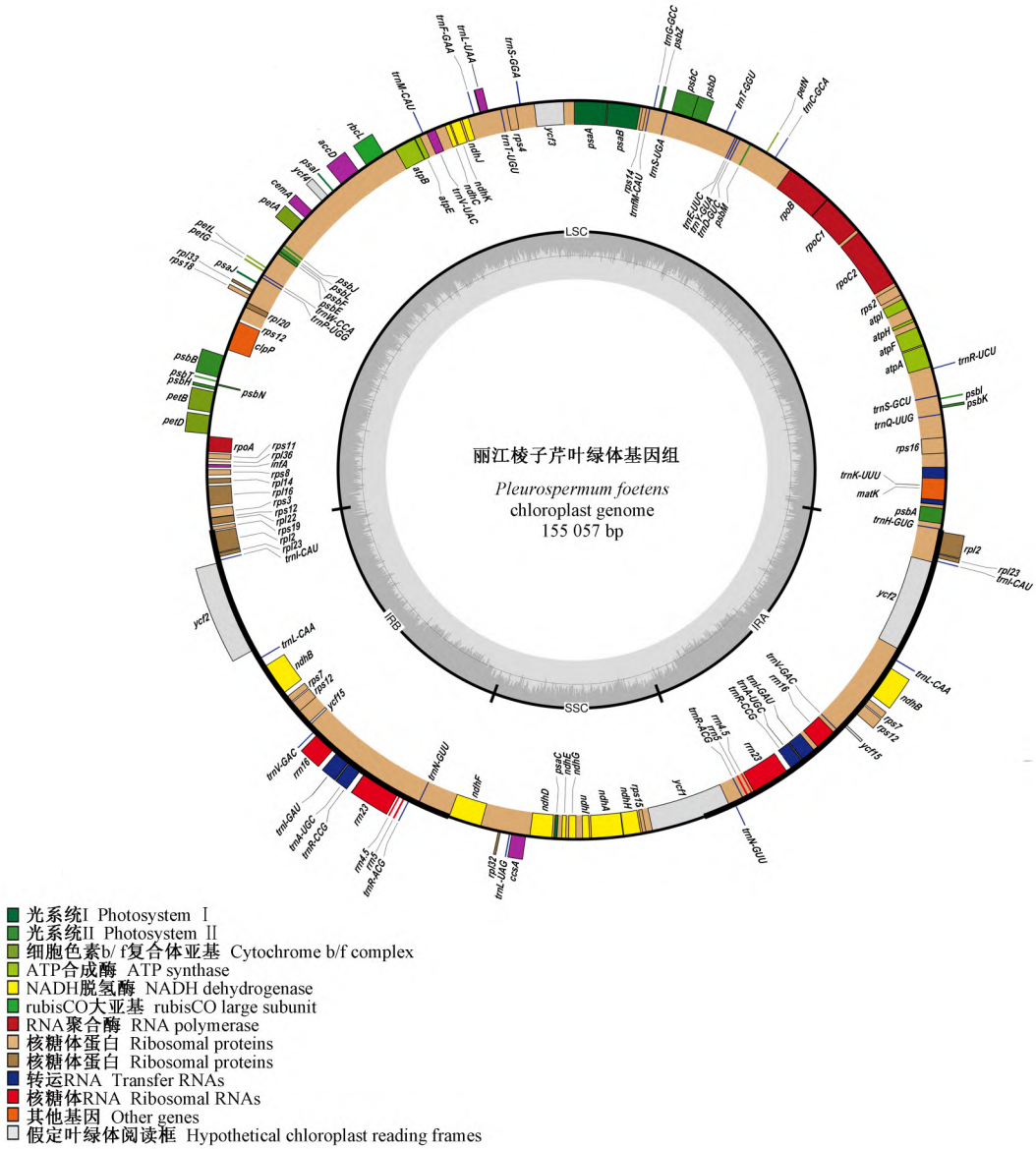
丽江棱子芹( *Pleurospermum foetens*) 隶属伞形科( Apiaceae) 芹亚科( Apioideae) 棱子芹属( *Pleurospermum*) , 为多年生草本植物, 株高 10~30 cm, 茎、叶和花序常带紫色, 有奇臭味( She et al., 2005) , 因此也称臭棱子芹。丽江棱子芹是中国特有高山植物, 主要分布于横断山区域与云南高原, 生境多为海拔 4 000 m 以上的高山流石滩, 是高山特殊冰缘生境植物之一。横断山区域作为生物多样性研究热点区域之一, 近年来被划入了“天空岛”范围( He and Jiang, 2014)。“天空岛”因海拔、生境与山体之间的距离等原因使“岛”上生长的物种产生地理隔离, 限制了世代中个体或种群的扩散与传播, 使不同的“岛屿”间物种交流减少, 易产生遗传漂变, 因此“天空岛”是很好的研究物种遗传机制及扩散历史的区域。丽江棱子芹作为横断山区域的典型冰缘带植物, 目前仅对其细胞分类学、花粉形态与核基因等进行了研究( 张勇等, 2013; 王会朋等, 2014; 孙文远等,

2020) , 还未对其叶绿体基因进行过具体描述与分析。丽江棱子芹所属的棱子芹属的分类一直受到争议, 但这些分类研究大多依赖于 ITS、叶绿体片段与形态学, 很少基于叶绿体基因组对其属下物种及近缘种开展系统发育研究。因此本研究采用高通量测序方法, 通过组装和注释对丽江棱子芹叶绿体基因组的组成、结构及系统发育等进行分析, 探究丽江棱子芹的叶绿体基因组结构, 可为丽江棱子芹的谱系地理学、棱子芹属及其近缘属的系统发育与亲缘关系提供理论基础。

## 1 结果

### 1.1 丽江棱子芹叶绿体基因组结构与特征

丽江棱子芹叶绿体基因组结构为典型的四分体结构( 图 1) , 其全长 155 057 bp, GC 总含量为 38.0%, 碱基 A 占 30.7%, 碱基 C 占 19.3%, 碱基 G 占 18.7%, 碱基 T 占 31.3%。其中, LSC 区长 85 579 bp, SSC 区长 17 418 bp, 二者被两段长度为 26 030 bp 的反向重复序列 IRa 和 IRb 分隔。IR 区的 GC 含量为 42.9%, 明显高于 LSC 区( 36.2%) 与 SSC 区的( 32.4%)。丽江棱子芹叶绿体基因组包含 87 个蛋白编码基因( protein coding genes, PCGs)、8 个核糖体 RNA( rRNA) 基因和 39 个转运 RNA( tRNA) 基因。87 个蛋白编码基因被分为三个大类。第一类是光合作用基因, 第二类是自我复制基因, 第三类其他基因和未知功能基因( 表 1)。其中核糖体大亚基基因 *rpl23* 和 *rpl2*, 核糖体小亚基基因 *rps12* 和 *rps7*, 转运 RNA 基因 *trnA-UGC*、*trnI-CAU*、*trnI-GAU*、*trnL-CAA*、*trnN-GUU*、*trnR-ACG*、*trnR-CCG* 和 *trnV-GAC*, NADH 脱氢酶亚基基因 *ndhB*, 未知功能基因 *ycf15*、*ycf2* 和 4 个核糖体 RNA 基因均有 2 次拷贝, 其余基因为单拷贝基因。



根据基因功能进行颜色编码。  
Color coding based on gene function.

图 1 丽江梭子芹叶绿体基因组图谱  
Figure 1 Gene map of chloroplast genome of *Pleurospermum foetens*

表 1 丽江梭子芹叶绿体基因组注释基因信息

| 功能基因分类<br>Functional gene classification | 基因种类<br>Category of genes                            | 基因名称<br>Gene name                                                                                                    |
|------------------------------------------|------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| 光合作用基因<br>Genes for photosynthesis       | ATP 合成酶亚基<br>Subunits of ATP synthase                | <i>atpA , atpB , atpE , atpF , atpH , atpI</i>                                                                       |
|                                          | 光系统 II 亚基<br>Subunits of photosystem II              | <i>psbA , psbB , psbC , psbD , psbE , psbF , psbH , psbI , psbJ , psbK , psbL , psbM , psbN , psbT , psbZ , ycf3</i> |
|                                          | NADH 脱氢酶亚基<br>Subunits of NADH-dehydrogenase         | <i>ndhA , ndhB<sup>*</sup> , ndhC , ndhD , ndhE , ndhF , ndhG , ndhH , ndhI , ndhJ , ndhK</i>                        |
|                                          | 细胞色素 b/f 复合体亚基<br>Subunits of cytochrome b/f complex | <i>petA , petB , petD , petG , petL , petN</i>                                                                       |
|                                          | 光系统 I 亚基<br>Subunits of photosystem I                | <i>psaA , psaB , psaC , psaI , psaJ</i>                                                                              |
|                                          | 二磷酸核酮糖羧化酶大亚基<br>Subunit of rubisco                   | <i>rbcL</i>                                                                                                          |

续表  
Continuing table

| 功能基因分类<br>Functional gene classification | 基因种类<br>Category of genes                          | 基因名称<br>Gene name                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------------------------|----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 自我复制基因<br>Self replication genes         | 核糖体 RNA 基因<br>Ribosomal RNA genes                  | <i>rrn16*</i> , <i>rrn23*</i> , <i>rrn4.5*</i> , <i>rrn5*</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|                                          | 核糖体大亚基<br>Large subunit of ribosome                | <i>rpl14</i> , <i>rpl16</i> , <i>rpl2*</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23*</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|                                          | DNA 依赖性 RNA 聚合酶<br>DNA dependent RNA polymerase    | <i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|                                          | 核糖体小亚基<br>Small subunit of ribosome                | <i>rps11</i> , <i>rps12*</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> , <i>rps18</i> , <i>rps19</i> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7*</i> , <i>rps8</i>                                                                                                                                                                                                                                                                                                                                                                                                       |
|                                          | 转运 RNA 基因<br>Transfer RNA genes                    | <i>trnA-UGC*</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-M-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <i>trnI-CAU*</i> , <i>trnI-GAU*</i> , <i>trnK-UUU</i> , <i>trnL-CAA*</i> , <i>trnL-UAA</i> , <i>trn-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUU*</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG*</i> , <i>trnR-CCG*</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC*</i> , <i>trnV-UAC</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i> |
|                                          | 其他基因<br>Other genes                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|                                          | 乙酰辅酶 A 羧化酶亚基<br>Subunit of acetyl CoA carboxylase  | <i>accD</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 未知功能基因<br>Unknown function genes         | c 型细胞色素合成基因<br>c type cytochrom synthesis gene     | <i>ccsA</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|                                          | 包膜蛋白<br>Envelop membrane protein                   | <i>cemA</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|                                          | ATP 依赖的蛋白酶亚基 P<br>ATP dependent protease subunit P | <i>clpP</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|                                          | 翻译起始因子<br>Translational initiation factor          | <i>infA</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|                                          | 成熟酶<br>Maturase                                    | <i>matK</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|                                          | 开放阅读框<br>Open reading frames                       | <i>ycf1</i> , <i>ycf2*</i> , <i>ycf4</i> , <i>ycf15*</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|                                          |                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

注: \* 为两次拷贝。  
Note: \* is two copies.

1.2 丽江梭子芹叶绿体基因组的密码子偏好性

密码子是蛋白质合成过程中起着关键作用的三联体碱基。在丽江梭子芹叶绿体基因组的分析中, 87 个蛋白编码基因共编码 26 075 个密码子, 在这些

密码子中, 编码异亮氨酸 Ile 的 ATT 数量最多, 半胱氨酸 Cys 的 TGC 数量是除终止子外编码最少的; 3 个终止密码子中 TAA 在蛋白编码基因中使用最多 (TAA: 50.6%; TAG: 26.4%; TGA: 23.0%) (表 2)。

表 2 丽江梭子芹叶绿体全基因组的密码子使用频率  
Table 2 Codons usage frequency of the complete chloroplast genome of *Pleurospermum foetens*

| 密码子<br>Codon | 氨基酸<br>Amino acid | 该氨基酸内<br>占比/%<br>The internal frequency<br>of this amino acid/% | 总氨基酸<br>中频率<br>Frequency in<br>all amino acid | 数量<br>Number | 密码子<br>Codon | 氨基酸<br>Amino acid | 该氨基酸内<br>占比/%<br>The internal frequency<br>of this amino acid/% | 总氨基酸<br>中频率<br>Frequency in<br>all amino acid | 数量<br>Number |
|--------------|-------------------|-----------------------------------------------------------------|-----------------------------------------------|--------------|--------------|-------------------|-----------------------------------------------------------------|-----------------------------------------------|--------------|
| TGA          | *                 | 23.0                                                            | 0.767                                         | 20           | CGA          | Arg               | 23.0                                                            | 14.113                                        | 368          |
| TAG          | *                 | 26.4                                                            | 0.882                                         | 23           | CAT          | Leu               | 13.7                                                            | 14.382                                        | 375          |
| TAA          | *                 | 50.6                                                            | 1.687                                         | 44           | TCA          | Ser               | 19.2                                                            | 14.765                                        | 385          |
| TGC          | Cys               | 24.4                                                            | 2.646                                         | 69           | AGT          | Ser               | 19.4                                                            | 14.919                                        | 389          |

续表  
Continuing table

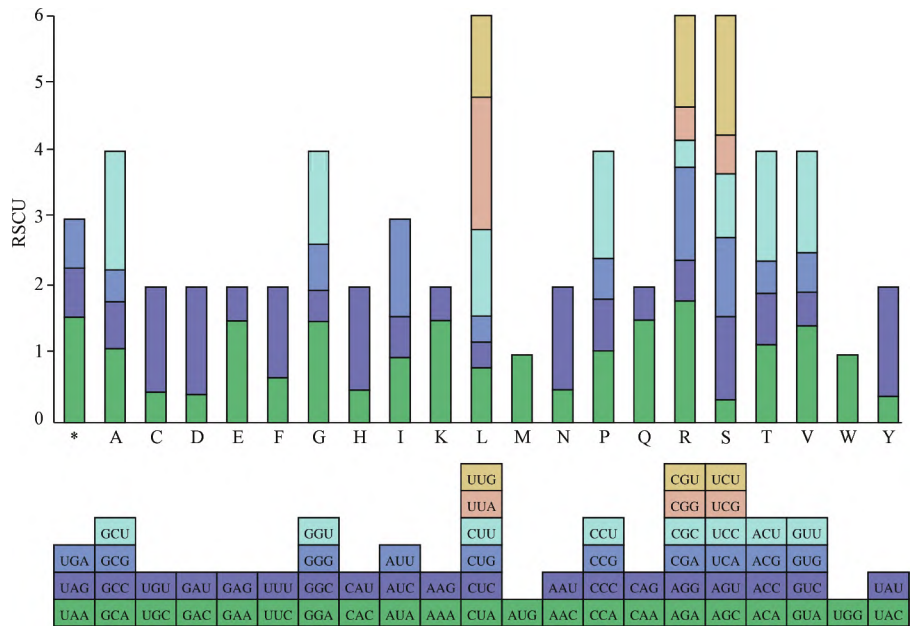
| 密码子<br>Codon | 氨基酸<br>Amino acid | 该氨基酸内<br>占比/%<br>The internal frequency<br>of this amino acid/% | 总氨基酸<br>中频率<br>Frequency in<br>all amino acid | 数量<br>Number | 密码子<br>Codon | 氨基酸<br>Amino acid | 该氨基酸内<br>占比/%<br>The internal frequency<br>of this amino acid/% | 总氨基酸<br>中频率<br>Frequency in<br>all amino acid | 数量<br>Number |
|--------------|-------------------|-----------------------------------------------------------------|-----------------------------------------------|--------------|--------------|-------------------|-----------------------------------------------------------------|-----------------------------------------------|--------------|
| CGC          | Arg               | 6.4                                                             | 3.912                                         | 102          | GCA          | Ala               | 27.8                                                            | 15.034                                        | 392          |
| AGC          | Ser               | 5.9                                                             | 4.564                                         | 119          | ACA          | Thr               | 29.2                                                            | 15.149                                        | 395          |
| CGG          | Arg               | 8.4                                                             | 5.177                                         | 135          | CCT          | Pro               | 38.5                                                            | 16.453                                        | 429          |
| CAC          | His               | 24.7                                                            | 5.906                                         | 154          | ATC          | Ile               | 20.7                                                            | 17.335                                        | 452          |
| GCG          | Ala               | 11.5                                                            | 6.213                                         | 162          | TGG          | Trp               | 100.0                                                           | 17.45                                         | 455          |
| AGG          | Arg               | 10.2                                                            | 6.29                                          | 164          | CAT          | His               | 75.3                                                            | 17.987                                        | 469          |
| CCG          | Pro               | 14.8                                                            | 6.328                                         | 165          | AGA          | Arg               | 30.5                                                            | 18.725                                        | 488          |
| ACG          | Thr               | 12.3                                                            | 6.405                                         | 167          | GTA          | Val               | 35.5                                                            | 19.367                                        | 505          |
| GTC          | Val               | 12.8                                                            | 6.98                                          | 182          | TTC          | Phe               | 35.2                                                            | 19.981                                        | 521          |
| CTG          | Leu               | 6.7                                                             | 7.018                                         | 183          | GTT          | Val               | 37.1                                                            | 20.288                                        | 529          |
| TAC          | Tyr               | 19.4                                                            | 7.21                                          | 188          | ACT          | Thr               | 39.3                                                            | 20.403                                        | 532          |
| CTC          | Leu               | 7.0                                                             | 7.325                                         | 191          | TTG          | Leu               | 20.1                                                            | 21.131                                        | 551          |
| GGC          | Gly               | 10.9                                                            | 7.517                                         | 196          | TCT          | Ser               | 29.0                                                            | 22.282                                        | 581          |
| TCG          | Ser               | 9.8                                                             | 7.555                                         | 197          | CTT          | Leu               | 21.8                                                            | 22.857                                        | 596          |
| GTG          | Val               | 14.6                                                            | 7.977                                         | 208          | GCT          | Gly               | 33.4                                                            | 23.011                                        | 600          |
| TGT          | Cys               | 75.6                                                            | 8.207                                         | 214          | GCT          | Ala               | 42.8                                                            | 23.164                                        | 604          |
| GAC          | Asp               | 20.5                                                            | 8.207                                         | 214          | ATG          | Met               | 100.0                                                           | 23.241                                        | 606          |
| CCC          | Pro               | 19.4                                                            | 8.284                                         | 216          | GGA          | Gly               | 38.3                                                            | 26.424                                        | 689          |
| CAG          | Gln               | 25.1                                                            | 9.012                                         | 235          | CAA          | Gln               | 74.9                                                            | 26.846                                        | 700          |
| GCC          | Ala               | 17.9                                                            | 9.664                                         | 252          | ATA          | Ile               | 32.3                                                            | 27.076                                        | 706          |
| ACC          | Thr               | 19.1                                                            | 9.933                                         | 259          | TAT          | Tyr               | 80.6                                                            | 30.029                                        | 783          |
| CCA          | Pro               | 27.2                                                            | 11.62                                         | 303          | GAT          | Asp               | 79.5                                                            | 31.908                                        | 832          |
| AAC          | Asn               | 24.7                                                            | 11.735                                        | 306          | TTA          | Leu               | 30.8                                                            | 32.368                                        | 844          |
| GGG          | Gly               | 17.4                                                            | 12.004                                        | 313          | AAT          | Asn               | 75.3                                                            | 35.743                                        | 932          |
| TCC          | Ser               | 16.5                                                            | 12.694                                        | 331          | TTT          | Phe               | 64.8                                                            | 36.855                                        | 961          |
| CGT          | Arg               | 21.5                                                            | 13.193                                        | 344          | GAA          | Glu               | 74.1                                                            | 37.852                                        | 987          |
| GAG          | Glu               | 25.9                                                            | 13.231                                        | 345          | AAA          | Lys               | 73.7                                                            | 39.271                                        | 1 024        |
| AAG          | Lys               | 26.3                                                            | 13.998                                        | 365          | ATT          | Ile               | 47.1                                                            | 39.463                                        | 1 029        |

注: \* 表示终止密码子。

Note: \* represents stop codon.

由于密码子的简并性, 每种蛋白质对密码子的选择不平等, 因此, 密码子的选择具有偏好, 这种情况叫作密码子的偏好性( codon usage bias) ( Hershberg and Petrov , 2008) 。如果密码子使用没有偏好, 则该密码子的相对同义密码子使用度( relative synonymous codon usage , RSCU) 等于 1。当某一密码子

的 RSCU 值大于 1, 则表明其使用频率相对较高( 赵耀 等, 2008) 。丽江棱子芹叶绿体基因组各氨基酸的 RSCU 分析中, RSCU 值最高的为 UUA, 最低的是 AGC( 图 2) 。在 RSCU 大于 1 的 30 个氨基酸中, 除 Leu( UUG) 对应密码子是 C/G 结尾外, 其他均以 A/ U 结尾( 表 3) 。



横坐标表示 20 个氨基酸和终止子,纵坐标表示 RSCU 值; 每个氨基酸的不同颜色对应下面相同颜色的密码子; A 表示丙氨酸( Ala) , C 表示半胱氨酸( Cys) , D 表示天冬氨酸( Asp) , E 表示谷氨酸( Glu) , F 表示苯丙氨酸( Phe) , G 表示甘氨酸( Gly) , H 表示组氨酸( His) , I 表示异亮氨酸( Ile) , K 表示赖氨酸( Lys) , L 表示亮氨酸( Leu) , M 表示甲硫氨酸( Met) , N 表示天冬酰胺( Asn) , P 表示脯氨酸( Pro) , Q 表示谷氨酰胺( Gln) , R 表示精氨酸( Arg) , S 表示丝氨酸( Ser) , T 表示苏氨酸( Thr) , V 表示缬氨酸( Val) , W 表示色氨酸( Trp) , Y 表示酪氨酸( Tyr) ; \* 表示终止密码子。

The abscissa represents 20 amino acids and terminators , while the ordinate represents the RSCU value; The different colors of each amino acid correspond to the codon of the same color below; A represents alanine ( Ala) , C represents cysteine ( Cys) , D represents asparticacid ( Asp) , E represents glutamicacid ( Glu) , F represents phenylalanine ( Phe) , G represents glycine ( Gly) , H represents histidine ( His) , I represents isoleucine ( Ile) , K represents lysine ( Lys) , L represents leucine ( Leu) , M represents methionine ( Met) , N represents asparagine ( Asn) , P represents proline ( Pro) , Q represents glutamine ( Gln) , R represents arginine ( Arg) , S represents serine ( Ser) , T represents threonine ( Thr) , V represents valine ( Val) , W represents tryptophan ( Trp) , Y represents tyrosine ( Tyr) ; \* represents stop codon.

图 2 丽江棱子芹基因组各氨基酸的 RSCU 分析

Figure 2 RSCU analysis of each amino acid in the chloroplast genome of *Pleurospermum foetens*

表 3 丽江棱子芹叶绿体全基因组的 RSCU 统计

Table 3 RSCU statistics of the complete chloroplast genome of *Pleurospermum foetens*

| 氨基酸缩写<br>Amino acid<br>abbreviation | 氨基酸<br>Amino acid | 密码子<br>Codon | RSCU    | 氨基酸缩写<br>Amino acid<br>abbreviation | 氨基酸<br>Amino acid | 密码子<br>Codon | RSCU    |
|-------------------------------------|-------------------|--------------|---------|-------------------------------------|-------------------|--------------|---------|
| S                                   | Ser               | AGC          | 0.349 0 | M                                   | Met               | AUG          | 1.000 0 |
| Y                                   | Tyr               | UAC          | 0.383 6 | W                                   | Trp               | UGG          | 1.000 0 |
| L                                   | Leu               | CUC          | 0.387 3 | P                                   | Pro               | CCA          | 1.059 2 |
| L                                   | Leu               | CUG          | 0.389 8 | A                                   | Ala               | GCA          | 1.093 2 |
| R                                   | Arg               | CGC          | 0.402 6 | S                                   | Ser               | UCA          | 1.153 7 |

续表  
Continuing table

| 氨基酸缩写<br>Amino acid<br>abbreviation | 氨基酸<br>Amino acid | 密码子<br>Codon | RSCU    | 氨基酸缩写<br>Amino acid<br>abbreviation | 氨基酸<br>Amino acid | 密码子<br>Codon | RSCU    |
|-------------------------------------|-------------------|--------------|---------|-------------------------------------|-------------------|--------------|---------|
| D                                   | Asp               | GAC          | 0.413 9 | T                                   | Thr               | ACA          | 1.162 6 |
| C                                   | Cys               | UGC          | 0.454 5 | L                                   | Leu               | UUG          | 1.197 0 |
| G                                   | Gly               | GGC          | 0.456 6 | S                                   | Ser               | AGU          | 1.214 2 |
| T                                   | Thr               | ACG          | 0.471 8 | L                                   | Leu               | CUU          | 1.290 0 |
| A                                   | Ala               | GCG          | 0.475 2 | R                                   | Arg               | CGU          | 1.330 4 |
| Q                                   | Gln               | CAG          | 0.484 1 | F                                   | Phe               | UUU          | 1.334 4 |
| H                                   | His               | CAC          | 0.485 0 | G                                   | Gly               | GGU          | 1.367 4 |
| R                                   | Arg               | CGG          | 0.490 2 | R                                   | Arg               | CGA          | 1.391 7 |
| N                                   | Asn               | AAC          | 0.491 1 | V                                   | Val               | GUA          | 1.425 9 |
| V                                   | Val               | GUC          | 0.493 0 | I                                   | Ile               | AUU          | 1.427 8 |
| E                                   | Glu               | GAG          | 0.496 6 | G                                   | Gly               | GGA          | 1.494 7 |
| K                                   | Lys               | AAG          | 0.501 7 | K                                   | Lys               | AAA          | 1.498 3 |
| V                                   | Val               | GUG          | 0.571 0 | *                                   | Ter               | UAA          | 1.500 0 |
| S                                   | Ser               | UCG          | 0.576 9 | E                                   | Glu               | GAA          | 1.503 4 |
| R                                   | Arg               | AGG          | 0.595 2 | N                                   | Asn               | AAU          | 1.508 9 |
| P                                   | Pro               | CCG          | 0.598 1 | V                                   | Val               | GUU          | 1.510 1 |
| I                                   | Ile               | AUC          | 0.605 6 | H                                   | His               | CAU          | 1.515 0 |
| F                                   | Phe               | UUC          | 0.665 6 | Q                                   | Gln               | CAA          | 1.515 9 |
| G                                   | Gly               | GGG          | 0.681 2 | C                                   | Cys               | UGU          | 1.545 5 |
| A                                   | Ala               | GCC          | 0.698 8 | P                                   | Pro               | CCU          | 1.578 4 |
| *                                   | Ter               | UGA          | 0.712 5 | D                                   | Asp               | GAU          | 1.586 1 |
| P                                   | Pro               | CCC          | 0.764 3 | T                                   | Thr               | ACU          | 1.597 3 |
| T                                   | Thr               | ACC          | 0.768 3 | Y                                   | Tyr               | UAU          | 1.616 4 |
| *                                   | Ter               | UAG          | 0.787 5 | A                                   | Ala               | GCU          | 1.732 9 |
| L                                   | Leu               | CUA          | 0.797 2 | S                                   | Ser               | UCU          | 1.766 2 |
| S                                   | Ser               | UCC          | 0.940 1 | R                                   | Arg               | AGA          | 1.789 9 |
| I                                   | Ile               | AUA          | 0.966 5 | L                                   | Leu               | UUA          | 1.938 8 |

注: \* 表示终止密码子。  
Note: \* represents stop codon.

1.3 丽江棱子芹叶绿体基因组的重复序列与 SSR 分析

重复序列可分为串联重复序列(tandam repeat sequence)和散在重复序列(interspersed repeat sequence)两类。通过 TRF 软件预测到的串联重复序列共 23 个,长 1 055 bp,其中长度最小的为 24 bp,

最大的为 97 bp,31~40 bp 的重复序列最多,有 6 个(表 4)。利用 REPuter 软件进行散在重复序列分析(表 5),在丽江棱子芹叶绿体基因组中发现了 17 个散在重复,全部是正向重复序列(forward repeat ,F),未检测到其他类型重复序列。

表 4 丽江棱子芹叶绿体全基因组的串联重复序列  
Table 4 Tandem repeat sequence across the complete chloroplast genome of *Pleurospermum foetens*

| 序号<br>Order number | 重复单元长度/bp<br>Repeating units length/bp | 重复单元数量<br>Repeating units number | 重复序列大小/bp<br>Repeating sequence size/bp | 所在位置<br>Location                                  |
|--------------------|----------------------------------------|----------------------------------|-----------------------------------------|---------------------------------------------------|
| 1                  | 17                                     | 2.5                              | 41                                      | IGS                                               |
| 2                  | 20                                     | 2.0                              | 39                                      | IGS                                               |
| 3                  | 20                                     | 2.0                              | 39                                      | IGS                                               |
| 4                  | 49                                     | 2.0                              | 97                                      | <i>rpoC2</i> , IGS( <i>rpoC2</i> 、 <i>rpoC1</i> ) |
| 5                  | 17                                     | 2.1                              | 34                                      | IGS                                               |
| 6                  | 12                                     | 2.2                              | 25                                      | IGS                                               |

续表

Continuing table

| 序号<br>Order number | 重复单元长度/bp<br>Repeating units length/bp | 重复单元数量<br>Repeating units number | 重复序列大小/bp<br>Repeating sequence size/bp | 所在位置<br>Location                               |
|--------------------|----------------------------------------|----------------------------------|-----------------------------------------|------------------------------------------------|
| 7                  | 13                                     | 2.0                              | 25                                      | <i>trnL-UAA</i>                                |
| 8                  | 17                                     | 2.0                              | 33                                      | IGS                                            |
| 9                  | 24                                     | 1.9                              | 45                                      | IGS                                            |
| 10                 | 21                                     | 2.0                              | 40                                      | <i>rps18</i>                                   |
| 11                 | 19                                     | 2.0                              | 37                                      | <i>psbT</i> , IGS( <i>psbT</i> 、 <i>psbN</i> ) |
| 12                 | 15                                     | 2.1                              | 30                                      | <i>petB</i> ( intron)                          |
| 13                 | 24                                     | 2.1                              | 49                                      | IGS                                            |
| 14                 | 21                                     | 3.2                              | 67                                      | <i>ycf2</i>                                    |
| 15                 | 15                                     | 3.0                              | 44                                      | <i>ycf2</i>                                    |
| 16                 | 9                                      | 5.9                              | 52                                      | <i>ycf2</i>                                    |
| 17                 | 18                                     | 3.3                              | 58                                      | <i>ycf2</i>                                    |
| 18                 | 12                                     | 2.1                              | 24                                      | IGS                                            |
| 19                 | 18                                     | 3.3                              | 58                                      | <i>ycf2</i>                                    |
| 20                 | 9                                      | 5.9                              | 52                                      | <i>ycf2</i>                                    |
| 21                 | 15                                     | 3.0                              | 44                                      | <i>ycf2</i>                                    |
| 22                 | 21                                     | 2.6                              | 53                                      | <i>ycf2</i>                                    |
| 23                 | 19                                     | 3.4                              | 69                                      | <i>ycf2</i>                                    |

注: 括号中的 IGS 表示内部指导序列 , intron 表示内含子区域。

Note: IGS in parentheses indicates the internal guidance sequence , and intron indicates the intron region.

表 5 丽江梭子芹叶绿体基因组散在重复序列

Table 5 Interspersed repeat sequences in the complete chloroplast genome of *Pleurospermum foetens*

| 重复起始点 1<br>Repeat start 1 | 重复类型<br>Repeat type | 大小/bp<br>Size/bp | 重复起始点 2<br>Repeat start 2 | E 值<br>E value | 所在基因名称<br>Gene name of location                                                | 基因区域<br>Region of gene |
|---------------------------|---------------------|------------------|---------------------------|----------------|--------------------------------------------------------------------------------|------------------------|
| 90 994                    | F                   | 30               | 91 009                    | 5. 28E-07      | <i>ycf2</i>                                                                    | IRb                    |
| 149 597                   | F                   | 30               | 149 612                   | 5. 28E-07      | <i>ycf2</i>                                                                    | IRa                    |
| 8 989                     | F                   | 30               | 36 623                    | 2. 30E-05      | <i>psbI</i> , <i>trnS-GCU</i> ( IGS) 、<br><i>psbC</i> , <i>trnS-UGA</i> ( IGS) | IRb , LSC              |
| 39 846                    | F                   | 30               | 42 070                    | 6. 43E-04      | <i>psaB</i> 、 <i>psaA</i>                                                      | LSC                    |
| 89 816                    | F                   | 30               | 89 858                    | 6. 43E-04      | <i>ycf2</i>                                                                    | IRb                    |
| 92 283                    | F                   | 30               | 92 301                    | 6. 43E-04      | <i>ycf2</i>                                                                    | IRb                    |
| 150 748                   | F                   | 30               | 150 790                   | 6. 43E-04      | <i>ycf2</i>                                                                    | IRa                    |
| 85 304                    | F                   | 31               | 85 328                    | 1. 78E-04      | IGS                                                                            | IRb                    |
| 92 261                    | F                   | 31               | 92 297                    | 1. 78E-04      | <i>ycf2</i>                                                                    | IRb                    |
| 108 474                   | F                   | 31               | 132 131                   | 1. 78E-04      | IGS                                                                            | IRb , IRa              |
| 148 312                   | F                   | 31               | 148 348                   | 1. 78E-04      | <i>ycf2</i>                                                                    | IRa                    |
| 21 317                    | F                   | 36               | 21 366                    | 2. 76E-07      | <i>rpoC2</i> 、 <i>rpoC1</i> ( IGS)                                             | LSC                    |
| 44 841                    | F                   | 39               | 99 301                    | 2. 62E-12      | <i>ycf3</i> ( intron)<br><i>rps12</i> ( intron)                                | LSC , IRb              |
| 44 841                    | F                   | 39               | 121 872                   | 2. 62E-12      | <i>trnL-UAA</i> ( intron)<br><i>ndhA</i> ( intron)                             | SSC , LSC              |
| 99 299                    | F                   | 42               | 121 870                   | 3. 50E-16      | <i>rps12</i> ( intron)<br><i>ndhA</i> ( intron)                                | IRb , SSC              |
| 92 257                    | F                   | 49               | 92 275                    | 1. 06E-14      | <i>ycf2</i>                                                                    | IRb                    |
| 148 308                   | F                   | 49               | 148 326                   | 1. 06E-14      | <i>ycf2</i>                                                                    | IRa                    |

简单重复序列( simple sequence repeat , SSR) 分析( 表 6) 共检测到 SSR 位点 34 个 , 其中单核 苷酸重复序列 27 个 , 双核苷酸重复序列 4 个 , 复合型重复序列 3 个 , 未检测到三 / 四 / 五核苷酸序

列。检测结果中 88.2% 的 SSR 位于内部指导序列( internal guide sequence , IGS) 和内含子( intron) 等非编码区域, 且 A/T 碱基在 SSR 中出现频率较高, 在 27 个单核苷酸 SSR 中, 包含 A 碱基的有 12 个, 包含 T 碱基的有 13 个。

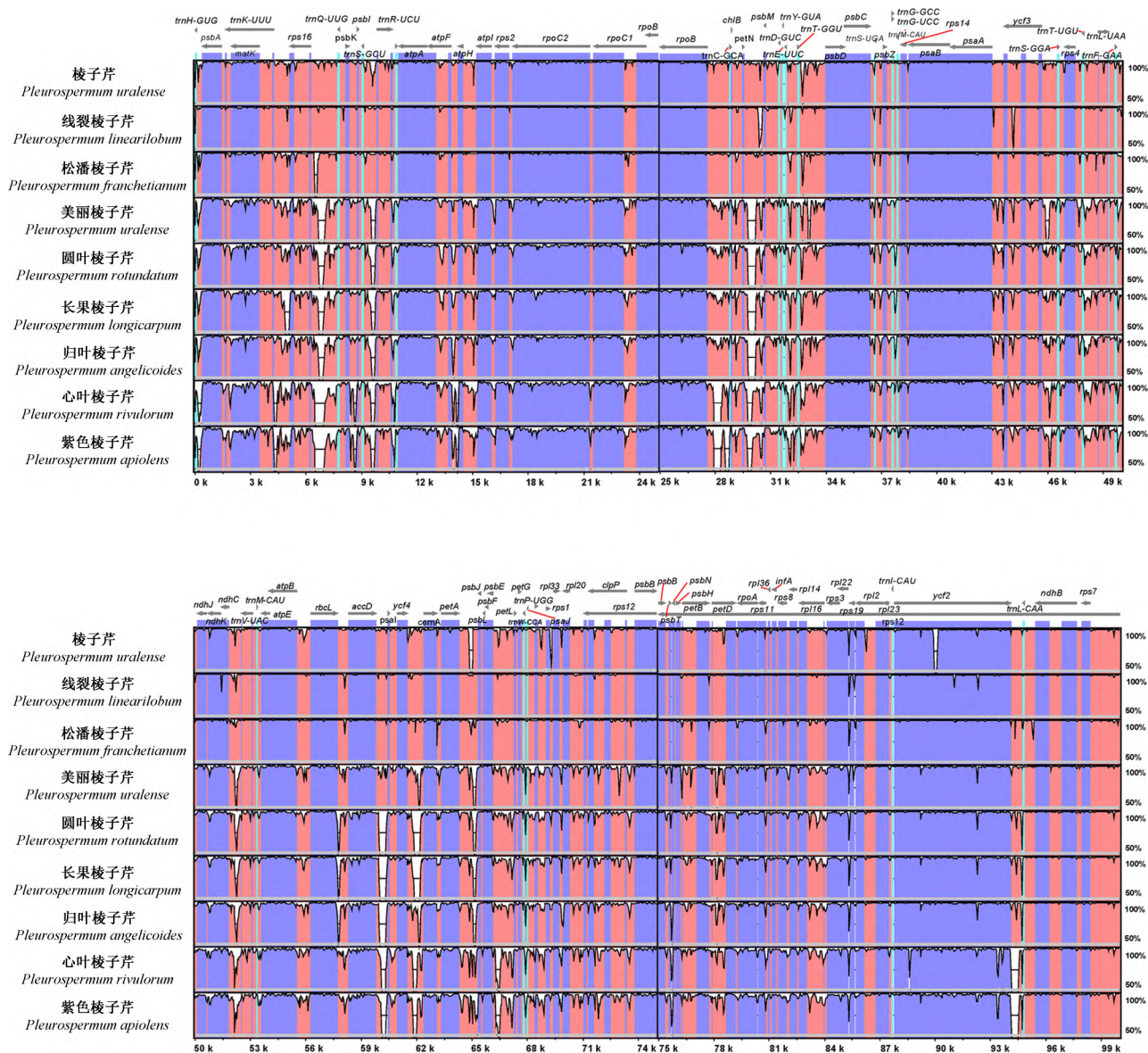
表 6 丽江棱子芹叶绿体全基因组 SSR 信息

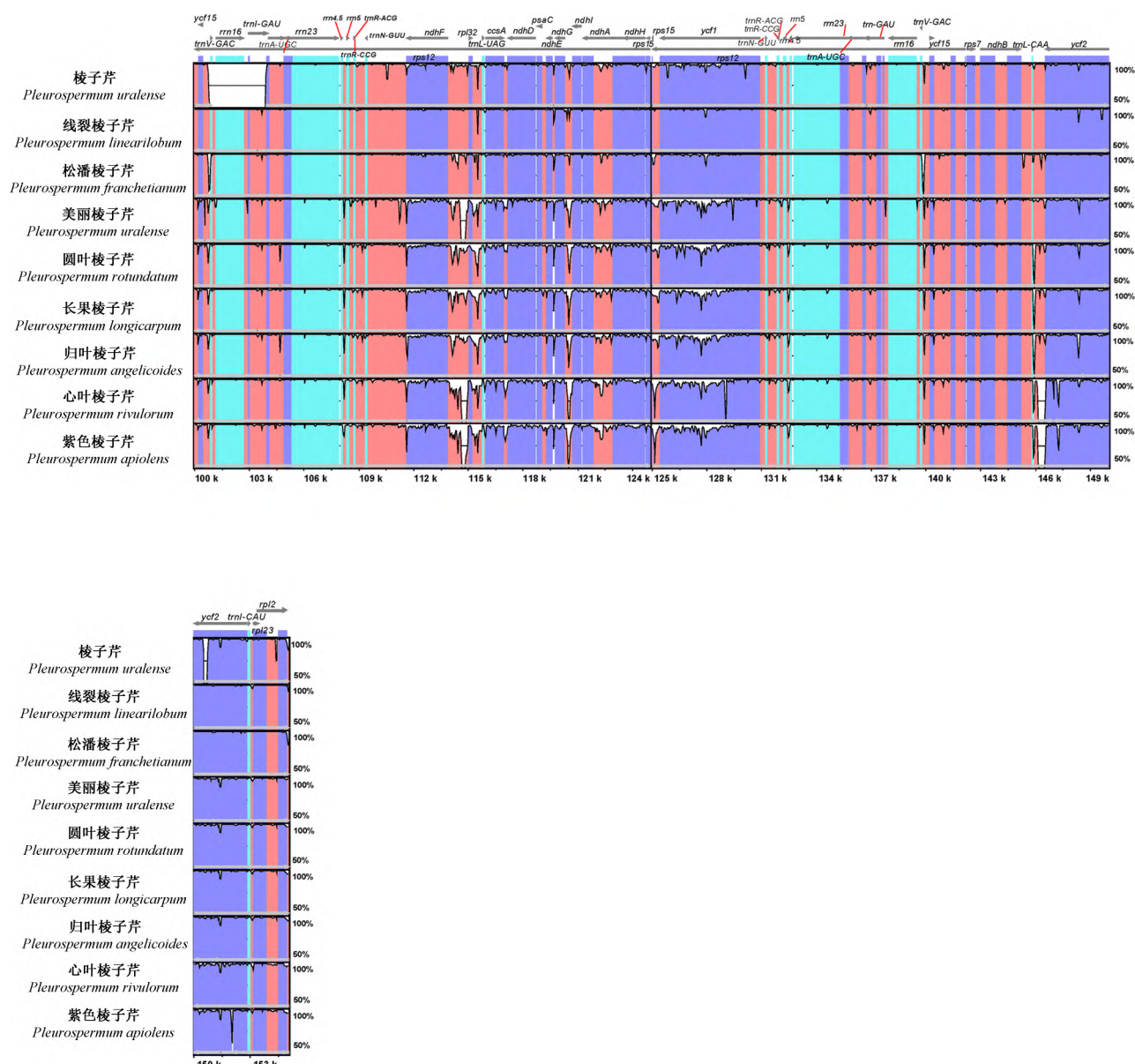
Table 6 Information of SSR in the complete chloroplast genome of *Pleurospermum foetens*

| 序号<br>Order number | 重复类型<br>SSR type | 简单重复序列<br>SSR                                                             | 大小<br>Size | 起始<br>Start | 终止<br>End | 位置<br>Location |
|--------------------|------------------|---------------------------------------------------------------------------|------------|-------------|-----------|----------------|
| 1                  | p1               | ( A) 11                                                                   | 11         | 6 649       | 6 659     | IGS            |
| 2                  | p1               | ( A) 10                                                                   | 10         | 6 962       | 6 971     | IGS            |
| 3                  | p1               | ( A) 11                                                                   | 11         | 7 592       | 7 602     | IGS            |
| 4                  | c                | ( T) 10a( T) 12                                                           | 23         | 9 494       | 9 516     | IGS            |
| 5                  | p1               | ( T) 10                                                                   | 10         | 13 255      | 13 264    | intron         |
| 6                  | p1               | ( A) 13                                                                   | 13         | 13 867      | 13 879    | IGS            |
| 7                  | p1               | ( T) 10                                                                   | 10         | 15 055      | 15 064    | IGS            |
| 8                  | c                | ( T) 12c( A) 12                                                           | 25         | 16 954      | 16 978    | <i>rps2</i>    |
| 9                  | p1               | ( T) 14                                                                   | 14         | 19 177      | 19 190    | <i>rpoC2</i>   |
| 10                 | p1               | ( A) 14                                                                   | 14         | 23 262      | 23 275    | intron         |
| 11                 | p1               | ( T) 12                                                                   | 12         | 26 911      | 26 922    | IGS            |
| 12                 | p1               | ( T) 10                                                                   | 10         | 33 285      | 33 294    | IGS            |
| 13                 | p1               | ( A) 11                                                                   | 11         | 33 486      | 33 496    | IGS            |
| 14                 | p2               | ( AT) 6                                                                   | 12         | 36 937      | 36 948    | IGS            |
| 15                 | c                | ( T) 10caaaattttttatctatcccgatcccaaatcatgtgtttctctgaagattgagagcaat( A) 10 | 82         | 45 433      | 45 514    | intron         |
| 16                 | p2               | ( AT) 7                                                                   | 14         | 49 800      | 49 813    | IGS            |
| 17                 | p1               | ( T) 10                                                                   | 10         | 55 447      | 55 456    | <i>atpB</i>    |
| 18                 | p1               | ( A) 13                                                                   | 13         | 56 025      | 56 037    | IGS            |
| 19                 | p1               | ( T) 11                                                                   | 11         | 60 568      | 60 578    | IGS            |
| 20                 | p1               | ( T) 13                                                                   | 13         | 83 564      | 83 576    | intron         |
| 21                 | p1               | ( T) 17                                                                   | 17         | 85 647      | 85 663    | IGS            |
| 22                 | p1               | ( A) 10                                                                   | 10         | 97 671      | 97 680    | IGS            |
| 23                 | p1               | ( T) 13                                                                   | 13         | 99 679      | 99 691    | IGS            |
| 24                 | p1               | ( T) 12                                                                   | 12         | 103 435     | 103 446   | intron         |
| 25                 | p1               | ( G) 10                                                                   | 10         | 104 664     | 104 673   | intron         |
| 26                 | p1               | ( T) 10                                                                   | 10         | 117 120     | 117 129   | IGS            |
| 27                 | p2               | ( TA) 7                                                                   | 14         | 120 287     | 120 300   | IGS            |
| 28                 | p2               | ( AT) 6                                                                   | 12         | 120 531     | 120 542   | IGS            |
| 29                 | p1               | ( A) 11                                                                   | 11         | 128 202     | 128 212   | <i>ycf1</i>    |
| 30                 | p1               | ( C) 10                                                                   | 10         | 135 964     | 135 973   | intron         |
| 31                 | p1               | ( A) 12                                                                   | 12         | 137 191     | 137 202   | intron         |
| 32                 | p1               | ( A) 13                                                                   | 13         | 140 946     | 140 958   | IGS            |
| 33                 | p1               | ( T) 10                                                                   | 10         | 142 957     | 142 966   | IGS            |
| 34                 | p1               | ( A) 17                                                                   | 17         | 154 974     | 154 990   | IGS            |

注: p 表示单个 SSR 类型, 其后的数字 1 和 2 表示构成序列的碱基数; c 表示复合 SSR 类型。

Note: p indicates single SSR type, followed by the numbers 1 and 2 indicating the number of bases forming the sequence; c indicates composite SSR type.





横坐标为序列位置, 纵坐标为相似度; 灰色箭头表示基因转录方向; 蓝色表示外显子; 青色表示内含子; 红色表示非编码区。

The x-axis is the position of sequences, the y-axis is similarity; Gray arrows represents the direction of gene transcription; Blue represents exon; Cyan represents intron; Red represents noncoding sequences.

图3 梭子芹属植物叶绿体全基因组比较

Figure 3 Comparison of the complete chloroplast genomes of *Pleurospermum*

对 10 种梭子芹属植物叶绿体基因组的 IR 区与 LSC 区、SSC 区进行边界分析, 其结果如图 4 显示, 10 种植物叶绿体基因组序列的长度差异较小, 除心叶梭子芹、紫色梭子芹和美丽梭子芹 (*P. amabile*) 的边界基因的位置和类型有较大的差异外, 其余几种基因组结构类似。在边界 LSC/IRb (JLB) 中分布的基因包括 *rps19*、*rpl22* 和 *rpl2*、*ycf2*。除心叶梭子芹、紫色梭子芹的 JLB 边界基因为横跨两区的 *ycf2* 基因, 其余均为 *rps19* 基因。其中归叶梭子芹 (*P. an-*

*gelicoides*) 的 *rps19* 基因完全位于 LSC 区。在边界 IRb/SSC (JSB) 上分布的基因均为 *ndhF*, 但基因的位置与大小有差异, 其中美丽梭子芹的 *ndhF* 基因距离边界较远, 并且基因大小明显小于其他种的。在边界 SSC/IRa (JSA) 上的边界基因均为 *ycf1*, 基因横跨两区, 长度与位置也大致相似。在 IRa/LSC (JLA) 边界上, 心叶梭子芹、紫色梭子芹的边界基因为 *trnL*, 距离上有较大差异。长果梭子芹的边界基因为 *rpl2*, 其余均为 *trnH* 基因。

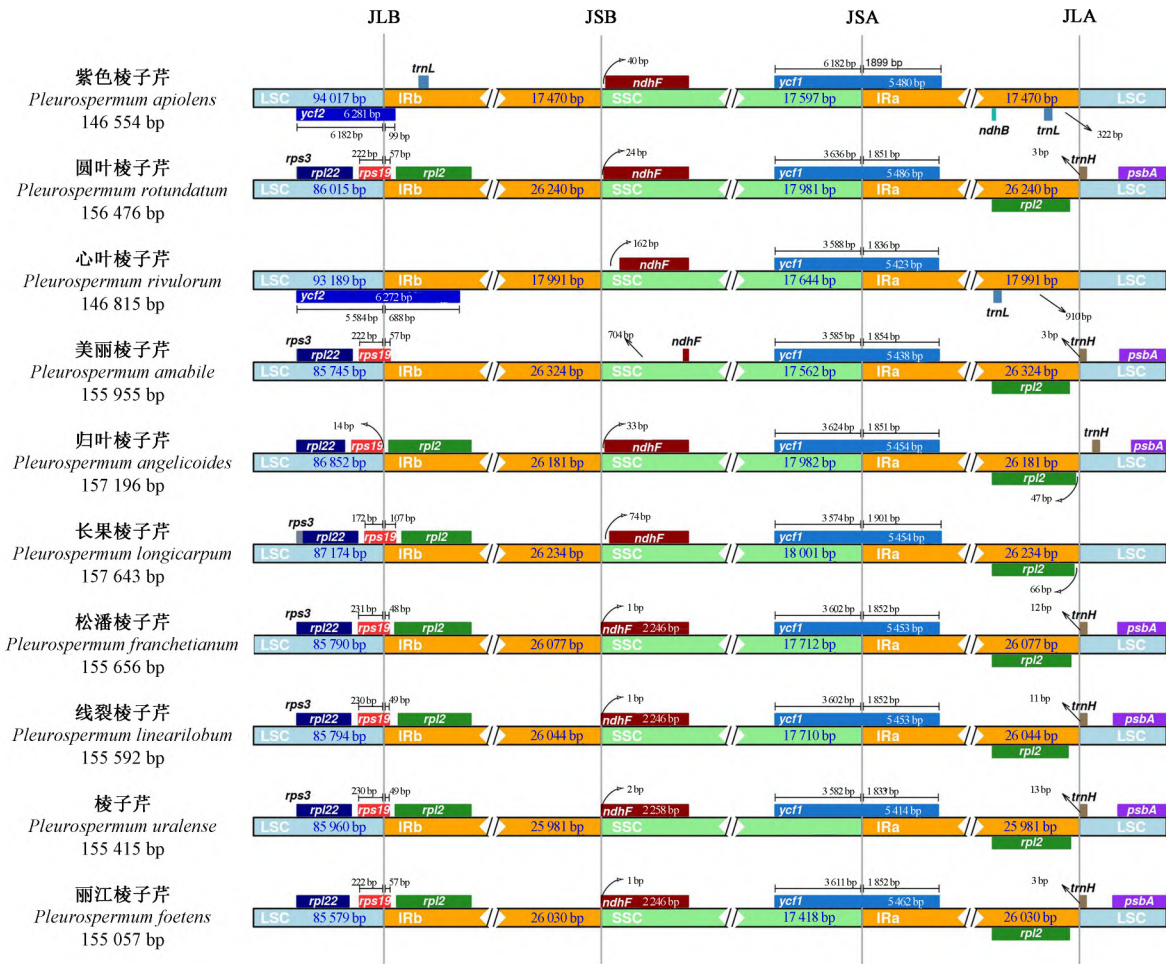


图4 棱子芹属植物叶绿体基因组 LSC、SSC 和 IR 边缘区的比较

Figure 4 Comparison of LSC , SSC , and IR border regions of the chloroplast genomes of *Pleurospermum*

### 1.5 系统进化分析

选取包括丽江棱子芹在内的 20 种植物，以白芨 (*Eleutherococcus trifolius*) 为外类群进行系统进化分析，构建的 ML 树主要分为三大支，棱子芹、松潘棱子芹、丽江棱子芹和线裂棱子芹聚为一支，其中丽江棱子芹与线裂棱子芹的亲缘关系最近，节点处支

持率达到 100%；美丽棱子芹、圆叶棱子芹 (*P. rotundatum*)、长果棱子芹和归叶棱子芹聚为一支，其中归叶棱子芹与长果棱子芹亲缘关系更近，节点处支持率为 97%；心叶棱子芹和紫色棱子芹聚为一支，距离芹亚科芫荽族 (*Coriandreae*) 的芫荽 (*Coriandrum sativum*) 更近 (图 5)。

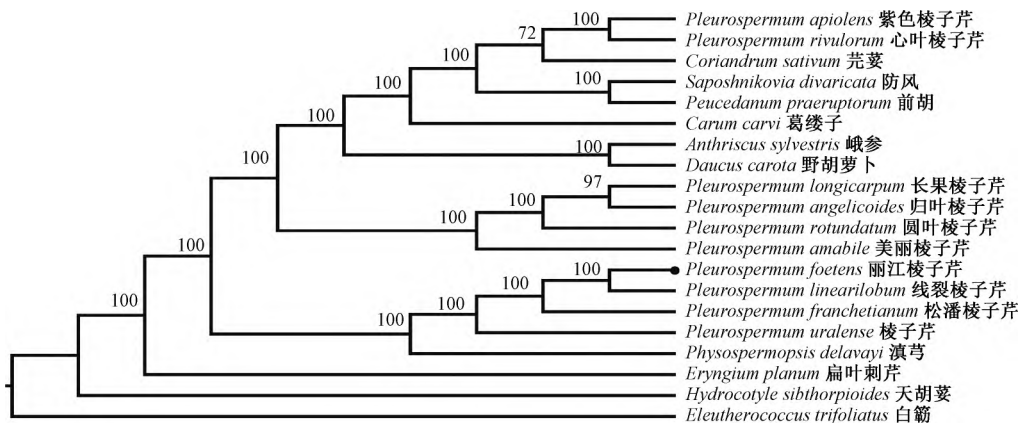


图5 基于 20 种伞形目物种叶绿体基因组构建的系统发育树

Figure 5 Phylogenetic tree constructed based on the chloroplast genomes of 20 Apiales species

## 2 讨论与结论

本研究基于高通量技术完成了对丽江梭子芹叶绿体基因组的测序、组装和注释, 研究结果显示, 丽江梭子芹叶绿体基因组全长 155 057 bp, 共编码 87 个基因, GC 含量为 38.0%, 具有典型的四分体结构。经过统计, 梭子芹属植物的叶绿体基因组长度变化不大, 大多在 146~155 kb, 其编码基因数量以及基因组的 GC 含量, 与已发表的梭子芹属植物叶绿体基因组( Guo et al., 2020) 相似, 由此分析梭子芹属植物叶绿体基因组结构较保守。

密码子偏好性的研究有助于叶绿体基因的表达, 通过基因工程提高某种目的基因的表达量( 梁菲菲, 2010)。本文对丽江梭子芹叶绿体基因组密码子偏好性进行分析, 结果表明, 编码异亮氨酸 Ile 的 ATT 数量最多, 终止密码子的选择也具有偏好性, 使用最多的是 TAA。在对丽江梭子芹的基因进行组学编辑时, 可较多地使用这几种密码子。并且, 密码子的偏好性与物种的亲缘关系有关, 亲缘关系近的物种, 所偏好的密码子相似。RSCU 分析中, RSCU 值大于 1 的相对同义密码子有 30 个, 其中以 A/T 碱基结尾的密码子有 29 个, 与大多数被子植物同义密码子组成相似, 均为 A/T 结尾, 如直刺变豆菜( *Sanicula orthacantha*)、苦马豆( *Sphaerophysa sal-sula*)、毛籽红山茶( *Camellia semiserrata*) 等( 陈志祥等, 2019; 苏丹丹等, 2022; 郑倩等, 2022)。

SSR 是基因组中 1~6 bp 的核苷酸串联序列, 广泛存在于基因组中。有研究表明位于非编码区的 SSR 比位于编码区的变化更大( Litt and Luty, 1989; Gandhi et al., 2010; 梅抗抗等, 2016)。目前, 叶绿体基因组的 SSR 分析在植物属间分类、生物地理学和群体遗传学的应用研究较为广泛( Bayly et al., 2013; Zhao et al., 2015)。在本文的分析中发现了两种 cpSSR 类型: 单核苷酸重复和二核苷酸重复, 且这些 SSR 中碱基重复单元大多为 A/T, 重复单元多位于 IGS 区。这一结果与伞形科的直刺变豆菜的结果( 陈志祥等, 2019) 相似, 均为 A/T 重复多。串联重复序列与散在重复序列的结果同样显示重复大多位于 IGS 区, 验证了前人研究结果。

在植物叶绿体基因组进化的过程中, IR 边界的收缩与扩张可以影响植物叶绿体基因组的长度与功能( Palmer and Thompson, 1982; Chaw et al., 2018), 也有一些物种出现了 IR 区部分缺失( Saski et al., 2005; 唐萍, 2015), 但本研究中梭子芹属植物的 IR

区结构保守, 均未发生收缩或扩张。10 种植物的边界基因均有 *ndhF* 与 *ycf1*, 除心叶梭子芹与紫色梭子芹在边界基因的类型、大小和位置有差异外, 其余几种梭子芹属植物的 IR 区相似。梭子芹属叶绿体基因组比对结果显示, 非编码区较编码区变异较大, LSC/SSC 区较 IR 区变异较大, 变异区域的 *matK*、*rpoC1*、*ycf1*、*rpoB-psbD*、*trnV-ndhC*、*ycf3-psaA*、*accD-cemA*、*ndhF-ndhD* 的 IGS 中变异位点数量较多, 这些差异大的序列可以用作 DNA 条形码备选区域, 以进行梭子芹属的分类鉴定。其中 *matK*、*ycf1* 在之前的研究中被选为陆生植物条形码( CBOL Plant Working Group, 2009; Dong et al., 2015)。

叶绿体基因组的系统发育结果中, 梭子芹属不能很好地聚为一支, 与梭子芹属一直存在分类争议的情况相同。但总体分支与韦瑾( 2020) 基于 ITS 分析所提出的划分标准相似, 韦瑾认为梭子芹、松潘梭子芹、丽江梭子芹和线裂梭子芹属梭子芹族, 美丽梭子芹属于东亚分支, 圆叶梭子芹、归叶梭子芹、长果梭子芹属于一支, 紫色梭子芹、心叶梭子芹各一支。本文中紫色梭子芹与心叶梭子芹聚为一支, 二者苞片为紫色, 形态上也有很强的相似性, 对于二者的系统发育位置还需要进行更多采样测序以进行验证。

综上所述, 基于叶绿体基因组结构与密码子偏好性、IR 区及全基因组比对分析可以发现, 丽江梭子芹的叶绿体基因组较为保守, 梭子芹属的系统进化关系依旧没有得到明确的界定, 因此将来还需要对更多梭子芹属植物的叶绿体基因组序列进行测定, 并结合形态解剖学、细胞分类学和分子生物学等多方面来对梭子芹属及其近缘属的系统发育关系进行研究。

## 3 材料与方法

### 3.1 实验材料

材料采集自云南省迪庆藏族自治州的大雪山垭口( 28°35'44" N, 99°51'33" E, 海拔 4 471 m), 选取生长状态良好、完整且无虫害脏污的新鲜叶片, 采摘后置于变性硅胶中进行干燥处理, 凭证标本( 标本号: NiuMa-147) 保存于中国科学院昆明植物研究所标本馆。叶绿体基因组数据信息已上传到 NCBI 数据库( 索引号为 OP-583945)。

### 3.2 叶绿体基因组的提取与测序

取已干燥的叶片, 采用 CTAB 法提取丽江梭子芹的叶绿体 DNA, 利用 1% 琼脂糖凝胶电泳分析

DNA 的纯度和完整性,使用 Qubit® DNA 试剂盒测定 DNA 浓度。利用 Agilent 2100 对文库的插入片段进行检测,待插入片段大小符合预期,使用 Q-PCR 方法对文库的有效浓度进行准确定量,以保证文库质量。质检后利用二代测序技术依靠 Illumina 高通量测序平台 Nova Seq 6000 对基因组测序,PE 读长为 150 bp。最终得到 2 G raw data, Q30 大于 80%。此过程由北京诺禾致源科技股份有限公司完成。

3.3 叶绿体基因组的组装与注释

测序得到的 Clean data 在 Linux 系统下使用 Get Organelle v.1.7.5.2( Jin et al.,2020) 进行组装,输出序列以 FASTA 格式存储。组装结果放入 NCBI 中进行 BLAST 查找相似度最高的序列( 线裂棱子芹),利用 CPGAVAS2 在线软件( Shi et al.,2019) 对序列进行注释,在 Geneious 软件( Kearse et al.,2012) 中对注释好的序列中存在问题的密码子位点进行手动校正后上传至 OGDRAW 在线工具( Lohse et al.,2007) 绘制丽江棱子芹的叶绿体基因组环状图谱。

3.4 密码子偏好性与重复序列、SSR 分析

利用 EMBOSS explorer 在线软件( <https://www.bioinformatics.nl/emboss-explorer/>)、CodonW 软件( Peden,1999) 检测丽江棱子芹 cpDNA 的密码子使用频率和相对密码子使用频率。密码子表选择通用密码子表,其他参数默认。使用 REPuter 软件进行正向、反向和回文 3 种类型的重复序列的检测,其中汉明距离设为 3,最小重复长度设为 30 bp。利用

Tandem Repeats Finder 软件分析串联重复序列( Benson,1999; Kurtz et al.,2001)。利用 MISA 软件( Thiel et al.,2003) 分析丽江棱子芹叶绿体全基因组 SSR 位点,不同重复单元最低重复次数参数设置为:单核苷酸重复次数  $\geq 10$ ,二核苷酸重复次数  $\geq 6$ ,三/四/五核苷酸重复次数  $\geq 5$ 。

3.5 棱子芹属植物种间基因组比较与 IR 区边界分析

将丽江棱子芹与从 NCBI 下载的已发表的 9 种棱子芹属植物( 表 7) 的叶绿体基因组进行比对分析。利用 Microsoft Excel 统计叶绿体基因组的结构信息,使用 mVISTA 在线工具( Mayor et al.,2000) 在 Multi-LAGAN 模式下进行 10 种棱子芹属植物的全基因组比对( 丽江棱子芹作为参考,在图中不显示)。利用 IRscope 可视化工具( Amiryousefi et al.,2018) 对 IR 区边界进行分析。

3.6 系统进化分析

从 NCBI 数据库下载棱子芹属及其近缘属叶绿体全基因组序列,外类群为白簕,共计 20 条物种数据,在 Geneious 软件中进行序列比对,提取 67 个共有 CDS,序列联合后在 IQ-TREE 软件( Nguyen et al.,2015) 中构建 ML 系统发育树,自举值( bootstrap value) 设置为 1 000 以确保系统发育树的稳定性。将建树结果置于 Fig Tree 网站( <http://tree.bio.ed.ac.uk/software/figtree/>) 中进行调整。

表 7 棱子芹属系统发育分析获取自 NCBI 数据库序列号信息  
Table 7 Phylogenetic analysis of the genus *Pleurospermum* obtained from the NCBI database  
accession number information

| 物种<br>Species                            | NCBI 索引号<br>NCBI accession number |
|------------------------------------------|-----------------------------------|
| 丽江棱子芹 <i>Pleurospermum foetens</i>       | OP_583945                         |
| 线裂棱子芹 <i>Pleurospermum linearilobum</i>  | NC_060751                         |
| 松潘棱子芹 <i>Pleurospermum franchetianum</i> | NC_060749                         |
| 棱子芹 <i>Pleurospermum uralense</i>        | NC_033343                         |
| 美丽棱子芹 <i>Pleurospermum amabile</i>       | MT_609922                         |
| 圆叶棱子芹 <i>Pleurospermum rotundatum</i>    | NC_060756                         |
| 心叶棱子芹 <i>Pleurospermum rivulorum</i>     | NC_060753                         |
| 归叶棱子芹 <i>Pleurospermum angelicoides</i>  | NC_060754                         |
| 长果棱子芹 <i>Pleurospermum longicarpum</i>   | NC_060755                         |
| 紫色棱子芹 <i>Pleurospermum apiolens</i>      | NC_060757                         |

## 作者贡献

张婕妤是本研究的实验设计和实验研究的执行人; 张婕妤和俞树良完成数据分析和论文初稿的写作; 孙文光参与实验设计和试验结果分析; 李志敏和孙文光是项目的构思者及负责人, 指导实验设计、数据分析和论文写作与修改。全体作者都已阅读并同意最终的文本。

## 致谢

感谢中国科学院昆明植物研究所马祥光博士对野外采样及标本鉴定提供的帮助。

## 参考文献

- 陈志祥, 姚雪莹, Stephen R. Downie, 等, 2019. 直刺变豆菜叶绿体全基因组及其特征. 生物多样性, 27(4): 366-372. [CHEN Z X, YAO X Y, DOWNIE S R, et al., 2019. Assembling and analysis of *Sanicula orthacantha* chloroplast genome. Biodiversity Science, 27(4): 366-372.]
- 梁菲菲, 2010. 密码子偏性的影响因素及研究意义. 畜牧与饲料科学, 31(1): 118-119. [LIANG F F, 2010. Influencing factors of codon bias and its research significance. Animal Husbandry and Feed Science, 31(1): 118-119.]
- 梅抗抗, 周力, 唐开宇, 等, 2016. 油橄榄分子标记及转录组学研究进展. 分子植物育种, 14(12): 3469-3478. [MEI K K, ZHOU L, TANG K Y, et al., 2016. Application of molecular markers and transcriptomics research progress in olive. Molecular Plant Breeding, 14(12): 3469-3478.]
- 苏丹丹, 刘玉萍, 刘涛, 等, 2022. 苦马豆叶绿体基因组结构及其特征分析. 植物研究, 42(3): 446-454. [SU D D, LIU Y P, LIU T, et al., 2022. Structure of chloroplast genome and its characteristics of *Sphaerophysa salsula*. Bulletin of Botanical Research, 42(3): 446-454.]
- 孙文远, 孙文光, 李志敏, 2020. 横断山高山冰缘带两种伞形科特有植物的核型分析. 云南师范大学学报(自然科学版), 40(2): 59-64. [SUN W Y, SUN W G, LI Z M, 2020. Karyomorphological study of two endemic plants (Apiaceae: *Pleurospermum*) from the alpine subnival belt in Hengduan Mountains. Journal of Yunnan Normal University (Natural Sciences Edition), 40(2): 59-64.]
- 唐萍, 2015. 蝶形花科植物叶绿体基因组研究进展. 南方农业学报, 46(1): 21-25. [TANG P, 2015. Advances in chloroplast genome of widely diversified Papilionaceae. Journal of Southern Agriculture, 46(1): 21-25.]
- 汪小全, 洪德元, 1997. 植物分子系统学近五年的研究进展概况. 植物分类学报, 35(5): 465-480. [WANG X Q, HONG D Y, 1997. Progress in molecular systematics of plants in recent five years. Acta Phytotaxonomica Sinica, 35(5): 465-480.]
- 王会朋, 马祥光, 郝鹏, 等, 2014. 中国棱子芹属 16 种 1 变种的果实特征及其分类学意义. 西北植物学报, 34(4): 711-719. [WANG H P, MA X G, GAO P, et al., 2014. Fruit anatomical features of sixteen species and one variety of *Pleurospermum* (Apiaceae) in China and their taxonomic significance. Acta Botanica Boreali-Occidentalia Sinica, 34(4): 711-719.]
- 韦瑾, 2020. 中国棱子芹属及其近缘类群的系统学研究: 兼论伞形科药用资源调研, 硕士学位论文. 昆明: 昆明医科大学. [WEI J, 2020. A phylogenetic study of *Pleurospermum* Hoffm. and related Genera in China: investigation of medicinal plants in Apiaceae, Thesis for MS. Kunming: Kunming Medical University.]
- 余涛, 蒲芬, 管芹, 等, 2022. 南欧大戟叶绿体基因组特征及其系统发育分析. 分子植物育种, 20(6): 1828-1837. [YU T, PU F, GUAN Q, et al., 2022. Chloroplast genome of *Euphorbia peplus* L. characteristic and phylogenetic analysis. Molecular Plant Breeding, 20(6): 1828-1837.]
- 张勇, 刘启新, 王利松, 等, 2013. 伞形科棱子芹属花粉形态特征及其演化意义. 植物资源与环境学报, 22(4): 29-37. [ZHANG Y, LIU Q X, WANG L S, et al., 2013. Pollen morphological characteristics of *Pleurospermum* Hoffm. (Apiaceae) and its evolution significance. Journal of Plant Resources and Environment, 22(4): 29-37.]
- 赵耀, 刘汉梅, 顾勇, 等, 2008. 玉米 waxy 基因密码子偏好性分析. 玉米科学, 16(2): 16-21. [ZHAO Y, LIU H M, GU Y, et al., 2008. Analysis of characteristic of codon usage in waxy gene of *Zea mays*. Journal of Maize Sciences, 16(2): 16-21.]
- 郑倩, 童一涵, 孔庆博, 等, 2022. 毛籽红山茶叶叶绿体基因组特征及其系统发育分析. 四川农业大学学报, 40(4): 574-582. [ZHENG Q, TONG Y H, KONG Q B, et al., 2022. Characterization of complete chloroplast genome and phylogenetic analysis of *Camellia trichosperma* Chang. Journal of Sichuan Agricultural University, 40(4): 574-582.]
- AMIRYOUSEFI A, HYVÖNEN J, POCZAI P, 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinform. (Oxf. Engl.), 34(17): 3030-3031.
- BAYLY M J, RIGAULT P, SPOKEVICIUS A, et al., 2013. Chloroplast genome analysis of Australian eucalypts – *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). Mol. Phylogenet. Evol., 69(3): 704-716.
- BENSON G, 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res., 27(2): 573-580.
- CAI Z Q, GUISENGER M, KIM H G, et al., 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences

- and novel DNA insertions. *J. Mol. Evol.*, 67( 6) : 696-704.
- CBOL Plant Working Group , 2009. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA* , 106 ( 31) : 12794-12797.
- CHAW S M , WU C S , SUDIANTO E , 2018. Evolution of gymnosperm plastid genomes. *Adv. Bot. Res.* 85: 195-222.
- DONG W P , XU C , LI C H , et al. , 2015. *ycf1* , the most promising plastid DNA barcode of land plants. *Sci. Rep.* , 5: 8348.
- GANDHI S G , AWASTHI P , BEDI Y S , 2010. Analysis of SSR dynamics in chloroplast genomes of Brassicaceae family. *Bioinformation* , 5( 1) : 16-20.
- GUISINGER M M , KUEHL J V , BOORE J L , et al. , 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements , repeats , and codon usage. *Mol. Biol. Evol.* , 28( 1) : 583-600.
- GUO M J , WEI J , LIU Z W , et al. , 2020. The complete chloroplast genome and phylogenetic analysis of *Pleurospermum amabile* Craib & W. W. Smith. *Mitochondrial DNA B* , 5 ( 3) : 2966-2968.
- HE K , JIANG X L , 2014. Sky Islands of southwest China. I : an overview of phylogeographic patterns. *Chin. Sci. Bull.* , 59 ( 7) : 585-597.
- HERSHBERG R , PETROV D A , 2008. Selection on codon bias. *Annu. Rev. Genet.* , 42: 287-299.
- JANSEN R K , RAUBESON L A , BOORE J L , et al. , 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Meth. Enzymol.* , 395: 348-384.
- JIN J J , YU W B , YANG J B , et al. , 2020. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* , 21( 1) : 1-31.
- KEARSE M , MOIR R , WILSON A , et al. , 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinform.* ( Oxf. Engl.) , 28( 12) : 1647-1649.
- KURTZ S , CHOUDHURI J V , OHLEBUSCH E , et al. , 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* , 29( 22) : 4633-4642.
- LIANG C L , WANG L , LEI J , et al. , 2019. A comparative analysis of the chloroplast genomes of four *Salvia* medicinal plants. *Engineering* , 5( 5) : 907-915.
- LITT M , LUTY J A , 1989. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat with in the cardiac muscle actin gene. *Am. J. Hum. Genet.* , 44 ( 3) : 397-401.
- LOHSE M , DRECHSEL O , BOCK R , 2007. OrganellarGenome-DRAW ( OGDRAW) : a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* , 52( 5-6) : 267-274.
- MAYOR C , BRUDNO M , SCHWARTZ J R , et al. , 2000. VIS-TA: visualizing global DNA sequence alignments of arbitrary length. *Bioinform.* ( Oxf. Engl.) , 16( 11) : 1046-1047.
- NGUYEN L T , SCHMIDT H A , VON HAESELER A , et al. , 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* , 32( 1) : 268-274.
- PALMER J D , 1985. Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* , 19: 325-354.
- PALMER J D , THOMPSON W F , 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* , 29( 2) : 537-550.
- PEDEN J F , 1999. Analysis of codon usage , Dissertation for Ph. D. Nottingham: University of Nottingham.
- SASKI C , LEE S B , DANIELL H , et al. , 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other Legume genomes. *Plant Mol. Biol.* , 59( 2) : 309-322.
- SHE M L , PU F D , PAN Z H , et al. , 2005. Apiaceae ( Umbelliferae) // Wu Z Y , Hong D Y , Raven P H , et al. , *Flora of China*. Beijing & St. Louis: Science Press & Missouri Botanical Garden Press: 14-45.
- SHI L C , CHEN H M , JIANG M , et al. , 2019. CPGAVAS2 , an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* , 47( W1) : W65-W73.
- THIEL T , MICHALEK W , VARSHNEY R , et al. , 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley ( *Hordeum vulgare* L.) . *Theor. Appl. Genet.* , 106( 3) : 411-422.
- WOLFE K H , LI W H , SHARP P M , 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial , chloroplast , and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* , 84 ( 24) : 9054-9058.
- ZHAO Y B , YIN J L , GUO H Y , et al. , 2015. The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. *Front. Plant Sci.* , 5: 696.

( 责任编辑 黄新元)  
( 责任副主编 刘云峰)