

DOI: 10.11913/PSJ.2095-0837.2022.60791

李佳, 李清. 美国黄松和蓝粉云杉叶绿体基因组序列测定与分析[J]. 植物科学学报, 2022, 40(6): 791~800

Li J, Li Q. Complete chloroplast genome sequence and analysis of *Pinus ponderosa* P. Lawson & C. Lawson and *Picea pungens* Engelm. [J]. Plant Science Journal, 2022, 40(6): 791~800

美国黄松和蓝粉云杉叶绿体基因组序列测定与分析

李佳*, 李清

(陕西学前师范学院生命科学与食品工程学院, 西安 710100)

摘要: 采用高通量测序技术获得了美国黄松(*Pinus ponderosa* P. Lawson & C. Lawson)和蓝粉云杉(*Picea pungens* Engelm.)的叶绿体全基因组序列。结果显示, 美国黄松和蓝粉云杉的叶绿体基因组大小分别为120 274和124 237 bp。美国黄松共注释到107个基因, 其中包含70个蛋白编码基因, 4个rRNA基因, 33个tRNA基因; 蓝粉云杉包含111个基因, 其中71个蛋白编码基因, 4个rRNA基因, 36个tRNA基因。美国黄松和蓝粉云杉分别鉴定到18和32个简单重复序列单元。基于19个松科植物和2个外类群共有的73个叶绿体基因的系统发育分析结果, 可将松科分为两大分支: 第1分支包含了雪松属(*Cedrus*)、金钱松属(*Pseudolarix*)、铁杉属(*Tsuga*)、油杉属(*Keteleeria*)和冷杉属(*Abies*); 第2分支包含了松属(*Pinus*)、银杉属(*Cathaya*)、云杉属(*Picea*)、落叶松属(*Larix*)和黄杉属(*Pseudotsuga*)。叶绿体基因组结构分析结果表明, 松科不同属间发生了多次重排, 并且是由包含

关键词: 松科; 美国黄松; 蓝粉云杉; 叶绿体基因组; 系统发育

中图分类号: Q943.2

文献标识码: A

文章编号: 2095-0837(2022)06-0791-10

Complete chloroplast genome sequence and analysis of *Pinus ponderosa* P. Lawson & C. Lawson and *Picea pungens* Engelm.

Li Jia*, Li Qing

(College of Life Sciences and Food Engineering, Shaanxi Xueqian Normal University, Xi'an 710100, China)

Abstract: In this study, the complete chloroplast genome sequences of *Pinus ponderosa* P. Lawson & C. Lawson and *Picea pungens* Engelm. were determined using next-generation sequencing technology. The genome sizes of *Pinus ponderosa* and *Picea pungens* were 120 274 bp and 124 237 bp, respectively. The chloroplast genome of *P. ponderosa* contained 107 genes, including 70 protein-coding genes, four ribosomal RNA (rRNA) genes, and 33 transfer RNA (tRNA) genes. The chloroplast genome of *P. pungens* contained 111 genes, including 71 protein-coding genes, four rRNA genes, and 36 tRNA genes. In total, 18 and 32 simple sequence repeats were identified in the *Pinus ponderosa* and *Picea pungens* chloroplast genomes, respectively. Phylogenetic analysis using 73 common chloroplast genes from 19 Pinaceae and two outgroups divided Pinaceae into two groups: the first clade included *Cedrus*, *Pseudolarix*, *Tsuga*, *Keteleeria*, and *Abies*; the second clade included

收稿日期: 2022-07-23, 修回日期: 2022-08-28。

基金项目: 陕西学前师范学院省级大学生创新创业训练计划项目(S202014390065); 陕西省教育厅专项科学计划(19JK0211)。This work was supported by grants from the Provincial Innovation and Entrepreneurship Training Program for College Students of Shaanxi Xueqian Normal University (S202014390065) and Science Foundation from the Education Department of Shaanxi Provincial Government (19JK0211)。

作者简介: 李佳(1988-), 女, 博士, 高级实验师, 研究方向为植物系统与进化(E-mail: anjing_880702@126.com)。

* 通讯作者(Author for correspondence. E-mail: anjing_880702@126.com)。

Pinus, *Cathaya*, *Picea*, *Larix*, and *Pseudotsuga*. Analysis of chloroplast genome structure identified several rearrangements between different genera of Pinaceae, mediated by a *trnS-GCU* small inverted repeat sequence.

Key words: Pinaceae; *Pinus ponderosa*; *Picea pungens*; Chloroplast genome; Phylogeny

美国黄松(*Pinus ponderosa* P. Lawson & C. Lawson)属于松科松属植物，其树皮粗厚，树干高大挺直，在黄土高原丘陵沟壑区生长良好，是城市森林生态体系建设中首选的外来树种，也是森林中重要的防火树种^[1]。蓝粉云杉(*Picea pungens* Engelm.)是松科云杉属植物，其新生叶为蓝绿色，是珍稀蓝色彩叶树种，树形优美，常作为园林树种，栽培于公园、广场、街道及庭院。蓝粉云杉的叶片也可作为重金属污染的监测器^[2]。近年来，关于美国黄松与蓝粉云杉的研究主要集中在形态解剖、发育、繁育种植等方面^[3-6]，但对于其遗传资源等方面的研究较少。迄今为止，美国黄松与蓝粉云杉的叶绿体全基因组序列还没有被报道。

叶绿体是植物进行光合作用的主要场所，对维持地球上的生命具有重要作用^[7]。陆生植物的叶绿体基因组结构和基因含量都非常保守。一般由一个环状的四分体结构组成，主要包括两段序列相同、方向相反的反向重复区(Inverted repeat, IR) 和将其隔开的大单拷贝区(Large single copy, LSC) 和小单拷贝区(Small single copy, SSC)^[8]。叶绿体基因组一般包含 110 ~ 130 个基因，主要是光合系统、转录以及翻译相关基因^[9]。尽管结构和基因含量很保守，但叶绿体基因组的大小变异相对比较大，主要由于 IR 区的收缩、扩张以及部分基因的丢失引起的^[10]。叶绿体基因组具有拷贝数多，重组发生率低，进化速率适中等特点，已被广泛应用于植物系统发育和进化研究^[11]。叶绿体系统发育学已成功用于解决苔藓、地钱、蕨类、裸子和被子植物等类群的系统发育关系^[12]。根据叶绿体基因组序列还对被子植物(Angiosperm phylogeny group IV, 2016) 和蕨类的分类系统 (Pteridophyte phylogeny group, 2016) 进行了修订^[7]。

作为裸子植物最大的科，松科包含了 11 个属大约 230 个物种，是北半球松柏林中重要的组成部分^[13, 14]。所有的松科植物构成一个单系群已经

得到了各方面的数据支持，包括形态、结构、生物化学特性、叶绿体基因组结构等^[15-18]。多数分子系统学研究支持将松科分为一个松亚科和一个冷杉亚科，其中松亚科包含银杉属(*Cathaya*)、落叶松属(*Larix*)、云杉属(*Picea*)、松属(*Pinus*)和黄杉属(*Pseudotsuga*) 5 个属；冷杉亚科包含冷杉属(*Abies*)、雪松属(*Cedrus*)、油杉属(*Keteleeria*)、长苞铁杉属(*Nothotsuga*)、金钱松属(*Pseudolarix*)和铁杉属(*Tsuga*) 6 个属^[19-21]。在松亚科内部，银杉属和松属先聚在一起，并且与云杉属形成姐妹群。在冷杉亚科内部，雪松属是基部类群，长苞铁杉属与铁杉属形成姐妹群并与金钱松属聚在一起^[22]。但是，还有少部分系统发育结果与上述研究存在分歧，如 Ni 等^[23]构建的 16 个松柏类的叶绿体全基因组序列系统发育树显示，松属并没有和银杉属聚在一起；雪松属与冷杉属、云杉属聚在一起形成一个枝。此外，目前还没有基于叶绿体基因组序列，覆盖松科多个属的较全面的系统发育研究。因此，本研究对美国黄松与蓝粉云杉的叶绿体基因组序列进行测定，并利用其序列信息对松科植物的系统发育关系进行重建。

NCBI 网站目前已公布了 87 种松科植物 (<https://www.ncbi.nlm.nih.gov/genome/browse#! /organelles/Pinaceae>) 的叶绿体全基因组序列，覆盖了 11 个属。研究发现，该科植物叶绿体基因组表现出一些独有的特征：(1) 其基因组均丢失了大的反向重复区，大部分物种残留了一些小的反向重复序列，其长度大约为 200 ~ 1300 bp^[24-27]，这些小的反向重复区可代替原有的大的反向重复序列，引起分子内的同源重组^[28]。还有一些物种没有检测到小的反向重复序列，例如雪岭云杉(*Picea schrenkiana* Fisch. & C. A. Mey.) 和鱼鳞云杉(*Picea jezoensis* (Siebold & Zucc.) Carrière)^[27]，在这些类群中，大的直接重复序列也可引起同源重组^[29]。(2) 已报道的所有松科植物的叶绿体基因组均丢失了 11 个完整的 *ndh* 基因(包括 *ndhA* ~ *ndhK*)，并且在欧洲云杉(*Picea abies*) 中也丢失了 11 个完整的 *ndh* 基因^[29]。

abies (L.) H. Karst.) 中发现一些 *ndh* 基因片段转移至了核基因组中^[30]。(3) 松科云杉属叶绿体基因组序列和结构虽然比较保守, 但是分子内重组和发生正选择作用的基因比较多^[29]。松科叶绿体基因组结构和基因含量相对多变, 因此, 本研究基于高通量测序技术获得了美国黄松和蓝粉云杉的叶绿体基因组全长序列, 并分析了其基因含量与重复序列, 探讨了松科内部各属的系统发育关系, 研究结果旨在为松科植物叶绿体基因组变异和遗传育种工作积累理论基础。

1 材料与方法

1.1 植物材料

美国黄松及蓝粉云杉的新鲜叶片分别采自陕西西安秦岭国家植物园与西安植物园, 采集后置于硅胶中干燥待用。

1.2 研究方法

1.2.1 DNA 提取及测序

采用植物基因组 DNA 提取试剂盒提取美国黄松及蓝粉云杉的总基因组 DNA。检测合格的 DNA 样本送测序公司先进行文库构建及库检, 使用 Illumina 高通量测序平台 NovaSeq 6000 进行双末端测序, 读长为 150 bp。

1.2.2 叶绿体基因组的组装与注释

测序完成后, 美国黄松和蓝粉云杉分别获得了 4.96 和 5.45 Gb 的高质量测序片段。采用 NovoPlasty v4.3 软件^[31]对测序的原始数据进行拼接, 使用 PGA 软件^[32]进行叶绿体基因组注释, 叶绿体基因组结构图利用在线工具 OGDRAW v1.3 (<https://chlorobox.mpiimp-golm.mpg.de/OGDraw.html>) 进行绘制。利用 BLAST 方法鉴定叶绿体基因组中包含的小反向重复区域。最后使用 NCBI 在线工具 Bankit 平台 (<https://submit.ncbi.nlm.nih.gov/about/bankit/>) 提交序列。

1.2.3 重复序列分析

利用在线工具 MISA v2.1 (<https://webblast.ipk-gatersleben.de/misa/>) 检测美国黄松与蓝粉云杉中的简单重复序列 (Simple sequence repeat, SSR), 参数设置为单核苷酸重复不少于 10 次重复, 2 核苷酸不少于 6 次, 3~6 核苷酸不少于 5

次。此外, 本研究还使用 Tandem repeat finder v4.09 (<https://tandem.bu.edu/trf/trf.html>) 软件对美国黄松和蓝粉云杉中的串联重复序列进行了检测。

1.2.4 系统发育分析

本研究提取 19 种松科植物及两个外类群宽叶罗汉松 (*Podocarpus latifolius* (Thunb.) R. Br. ex Mirb.) 和南洋杉 (*Araucaria cunninghamii* Aiton ex D. Don) 共有的 69 个蛋白编码基因及 4 个 rRNA 基因序列(附表 1¹⁾), 然后利用 Cluster W 软件分别进行序列比对, 最后合并为一条序列用于系统发育分析。基于最大似然法, 采用 IQ-TREE v2.0 软件^[33]进行系统发育分析, 采用软件计算的最优模型, 自展值设为 1000。

1.2.5 叶绿体基因组结构分析

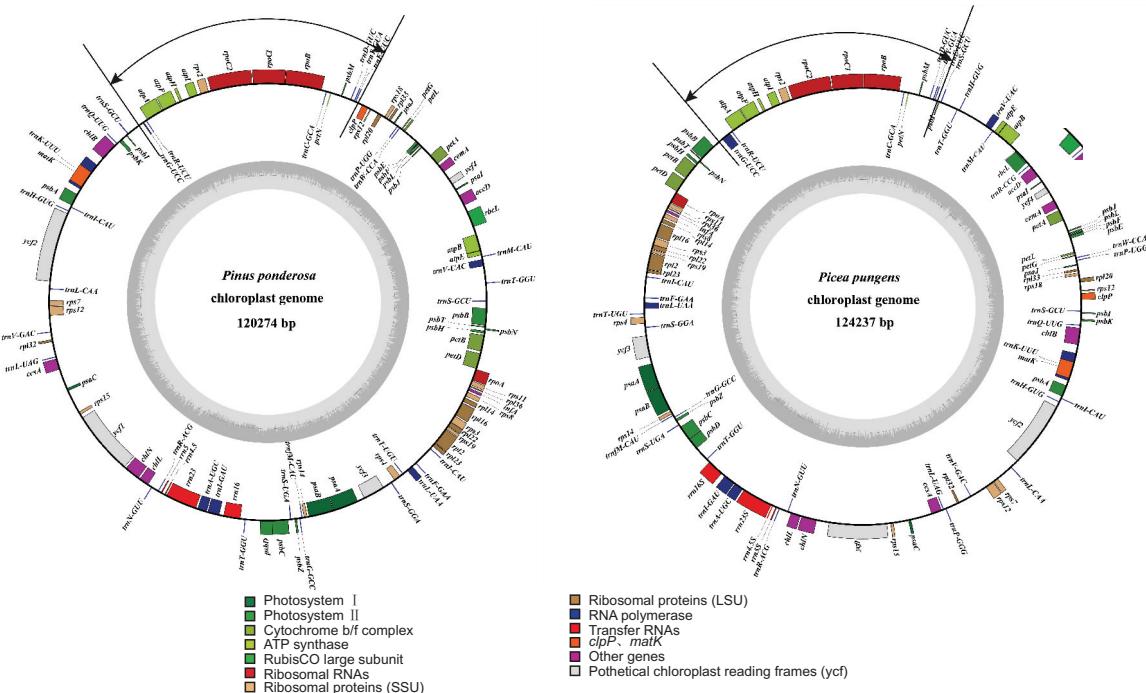
选取松科 10 个属各一个物种及蓝粉云杉、美国黄松进行叶绿体基因组重排分析。手动将所有物种叶绿体基因组序列调整为以 *psbA* 基因为起始, 然后采用 MAUVE v2.3.1 软件^[34]进行序列比对。最后使用在线工具 MGR 2.0.1 (<http://grimm.ucsd.edu/MGR/pubs.html>) 计算 12 种植物之间叶绿体基因组的重排次数。

2 结果与分析

2.1 美国黄松与蓝粉云杉的叶绿体基因组基本特征

序列拼接结果显示, 美国黄松与蓝粉云杉拼接的覆盖率分别是 178 x 和 187 x (登录号分别为 OL547484 和 OL547483)。二者叶绿体基因组全长分别为 120 274 bp 和 124 237 bp (图 1)。与其他松科植物类似, 两种植物叶绿体基因组均不包含典型的 IR 区, 因此也不具有四分体结构。美国黄松共注释到 107 个基因, 包含 70 个蛋白编码基因, 4 个 rRNA 基因, 33 个 tRNA 基因 (其中 *trnL-CAU*、*trnS-GCU* 和 *trnT-GGU* 为双拷贝基因)。蓝粉云杉共注释到 111 个基因, 包含 71 个蛋白编码基因 (包括双拷贝基因 *psbI*), 4 个 rRNA 基因, 36 个 tRNA 基因 (其中 *trnH-GUG*、*trnL-CAU*、*trnS-GCU*、*trnT-GGU* 有两个拷贝)。美国黄松和蓝粉云杉均有 12 个基因包含 1 个内含子, 2 个基因包含两个内含子 (表 1)。美国黄松和蓝粉

1) 如需查阅附表内容请登录《植物科学学报》网站 (<http://www.plantscience.cn>) 查看本期文章。



环内侧的基因顺时针转录，环外侧的基因逆时针转录；图中圆弧所示为美国黄松与蓝粉云杉基因排列顺序相同部分。Genes shown inside and outside circle are transcribed clockwise and counterclockwise, respectively. Curvature in the figure denotes gene arrangement was the same between *Pinus ponderosa* and *Picea pungens*.

图 1 美国黄松与蓝粉云杉叶绿体基因组结构

Fig. 1 Complete chloroplast genome map of *Pinus ponderosa* and *Picea pungens*

云杉的叶绿体基因组总 GC 含量分别为 38.46% 和 38.73%。

对本研究测序的蓝粉云杉、美国黄松与目前已公布的其他松科植物的叶绿体基因组序列进行比较(附表 1)，可以看出，麦吊云杉(*Picea brachytyla* (Franch.) E. Pritz.)的叶绿体基因组长度是最长的，为 124 291 bp，而且云杉属中除了巨云杉(*Picea sitchensis* Trautv. & G. Mey., 120 176 bp)外，其他物种长度也基本都在 124 kb 左右。其次，银杉(*Cathaya argyrophylla* Chun & Kuang)是所有松科中叶绿体基因组长度最短的(107 122 bp)。松科所有物种均不包含 *ndh* 基因，而宽叶罗汉松和南洋杉包含了 *ndhA*、*ndhB*、*ndhC* 等 11 个 *ndh* 基因。此外，还有两个蛋白编码基因 *psaM* 和 *psbI* 的拷贝数在松科植物中比较多变。*psaM* 基因在落叶松属、金钱松属、黄杉属、铁杉属发生了丢失，但在冷杉属、雪松属、油杉属和美国黄松中存在两个拷贝。*psbI* 基因也出现类似的情况，雪松属、油杉属、落叶松属、云杉属、金钱松属包含两个拷贝的 *psbI* 基因，黄杉属甚至包含了 3 个拷贝。引起 *psaM* 和 *psbI* 基因拷

贝数不同的主要原因是它们位于松科残留的反向重复区。

2.2 重复序列分析

SSR 序列的变异对于了解种群遗传多样性、系统发育和生物地理等研究可提供重要信息，因此，本研究对蓝粉云杉和美国黄松的 SSR 序列进行了分析。分析结果显示，蓝粉云杉总共包含 32 个 SSR (图 2)，其中 23 个为单碱基，8 个为双碱基，1 个 3 碱基重复。从分布的区域上来看，大部分的 SSR 都位于基因间隔区(24 个)，6 个位于内含子区域，分布在 *rpoC1*(2 个)、*trnS-CGA*(2 个)、*rpl16*、*ycf3* 基因上，仅有 2 个位于 *accD*、*psbZ* 基因的编码区。美国黄松总共包含 18 个简单重复序列单元，15 个为单碱基重复，3 个双碱基重复；美国黄松叶绿体基因组中不包含其他长度的简单重复序列。在分布区域上与蓝粉云杉类似，14 个 SSR 都位于基因间隔区，其中两个位于 *ccsA* 到 *psaC* 基因的间隔区；3 个位于内含子区域，分别位于 *rpoC1*、*rps12*、*ycf3* 基因上；仅有 1 个位于 *rpl32* 的基因编码区。

表 1 蓝粉云杉叶绿体基因组包含的基因
Table 1 Genes contained in *Picea pungens* chloroplast genome

基因功能分类 Gene function category	基因种类 Gene category	基因名称 Gene name
自我复制	核糖体 RNA 基因	<i>rrn16</i> 、 <i>rrn23</i> 、 <i>rrn4.5</i> 、 <i>rrn5</i>
	转运 RNA 基因	<i>trnA-UGC</i> [*] 、 <i>trnC-GCA</i> 、 <i>trnD-GUC</i> 、 <i>trnE-UUC</i> 、 <i>trnF-GAA</i> 、 <i>trnfM-CAU</i> 、 <i>trnG-GCC</i> 、 <i>trnG-UCC</i> [*] 、 <i>trnH-GUG</i> ^a 、 <i>trnI-CAU</i> ^a 、 <i>trnI-GAU</i> [*] 、 <i>trnK-UUU</i> [*] 、 <i>trnL-CAA</i> 、 <i>trnL-UAA</i> [*] 、 <i>trnL-UAG</i> 、 <i>trnM-CAU</i> 、 <i>trnN-GUU</i> 、 <i>trnP-GGG</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> ^a 、 <i>trnS-GGA</i> 、 <i>trnS-UGA</i> 、 <i>trnT-GGU</i> ^a 、 <i>trnT-UGU</i> 、 <i>trnV-GAC</i> 、 <i>trnV-UAC</i> [*] 、 <i>trnW-CCA</i> 、 <i>trnY-GUA</i>
	RNA 聚合酶亚基	<i>rpoA</i> 、 <i>rpoB</i> 、 <i>rpoC1</i> [*] 、 <i>rpoC2</i>
	核糖体小亚基(SSU)	<i>rps2</i> 、 <i>rps3</i> 、 <i>rps4</i> 、 <i>rps7</i> 、 <i>rps8</i> 、 <i>rps11</i> 、 <i>rps12</i> ^{**} 、 <i>rps14</i> 、 <i>rps15</i> 、 <i>rps18</i> 、 <i>rps19</i>
	核糖体大亚基(LSU)	<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>
	光系统 I	<i>psaA</i> 、 <i>psaB</i> 、 <i>psaC</i> 、 <i>psaI</i> 、 <i>psaJ</i>
	光系统 II	<i>psbA</i> 、 <i>psbB</i> 、 <i>psbC</i> 、 <i>psbD</i> 、 <i>psbE</i> 、 <i>psbF</i> 、 <i>psbH</i> 、 <i>psbI</i> ^a 、 <i>psbJ</i> 、 <i>psbK</i> 、 <i>psbL</i> 、 <i>psbM</i> 、 <i>psbN</i> 、 <i>psbT</i> 、 <i>psbZ</i>
	光合作用相关	<i>petA</i> 、 <i>petB</i> [*] 、 <i>petD</i> [*] 、 <i>petG</i> 、 <i>petL</i> 、 <i>petN</i>
其他基因	细胞色素 b/f 复合物	<i>atpA</i> 、 <i>atpB</i> 、 <i>atpE</i> 、 <i>atpF</i> [*] 、 <i>atpH</i> 、 <i>atpI</i>
	ATP 合成酶	<i>chlB</i> 、 <i>chlL</i> 、 <i>chlN</i>
	叶绿素生物合成	<i>rbcL</i>
	二磷酸核酮糖羧化酶大亚基	
	成熟酶	<i>matK</i>
	被膜蛋白	<i>cemA</i>
	乙酰辅酶 A 羧化酶	<i>accD</i>
未知功能基因	细胞色素 C 的合成基因	<i>ccsA</i>
	蛋白酶	<i>clpP</i>
	转录启动因子	<i>infA</i>
	假定的保守叶绿体开放阅读框	<i>ycf1</i> 、 <i>ycf2</i> 、 <i>ycf3</i> ^{**} 、 <i>ycf4</i>

注: ^a代表基因有两个; * 代表基因含有 1 个内含子; ** 代表基因含有 2 个内含子。美国黄松与蓝粉云杉相比, *psbI*、*trnH-GUG* 基因只有 1 个拷贝, *trnR-CCG*、*trnP-GGG* 两个 tRNA 基因发生了丢失。

Notes: ^a indicates gene has two copies; * indicates gene contains one intron; ** indicates gene contains two introns. Compared with *Picea pungens*, only one copy of *psbI* and *trnH-GUG* were identified and *trnR-CCG* and *trnP-GGG* genes were lost in *Pinus ponderosa* chloroplast genome.

串联重复序列分析结果表明, 蓝粉云杉总共包含 24 个串联重复序列, 其中位于基因间隔区的有 12 个, 内含子区域有 1 个, 蛋白编码区的有 11 个(图 2)。蓝粉云杉中最长的串联重复序列为 293 bp, 一段 33 bp 的序列重复了 9 次, 位于 *chlN* 到 *ycf1* 基因的间隔区。美国黄松中共检测到 21 段串联重复序列, 其中位于蛋白编码区和基因间隔区的各 10 个, 内含子区域的 1 个(图 2)。最长的一段串联重复长度为 172 bp, 长度 55 bp 的序列重复了 3 次, 位于 *chlL* 到 *trnN-GUU* 基因的间隔区。

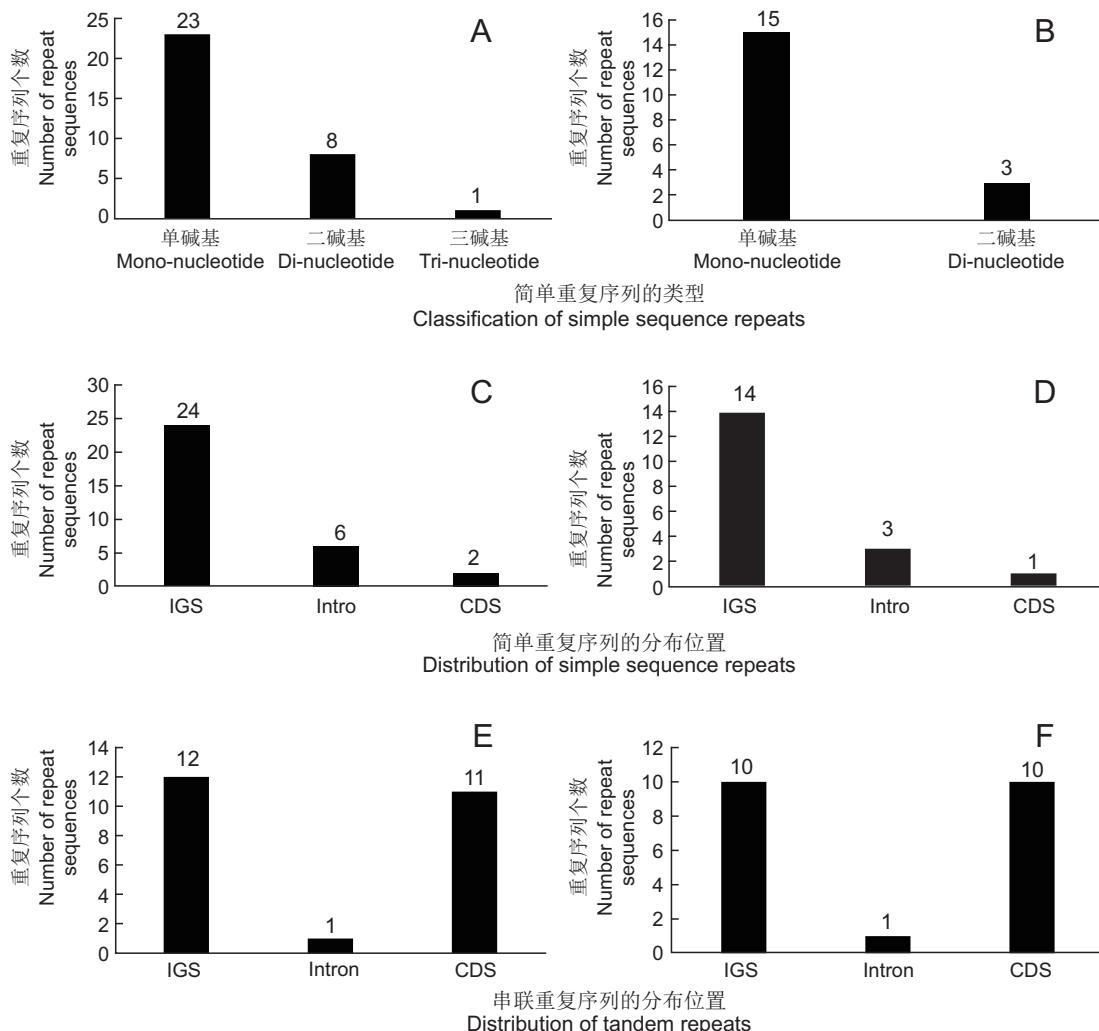
2.3 松科植物的系统发育分析

本研究基于 19 个松科及 2 个外类群宽叶罗汉松和南洋杉重建了松科植物的系统发育关系, 73 个基因合并后的序列总长为 42 627 bp, 其中包含了 13 898 个多态性位点。系统进化树分析结果显示, 松科内部被分为两大主要分支, 第 1 分支主要

包含雪松属、金钱松属、铁杉属、油杉属和冷杉属, 雪松属为第 1 分支的基部类群, 油杉属和冷杉属亲缘关系较近; 铁杉属与金钱松属聚在了一起(图 3)。第 2 分支包含松属、银杉属、云杉属、落叶松属和黄杉属, 其中, 松属和银杉属亲缘关系较近, 再与云杉属聚在一起, 落叶松属和黄杉属组成了姐妹群。本研究构建的系统发育树中, 同一个属的植物均聚在了一起。本研究测序的蓝粉云杉与欧洲云杉和云杉(*Picea asperata* Mast.)共同组成的枝聚在一起, 而美国黄松与火炬松(*Pinus taeda* L.)亲缘关系最近。

2.4 松科叶绿体基因组结构重排分析

叶绿体基因组重排分析结果表明, 松科叶绿体基因组结构可分为 4 种类型(图 4), 欧洲冷杉(*Abies alba* Mill.)、欧洲云杉、蓝粉云杉及台湾黄杉(*Pseudotsuga sinensis* Dode)的叶绿体基因组结构一致(第 I 种); 银杉、雪松(*Cedrus deodara* Dode)的叶绿体基因组结构重排程度较低(第 II 种); 美国黄松的叶绿体基因组结构重排程度最高(第 III 种); 而蓝粉云杉的叶绿体基因组结构重排程度居于两者之间(第 IV 种)。



CDS: 蛋白质编码区; IGS: 基因间隔区; Intron: 内含子区。
CDS: Coding sequence, IGS: Intergenic spacer.

图 2 蓝粉云杉(A、C、E)与美国黄松(B、D、F)中包含的简单重复序列的类型、分布位置及串联重复序列的分布位置
Distribution and classification of simple and tandem repeat sequences in *Picea pungens* (A, C, E) and *Pinus ponderosa* (B, D, F)

Fig. 2 Distribution and classification of simple and tandem repeat sequences in *Picea pungens* (A, C, E) and *Pinus ponderosa* (B, D, F)

(Roxb. Ex D. Don) G. Don)、铁坚油杉(*Keteleeria davidiana* (C. E. Bertrand) Beissn.)、金钱松(*Pseudolarix amabilis* (J. Nelson) Rehder)、铁杉(*Tsuga chinensis* (Franch.) E. Pritz.)的叶绿体基因组结构一致(第Ⅱ种);欧洲落叶松(*Larix decidua* DC. Mill.)的叶绿体基因组结构单独为一种类型(第Ⅲ种);美国黄松和火炬松的叶绿体基因组结构一致(第Ⅳ种)。第Ⅰ种叶绿体基因组结构与第Ⅱ种之间存在1次重排,即共线性片段(Locally collinear blocks, LCB)3+共线性片段2发生1次反转。第Ⅱ种与第Ⅲ种存在2次重排;第Ⅲ种与第Ⅳ种存在1次重排。对比图3可以看出,系统发育关系较近的类群其叶绿体基因组结构却不一样。

定相同,如松属与银杉属为姐妹群,但是它们的叶绿体基因组结构却分别属于第Ⅳ与第Ⅱ种类型。因此,推测叶绿体基因组结构在松科不同类群之间发生了多次独立的重排。

本研究还对12种松科植物包含的小反向重复序列(Small inverted repeat, sIR)进行了鉴定,结果发现,12种植物均有一段包含 $trnS$ -GCU和 $trnI$ -CAU基因的小反向重复序列, $trnS$ -GCU这段反向重复在台湾黄杉中重复了3次(图4)。云杉属的欧洲云杉、蓝粉云杉还有一段长度为500 bp以上的包含 $trnG$ -UCC的反向重复序列。我们对这些sIR在叶绿体基因组上的位置进行了标注(图4),可以看出, $trnS$ -GCU-sIR、 $trnG$ -UCC-sIR均

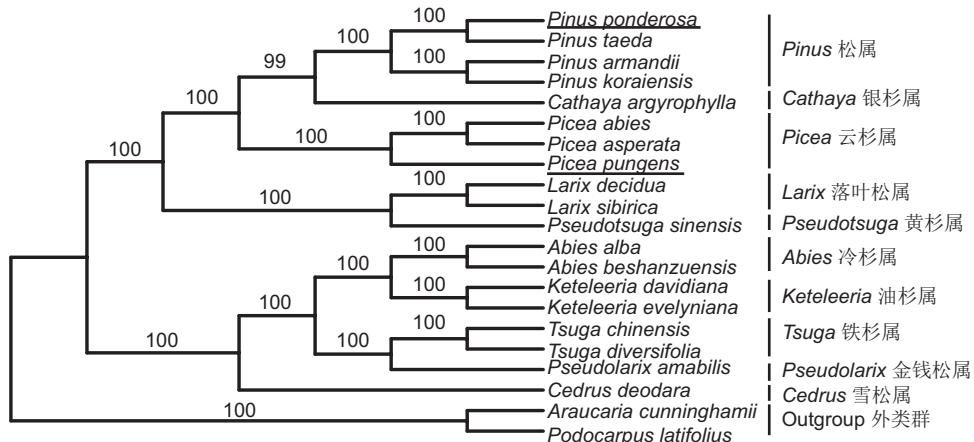
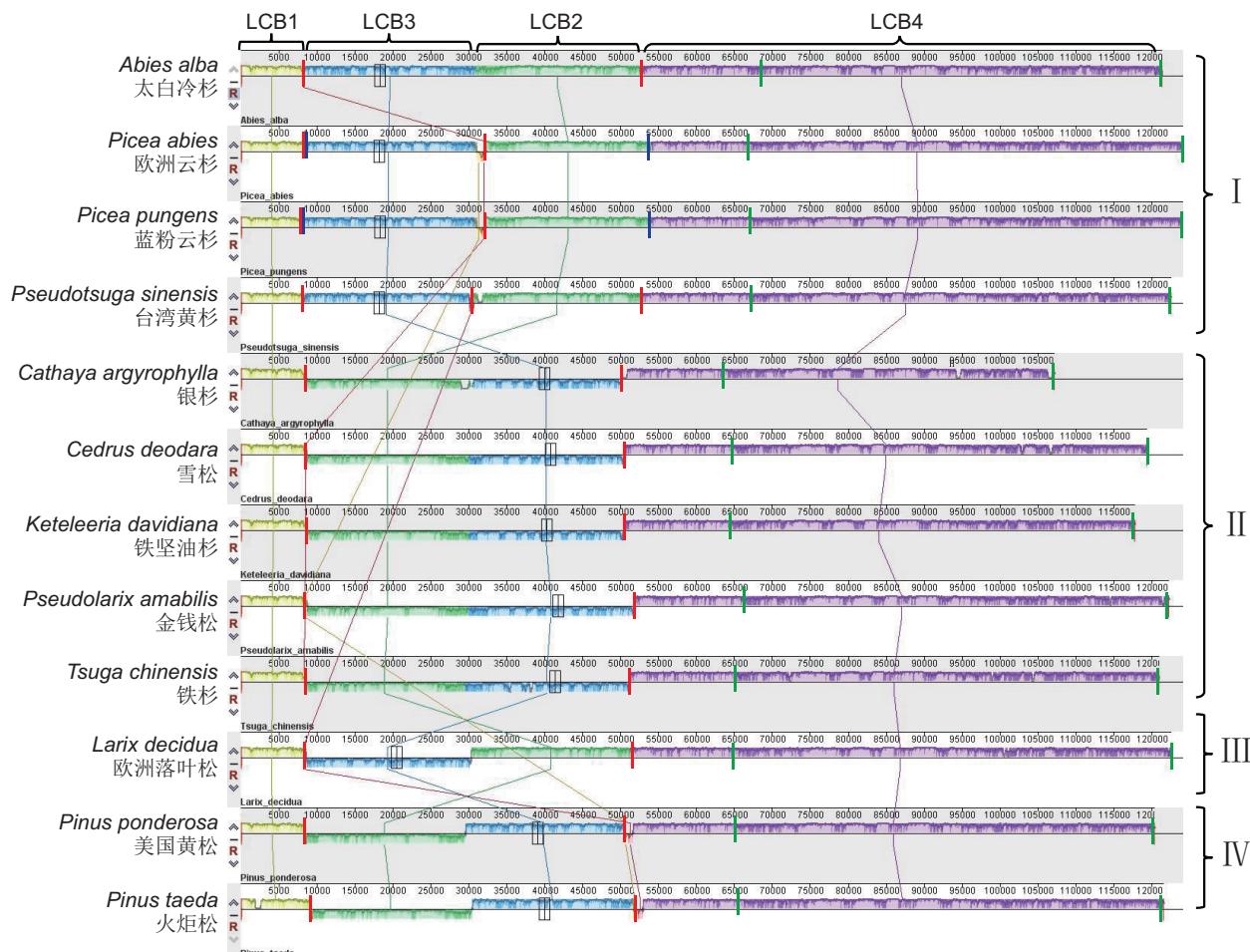


图3 基于73个共有叶绿体基因构建的最大似然树
Fig. 3 Maximum-likelihood tree of 73 chloroplast genes



红色、蓝色、绿色的小矩形条分别代表了包含 *trnS-GCU*、*trnG-UCC* 以及 *trnl-CAU* 基因的小反向重复序列。
Red, blue, and green rectangles represent location of small inverted repeat sequences of *trnS-GCU*, *trnG-UCC*, and *trnl-CAU*.

图4 12种松科植物叶绿体基因组的共线性分析
Fig. 4 Collinearity analysis of chloroplast genomes from 12 Pinaceae species

位于重排的末端，并且位于这两个 sIR 之间的 LCB3、LCB2 在松科叶绿体基因组进化过程中发生了多次反转。说明 *trnS-GCU-sIR*、*trnG-UCC-sIR* 具有介导同源重组的能力。但是，包含 *trnl-CAU* 基因的小反向重复序列位于共线性区块内部，说明 *trnl-CAU*-sIR 没有介导同源重组的能力。因此，松科残留的 *trnS-GCU-sIR*、*trnG-UCC-sIR* 代替了原有的大反向重复区的功能，引起了松科叶绿体基因组结构的多变。

3 讨论

陆生植物的叶绿体基因组结构一般为典型的四分体结构，即包含一段大单拷贝区，小单拷贝区和两段反向重复区。本研究中，蓝粉云杉与美国黄松的叶绿体基因组结构与已公布的松科植物的类似，均丢失了大的反向重复区，并且不再具有四分体结构。但美国黄松与蓝粉云杉叶绿体基因组中分别包含一段长度为 846 和 660 bp 的小反向重复序列。这段 sIR 均包含了完整的 *trnS-GCU* 基因，蓝粉云杉中还包含了完整的 *psbl* 基因。*trnS-GCU-IR* 在其他松科植物中也存在，说明该段反向重复为松科植物共有。并且 *trnS-GCU-IR* 可以替代传统的 IR 介导同源重组，引起叶绿体基因组同分异构体的产生^[21, 28, 35]。本研究对 12 种松科植物的叶绿体基因组重排结果也显示，*trnS-sIR* 可以引起松科各个属之间叶绿体基因组的重排。此外，广义柏类残留的 *trnQ-sIR* 也具有此功能^[36]。说明松科植物叶绿体基因组虽然均丢失了典型的 IR 区，却出现了一些小的 IR 替代原始 IR 的功能，促进了叶绿体基因组结构的遗传多样性。

本研究中，我们以宽叶罗汉松和南洋杉为外类群，使用 73 个共有的叶绿体基因（包括 69 个蛋白编码基因和 4 个 rRNA 基因），基于最大似然法重建了松科的系统发育关系。结果可将松科分为两大主要分支，一支包括松属、银杉属等 5 个属；另一分支包含铁杉属、金钱松属等 5 个属。上述系统发育关系不仅得到了形态解剖数据的支持，如树脂道的位置和数量，种皮中树脂囊的有无，种蛋白的免疫性等^[16, 37, 38]；而且也与叶绿体基因组序列^[10, 21, 39, 40]和转录组数据^[22]的结果一致。银杉属的系统位置一直以来都存在争议，主要包括与松属^[21]、云杉属^[10]或松属+云杉属^[41]互为姐妹群 3

种说法，但是 Ran 等^[22]基于转录组数据确定银杉属与松属互为姐妹群，与本研究结果一致。Ni 等^[42]根据 17 个物种共有的 56 个叶绿体基因进行系统发育重建的结果也与本研究一致，但该研究使用叶绿体全基因组序列进行分析时，发现松属并没有和银杉属聚在一起，而是单独成为一个分支；冷杉属、雪松属和云杉属聚在一起形成一个枝。原因可能是他们的分析中包含基因间区的序列，这部分区域通常受到的选择压力较小，进化速率更快，容易造成系统发育的误差。此外，该研究只分析了松属、冷杉属、雪松属、云杉属 4 个属，采样的不完全也可导致系统发育分析结果的偏差。本研究分析的物种包括了松科 11 个属中的 10 个属，仅不含长苞铁杉属。长苞铁杉 (*Nothotsuga longibracteata* (W. C. Cheng) H. H. Hu ex C. N. Page) 的叶绿体基因组序列虽然已公布（登录号为 NC_037407.1），但我们在分析中发现该物种缺失了 *ccsA*、*chIIB*、*matK*、*psaM*、*psbA*、*psbK*、*rpl32*、*rps15* 共 8 个蛋白编码基因；并且，与其他松科植物不同，长苞铁杉含有 34 239 bp 长的反向重复区，IR 区包含了 29 个蛋白编码基因和 11 个 tRNA 基因^[24]，考虑到该序列比较特殊，因此本研究的分析并没有选用该序列。

综上所述，本研究基于 73 个共有的叶绿体基因得到的松科属间的系统发育关系比较准确可靠。未来还需要测序更多松科植物的叶绿体基因组，尤其是长苞铁杉属，并且需要结合形态、解剖、转录组、叶绿体及核基因组等多种数据对松科植物的系统发育关系进行深入全面的研究。

参考文献：

- [1] Partelli-Feltrin R, Smith AMS, Adams HD, Kolden CA, Johnson DM. Short-and long-term effects of fire on stem hydraulics in *Pinus ponderosa* saplings [J]. *Plant Cell Environ*, 2021, 44(3): 696–705.
- [2] Çetin M, Çobanoğlu O. The possibilities of using blue spruce (*Picea pungens* Engelm.) as a biomonitor by measuring the recent accumulation of Mn in its leaves [J]. *Kastamonu Univ J Eng Sci*, 2019, 5(1): 43–50.
- [3] Montagnoli A, Terzaghi M, Chiatante D, Scippa GS, Lasserre B, Dumroese RK. Ongoing modifications to root system architecture of *Pinus ponderosa* growing on a sloped site revealed by tree-ring analysis [J]. *Dendrochronologia*, 2019, 58: 125650.

- [4] Remke MJ, Hoang T, Kolb T, Gehring C, Johnson NC, Bowker MA. Familiar soil conditions help *Pinus ponderosa* seedlings cope with warming and drying climate [J]. *Restor Ecol*, 2020, 28(S4) : S344–S354.
- [5] Breygina M, Maksimov N, Polevova S, Evmenyeva A. Bipolar pollen germination in blue spruce (*Picea pungens*) [J]. *Protoplasma*, 2019, 256(4) : 941–949.
- [6] Tao J, Chen SG, Qin CY, Li QM, Cai JF, et al. Somatic embryogenesis in mature zygotic embryos of *Picea pungens* [J]. *Sci Rep*, 2021, 11(1) : 19072.
- [7] Daniell H, Jin SX, Zhu XG, Gitzendanner MA, Soltis DE, Soltis PS. Green giant – a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny [J]. *Plant Biotechnol J*, 2021, 19(3) : 430–447.
- [8] Mower JP, Vickrey TL. Structural diversity among plastid genomes of land plants [J]. *Adv Bot Res*, 2018, 85 : 263–292.
- [9] Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 2016, 17(1) : 134.
- [10] Lubna, Asaf S, Khan AL, Jan R, Khan A, et al. The dynamic history of gymnosperm plastomes: insights from structural characterization, comparative analysis, phylogenomics, and time divergence [J]. *Plant Genome*, 2021, 14(3) : e20130.
- [11] Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history [J]. *Am J Bot*, 2018, 105(3) : 291–301.
- [12] Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes [J]. *BMC Evol Biol*, 2014, 14 : 23.
- [13] Neale DB, Wheeler NC. The conifers [M]//Neale DB, Wheeler NC, eds. The Conifers: Genomes, Variation and Evolution. Cham: Springer, 2019: 1–21.
- [14] Farjon A. The kew review: conifers of the world [J]. *Kew Bull*, 2018, 73(1) : 8.
- [15] Stull GW, Qu XJ, Parins-Fukuchi C, Yang YY, Yang JB, et al. Gene duplications and phylogenomic conflict underlie major pulses of phenotypic evolution in gymnosperms [J]. *Nat Plants*, 2021, 7(8) : 1015–1025.
- [16] García Esteban L, de Palacios P, García-Iruela A, García-Fernández F, García-Esteban L, de Vega DG. Comparative wood anatomy in Pinaceae with reference to its systematic position [J]. *Forests*, 2021, 12(12) : 1706.
- [17] Gernandt DS, Reséndiz Arias C, Terrazas T, Dugua XA, Willyard A. Incorporating fossils into the Pinaceae tree of life [J]. *Am J Bot*, 2018, 105(8) : 1329–1344.
- [18] Ran JH, Shen TT, Wang MM, Wang XQ. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms [J]. *Proc Roy Soc B Biol Sci*, 2018, 285(1881) : 20181012.
- [19] Gernandt DS, Magallón S, Geada López G, Zerón Flores O, Willyard A, Liston A. Use of simultaneous analyses to guide fossil-based calibrations of pinaceae phylogeny [J]. *Int J Plant Sci*, 2008, 169(8) : 1086–1099.
- [20] Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies [J]. *Genome Biol Evol*, 2010, 2 : 504–517.
- [21] Sudianto E, Wu CS, Lin CP, Chaw SM. Revisiting the plastid phylogenomics of Pinaceae with two complete plastomes of *Pseudolarix* and *Tsuga* [J]. *Genome Biol Evol*, 2016, 8(6) : 1804–1811.
- [22] Ran JH, Shen TT, Wu H, Gong X, Wang XQ. Phylogeny and evolutionary history of Pinaceae updated by transcriptomic analysis [J]. *Mol Phylogenet Evol*, 2018, 129 : 106–116.
- [23] Ni ZX, Ye YJ, Bai TD, Xu M, Xu LA. Complete chloroplast genome of *Pinus massoniana* (Pinaceae): Gene rearrangements, loss of *ndh* genes, and short inverted repeats contraction, expansion [J]. *Molecules*, 2017, 22(9) : 1528.
- [24] Chen L, Li LN, Yang GD, Qian HR, Li MZ. Characterization of the complete chloroplast genome sequence of *Tsuga longibracteata* W. C. Cheng (Pinaceae) [J]. *Conserv Genet Resour*, 2019, 11(2) : 117–120.
- [25] Li ZH, Zhu J, Yang YX, Yang J, He JW, Zhao GF. The complete plastid genome of Bunge's pine *Pinus bungeana* (Pinaceae) [J]. *Mitochondrial DNA Part A*, 2016, 27(4) : 2971–2972.
- [26] Li ZH, Qian ZQ, Liu ZL, Deng TT, Zu YM, et al. The complete chloroplast genome of Armand pine *Pinus armandii*, an endemic conifer tree species to China [J]. *Mitochondrial DNA Part A*, 2016, 27(4) : 2635–2636.
- [27] Li JF, Xu B, Yang Q, Liu ZL. The complete chloroplast genome sequence of *Picea schrenkiana* (Pinaceae) [J]. *Mitochondrial DNA Part B*, 2020, 5(3) : 2191–2192.
- [28] Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM. Comparative chloroplast genomes of pinaceae: insights into the mechanism of diversified genomic organizations [J]. *Genome Biol Evol*, 2011, 3 : 309–319.
- [29] Sullivan AR, Schiffthaler B, Thompson SL, Street NR, Wang XR. Interspecific plastome recombination reflects ancient reticulate evolution in *Picea* (Pinaceae) [J]. *Mol Biol Evol*, 2017, 34(7) : 1689–1701.
- [30] Ranade SS, García-Gil MR, Rosselló JA. Non-functional

- plastid *ndh* gene fragments are present in the nuclear genome of Norway spruce (*Picea abies* L. Karsch): insights from in silico analysis of nuclear and organellar genomes [J]. *Mol Genet Genomics*, 2016, 291(2): 935–941.
- [31] Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data[J]. *Nucleic Acids Res*, 2017, 45(4): e18.
- [32] Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes[J]. *Plant Methods*, 2019, 15(1): 50.
- [33] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era[J]. *Mol Biol Evol*, 2020, 37(5): 1530–1534.
- [34] Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement[J]. *PLoS One*, 2010, 5(6): e11147.
- [35] Tsumura Y, Suyama Y, Yoshimura K. Chloroplast DNA inversion polymorphism in populations of *Abies* and *Tsuga* [J]. *Mol Biol Evol*, 2000, 17(9): 1302–1212.
- [36] 李佳, 姜斌, 苏应娟, 王艇. 白豆杉的叶绿体基因组结构与系统进化分析[J]. 植物科学学报, 2021, 39(1): 5–13.
Li J, Jiang B, Su YJ, Wang T. Structure and phylogenetic analysis of the *Pseudotaxus chienii* (W. C. Cheng) W. C. Cheng chloroplast genome [J]. *Plant Science Journal*, 2021, 39(1): 5–13.
- [37] 姚庆渭, 黄鹏成. 松科各属种子的研究[J]. 南京林业大学学报(自然科学版), 1980(1): 28–42.
Yao QW, Huang PC. A research of the seeds of genera in Pinaceae [J]. *Journal of Nanjing Forestry University (Natural Sciences Edition)*, 1980(1): 28–42.
- [38] Esteban LG, De Palacios P. Comparative wood anatomy in Abietoideae (Pinaceae)[J]. *Bot J Linn Soc*, 2009, 160(2): 184–196.
- [39] Kim SC, Lee JW, Lee MW, Baek SH, Hong KN. The complete chloroplast genome sequences of *Larix kaempferi* and *Larix olgensis* var. *koreana* (Pinaceae)[J]. *Mitochondrial DNA Part B*, 2018, 3(1): 36–37.
- [40] Hao ZD, Cheng TL, Zheng RH, Xu HB, Zhou YW, et al. The complete chloroplast genome sequence of a relict conifer *Glyptostrobus pensilis*: comparative analysis and insights into dynamics of chloroplast genome rearrangement in cupressophytes and Pinaceae [J]. *PLoS One*, 2016, 11(8): e0161809.
- [41] Gernandt DS, Holman G, Campbell C, Parks M, Mathews S, et al. Phylogenetics of extant and fossil Pinaceae: methods for increasing topological stability[J]. *Botany*, 2016, 94(9): 863–884.
- [42] Ni ZX, Ye YJ, Xu M, Xu LA. Comparison among three methods for obtaining chloroplast genome sequences from the conifer *Pinus massoniana*[J]. *Genomics*, 2020, 112(3): 2459–2466.

(责任编辑: 周媛)