

油松叶绿体DNA间隔序列特点及系统学意义*

李 焜^{1**} 柴宝峰²

(¹山西财经大学环境经济学院 太原 030006)

(²山西大学生物技术研究所 太原 030006)

摘 要 对全国分布区12个天然油松种群进行取样,测定了叶绿体DNA *trnT-trnL*、*trnS-trnG*、*trnL-trnF*基因间隔序列,分析了序列碱基组成特点及在松科系统分类中的意义.测序结果显示,3个片段碱基序列均富含A/T,长度依次为448 bp、636 bp和421 bp,所测片段在种群水平上十分保守,没有发现具有种内鉴别意义的碱基变异.但此片段适合于松科属间和种间的系统学分析,运用PHYLP软件构建松科植物的邻接系统树,支持松科分为两个亚科的分类系统,拓扑结构表明松属与银杉属的关系最近.由于叶绿体DNA序列长度适中,扩增和测序较为容易,适合作为研究松科植物系统进化较为理想的分子标记.图2表2参24

关键词 油松; cpDNA间隔序列; 系统进化; 分子标记

CLC S791.255.01 : Q949

Characteristics and Systematic Significance of Chloroplast DNA Sequences of *Pinus tabulaeformis**

LI Cui^{1**} & CHAI Baofeng²

(¹Faculty of Environment Economics, Shanxi University of Finance and Economic, Taiyuan 030006, China)

(²Institute of Biotechnology, Shanxi University, Taiyuan 030006, China)

Abstract The samples of *Pinus tabulaeformis* were collected from 12 populations occurring in whole China within its chloroplast DNA distribution range, and their noncoding sequences of chloroplast DNA *trnT-trnL*, *trnS-trnG* and *trnL-trnF* were sequenced. The characteristics of base compositions and the significance of the sequences in taxonomy of Pinaceae were analyzed by using the sequence data. The results indicated: (1) The three cpDNA sequences, 448 bp, 636 bp and 421 bp length, were characterized with a high A+T content; and (2) the alignments of the three cpDNA sequences in *P. tabulaeformis* were identical and conservative, suggesting that the DNA sequences were not suitable for analyzing the intraspecific relationship of *P. tabulaeformis*, but they were found useful for analyzing interspecific and intergeneric relationships within the family. The NJ (neighbour-joining) tree constructed by using PHYLP software package supported that Pinaceae could be divided into two subfamilies, revealing the closer relationship between *Pinus* and *Cathaya*. Because of their moderate length and easy amplification, the three DNA sequences were considered as ideal molecular markers to study the evolutionary systematics of Pinaceae. Fig 2, Tab 2, Ref 24

Keywords *Pinus tabulaeformis*; cpDNA spacer sequence; systematic evolution; molecular marker

CLC S791.255.01 : Q949

松科 (Pinaceae Lindl.) 含10属约235种,是现存裸子植物中最大的一个类群,也是最为古老的一个类群.为了阐明松科植物的系统发育关系,国内外学者从形态学、解剖学、古植物学等方面进行了很多研究,认为松科属间界限明确,但属间和种间演化关系存在分歧,其原因可能是地质历史时期松科曾是很庞大的类群,中生代以后其中的许多类群相继灭绝,近代松科植物只是少数喜温性植物的后裔,属间存在很大的间断性^[1].由于松科植物种间基因交流频繁,分子进化和形态进化不平行,作为分类依据的枝条和叶缘等在进化中发生了很大的变化,利用传统方法构建松科演化的系列过程较为困难.

随着生物技术的快速发展及向生命科学领域的强力渗透,分类学家越来越重视分子数据在系统发育中的应用,期望以此来澄清分类与进化中的遗留问题^[2-4].叶绿体DNA (cpDNA) 非编码区不参与基因的转录翻译过程,在功能上限制较少,比编码区能提供更多的遗传变异的信息,适用于不同阶元的系统发育、谱系地理和遗传结构的研究^[5-8];但是由于同一cpDNA片段在不同植物类群中进化速率存在差异,所以选择合适的DNA序列是一个至关重要的问题.十分稳定保守、没有碱基变异的序列片段不能为系统学研究提供信息;有碱基变异的序列也不一定理想解决所存在的问题,如ITS序列在松科属间和属内均有相当的变异,尤其长度变异显著,在松科中的变异幅度为1 550~3 125 bp^[9],信息含量丰富,理论上讲,通过测序—排序—比对,就可针对性地解决目的问题,但是过长的片段给排序带来困难,限制了其在松科属间系统发育中的应用;5S rDNA基因长度过短,而且极度保守,不能提供足够的系统学信息^[10];也有一些片段长度适中,但其结构比较复杂^[11],且A/T含量太高,PCR扩增就存

收稿日期: 2009-02-12 接受日期: 2009-11-30

*山西省回国留学人员科研项目 (No. 200846) 和国家自然科学基金项目 (No. 30770294) 资助 Supported by the Science Foundation for Return Scholars of Shanxi Province, China (No. 200846) and the National Natural Science Foundation of China (No. 30770294)

**通讯作者 Corresponding author (E-mail: aihl@sxu.edu.cn)

在一定困难。

本研究通过筛选油松 (*Pinus tabulaeformis* Carr.) 叶绿体基因间隔序列引物, 寻找适合于松科植物系统分析的目的基因片段, 分析了扩增产物的序列组成特征, 并与GenBank中松科其它属植物的相应序列一起进行了比较, 推断了松科植物各属间的演化关系, 探讨了叶绿体基因间隔序列应用于松科植物系统进化分析的可行性。

1 材料与方 法

1.1 试验材料

样品收集自全国12个不同的油松种群, 于山西省境内的历山、紫团山、灵空山、芦芽山和关帝山各选择1个油松种群, 另外选取地理距离较远的, 分别位于甘肃天水、青海互助、内蒙宁城、四川理县、辽宁千山、河北雾灵山和陕西洛南其它7个种群作为研究对象, 每个种群选取1~3个代表性植株, 取针叶提取DNA, 进行目的基因的扩增和测序分析。取样地自然概况见表1。

表1 取样地自然概况

Table 1 Natural conditions and locations of all the populations of *P. tabulaeformis* studied in this study

样地 Population code	经度/纬度 Longitude/ latitude	样地 Population code	经度/纬度 Longitude/ latitude	样地 Population code	经度/纬度 Longitude/ latitude
山西历山 (LS) Mt. Lishan, Shanxi	111°58'E/35°25'N	山西关帝山 (GD) Mt. Guandi, Shanxi	111°30'E/37°47' N	四川理县 (SC) Lixian, Sichuan	103°15'E/31°23' N
山西灵空山 (LK) Mt. Lingkong, Shanxi	112°13'E/36°37' N	甘肃天水 (GS) Tianshui, Gansu	105°57'E/34°33' N	辽宁千山 (LN) Mt. Qianshan, Liaoning	123°04'E/40°55' N
山西紫团山 (ZT) Mt. Zituan, Shanxi	113°26'E/35°53' N	青海互助 (QH) Huzhu, Qinghai	102°00'E/37°18' N	陕西洛南 (SX) Luonan, Shaanxi	110°15'E/34°11' N
山西芦芽山 (LY) Mt. Luya, Shanxi	111°53'E/38°35' N	内蒙宁城 (NM) Ningcheng, Inner Mongolia	118°27'E/41°23' N	河北雾灵山 (HB) Mt. Wuling, Hebei	117°17'E/40°29' N

1.2 试验方法

采用改进的 CTAB 法提取油松针叶总DNA^[12]。扩增引物参考Taberlet和Hamilton的设计^[13~14], 筛选出3对适合于扩增油松叶绿体目的基因的引物, 分别为: *trnT-trnL* (F: 5'- CATTACAAATGCGATGCTCT-3', R: 5'- TCTACCGATTTTCGCCATATC-3'), *trnS-trnG* (F: 5'-GCCGCTTTAGTCCACTCAGC-3', R: 5'-GAACGAATCACACTTTTACCAC-3'), *trnL-trnF* (F: 5'-GGTTCAAGTCCCTCTATCCC-3', R: 5'-ATTTGAACCTGGTGACACGAG-3')。PCR扩增反应液总体积为15 μL体系, 包括10×PCR 缓冲液 1.5 μL, 2.0 mmol/L dNTP 1.5 μL, 2.0 μmol/L的引物各1.5 μL, TaqDNA聚合酶0.5 U, DNA模板50 ng, 加ddH₂O至15 μL。PCR扩增条件: 94 °C预变性5 min; 94 °C 30 s, 55 °C 1 min, 72 °C 1.5 min, 35个循环; 72 °C延伸7 min。用低熔点的琼脂糖回收PCR扩增产物的目的片段, 纯化回收产物, 与pGEM-T easy载体 (Promega公司) 连接, 转化DH5α感受态细胞, 进行蓝白斑筛选。提取阳性克隆的质粒进行鉴定, 选取含目的基因片段的质粒进行测序。

序列对比采用BioEdit 软件完成^[15], 并加以适当的手工校对。排好的序列用PHYLIP (3.66) 软件包进行系统发育分析^[16], 邻接法 (Neighbor-joining) 分析时, 选取裸子植物苏铁 (*Cycas revolute*) 和银杏 (*Ginkgo biloba*) 为外类群^[17], 各节点

的置信度以自举检验分析, bootstrap进行100次重复。

2 结果

2.1 油松cpDNA基因间隔区序列特征

测定了12个油松种群叶绿体 *trnT-trnL*、*trnS-trnG*、*trnL-trnF* 基因间隔序列, 以BioEdit软件进行对位排列, 结果显示这3个片段十分保守, 序列在种群间具100%的同源性。cpDNA *trnT-trnL* 序列全长448 bp, 核苷酸的组成为A (149)、T (121)、G (108)、C (70), A+T共270个碱基, 占60.27%; cpDNA *trnS-trnG* 序列长度为636 bp, 核苷酸的组成为A (210)、T (202)、G (114)、C (110), A+T共412个碱基, 占64.78%; cpDNA *trnL-trnF* 序列全长421 bp, 核苷酸的组成为A (131)、T (146)、G (67)、C (77), A+T共277个碱基, 占65.79%。3个片段均富含A/T, 许多非编码区和假基因序列都具此性质^[18], 油松叶绿体基因间隔序列符合非编码区碱基组成的一般规律。

本试验选取的12个种群, 基本覆盖天然油松的主要分

布区, 测序结果没有发现具有种内鉴别意义的变异位点。通过GenBank查询, 下载与叶绿体 *trnT-trnL* 片断相关的松科植物15种 (表2)。

表2 来源于GenBank的物种名称和序列注册号
Table 2 Species and their accession numbers from GenBank

种名 Species	序列号 Access No.	种名 Species	序列号 Access No.
<i>Pinus thunbergii</i>	D17510	<i>Pseudotsuga sinensis</i>	AF440504
<i>P. tropicalis</i>	AB097065	<i>P. menziesii</i>	AF440503
<i>P. strobus</i>	DQ010640	<i>Abies kawakamii</i>	DQ112528
<i>P. bungeana</i>	DQ010641	<i>Keteleeria evelyniana</i>	EF395417
<i>Picea meyeri</i>	DQ010623	<i>Tsuga mertensiana</i>	EF395447
<i>P. glauca</i>	DQ010605	<i>Cedrus deodara</i>	DQ983638
<i>Cathaya argyrophylla</i>	DQ010639	<i>Ginkgo biloba</i> (outgroup 1)	AY145323
<i>Larix kaempferi</i>	DQ087863	<i>Cycas revolute</i> (outgroup 2)	AF531181
<i>L. occidentalis</i>	DQ087866		

油松叶绿体 *trnT-trnL* 的间隔序列长度为448 bp, 为了判断序列间的同源性, 将无法排序的部分舍弃, 只比较相对保守的部分。油松可比较序列为413 bp, 共同进行分支分析的其它物种的序列长度介于395~419 bp, 显示了 *trnT-trnL* 基因片段在松科属间的长度多态性。为了便于比较, 各属分别选取一个物种, 共9种植物, 分属于松科9属, 序列对比结果如图1。该DNA片段在许多位点发生插入、缺失和碱基的替换, 尤其是一些有规律的碱基变化, 反映着物种的进化历史。

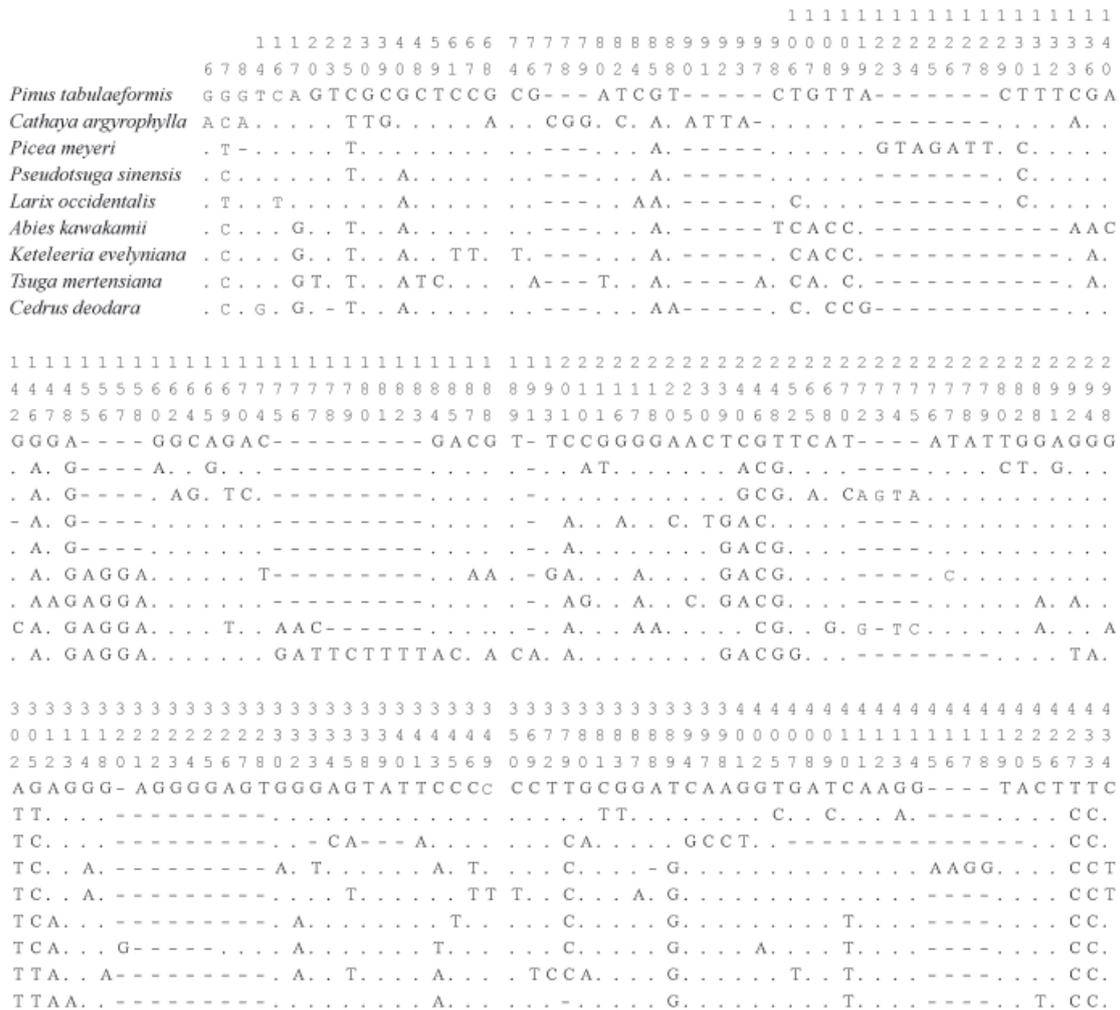


图1 松科9种植物的trnT-trnL基因间隔序列差异比较

Fig. 1 The sequences alignment of the cpDNA trnT-trnL regions among 9 species of Pinaceae and their variable sites based on the alignment 数字表示序列位点, 黑点表示同油松序列一致 The numbers indicate the sequence sites, and the dots show the identical bases to that of P. tabulaeformis

2.2 基于trnT-trnL序列构建邻接系统树

序列对比结果显示, 松属中的华山松 (Pinus armandii)、白皮松 (P. bungeana) 和红松 (P. koraiensis) 序列具有100%同源性, 云杉属 (Picea) 中的青杆 (P. wilsonii) 和白杆 (P. meyeri) 100%同源. 构建系统树时, 序列同源的物种选取一种, 16个松科植物trnT-trnL的邻接系统树见图2. 从图中可以看出, (1) 松科9属植物分为两个主要分支: 一个分支是冷杉属 (Abies)、油杉属 (Keteleeria)、雪杉属 (Cedrus)、铁杉属 (Tsuga) 的物种, 其中油杉属与冷杉属更近缘; 另一分支由两个亚分支组成, 一个亚分支是黄杉属 (Pseudotsuga)、落叶松属 (Larix) 的物种, 一个亚分支是松属、银杉属 (Cathaya)、云杉属的物种, 而且松属与银杉属的关系近于与云杉属的关系; (2) 松属的5个种形成一个单系群, 获得很高的bootstrap值, 强烈支持松属种间的亲缘关系. 分属于黄杉属、云杉属和落叶松属的物种也聚为一支, 且得到100%的有力支持, 来自同一属的物种均能很好地组成一个单系群. 进一步验证松科属间具有明显的界限, 同时也表明此序列片段

适用于松科种以上系统发育研究; (3) 系统树同时显示出松属是与古老的外种群距离最远的一个类群.

2.3 叶绿体DNA trnS-trnG和trnL-trnF序列的系统学意义

叶绿体DNA trnS-trnG和trnL-trnF片段显示了与trnT-trnL共同的特征, 物种水平上同样十分稳定. 通过GenBank查询, 与trnS-trnG片段相关的松科植物分属于松科3属共14种; 与trnL-trnF相关的仅有松属植物9个物种, 由于3个片段不能作联合分析, 在此仅作简单分析. 油松叶绿体trnS-trnG的可比较序列为620 bp, 一起分析的其它属物种序列长度变化范围较大, 而且随着分类阶层的提高, trnS-trnG序列的差异越明显. 本研究涉及到的物种中, 凡属同一物种的个体均具有完全一致的序列, 此片段序列长度多态性主要体现在属间. 基于L. Peroffskyana作为外种群构建的系统邻接树, 同属物种同源性很高. 油松trnL-trnF全序列长为421 bp, 舍弃无法排序的部分, 可比较序列为388 bp, 共同进行分析的其它物种的序列长度介于382-393 bp之间, 显示了trnL-trnF片段在松科属间的

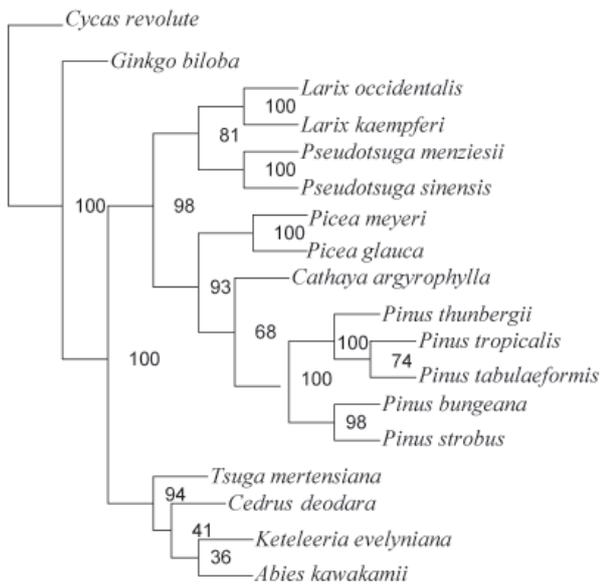


图2 基于松科16种植物的 $trnT-trnL$ 间隔序列运用PHYLIP (3.66) 构建的邻接树

Fig. 2 Phylogenetic relationships of 16 species in Pinaceae 分枝上的数字表示100次重复的自展值,用苏铁和银杏作为外种群 *C. revolute* and *G. biloba* were used as outgroups. The neighbour-joining tree was computed by PHYLIP based on 100 replicates. The bootstrap supporting values were indicated in the branches

长度多态性(未列出).

3 讨论

3.1 油松cpDNA间隔序列的系统学意义

油松是中国北方广泛分布的针叶树种,天然分布范围介于北纬 $31^{\circ}00' \sim 41^{\circ}00'$ 、东经 $101^{\circ}30' \sim 124^{\circ}25'$. 本研究所采样品跨约21个经度带、10个纬度带,3对引物所测序列均无任何碱基变异,在物种水平稳定保守,但在种间和属间存在丰富的碱基差异,说明所测序列的保守性是相对的. 3个cpDNA片段适合作为物种以上的遗传结构分析和系统学研究.

表型性状由于受环境因素的影响可能呈现不稳定性^[19],松科植物表型进化与分子进化不同步,依据表型进行研究的可靠性和有效性会受到影响. 而且松科许多植物核型一致性较高,染色体的变化主要表现在分子水平的结构变异和序列不同,对于松科植物的起源、物种间的亲缘关系等问题的深入了解,很难通过常规的遗传标记方法得到解决. 油松叶绿体 $trnL-trnF$ 、 $trnT-trnL$ 和 $trnS-trnG$ 序列片段长度适中,扩增和测序都很容易,引物在同科不同属间也具有很高的通用性,所以认为这3个片段是研究松科植物属间和部分种间系统学关系的较为理想的分子标记.

3.2 松科植物各属间的关系

《中国植物志》根据长、短枝及叶等形态学性状将松科植物划分为3个亚科. 松亚科 (Pinoideae) 仅松属一个属,落叶松亚科 (Laricoideae) 含雪松属、金钱松属 (*Pseudolarix*) 和落叶松属,冷杉亚科 (Abietoideae) 含油杉属、冷杉属、铁杉属、云杉属、银杉属、黄杉属^[20]. 李楠在综合了松科属间36个分类性状,筛选出有序进化性状20个,认为松科分为3个类群较为合适^[21]. Hu等根据解剖学、传粉学和形态学等特

征将松科分为两个类群,一个类群包括云杉属、落叶松属、黄杉属、银杉属和松属,另一个类群包括铁杉属、冷杉属、雪松属、油杉属和金钱松属^[22]. 这一结果也得到3个基因片段 ($marK$ 、 $nad5$ 、 $4CL$) 联合分析的支持^[17]. 我们的聚类结果也认为松科分为这两个类群是合适的,但是亚分支中属间关系有所不同,如图2中松属~银杉属~云杉属亚分支中,松属与银杉先聚为一支,再与云杉相聚,即油松与银杉的关系近于与云杉的关系. 比较松属的胚胎发育、花粉形态、染色体数目和核型等方面的特征,均与银杉属接近^[23],其它属间关系同已往的分类也有较好的吻合,说明松科植物叶绿体 $trnT-trnL$ 间隔序列包含着松科进化的分子信息.

油松cpDNA $trnT-trnL$ 、 $trnS-trnG$ 和 $trnL-trnF$ 间隔序列在一定程度上反映了松科系统关系,但由于只涉及整个基因组很少的一部分,作为系统进化分析有一定的局限性;且由于3个片段不能进行联合分析,使其信息含量降低. 只有收集更多属的物种,设计更多的引物,将多个序列拼接起来进行联合分析,使构建的基因树更真实地反映物种树^[24],对揭示松科物种的形成及适应机制,确定各属间的演化关系可能更有价值.

致谢 本研究的实验工作在山西大学化学生物学与分子工程教育部重点实验室完成,感谢梁爱华教授给予热情指导. 部分样品得到山西大学黄土高原研究所的支持. 在山西省境内野外调查和采样时得到当地林业局领导和专业技术人员的全力帮助.

References

- Li N (李楠). Studies on the geographic distribution, origin and dispersal of the family Pinaceae Lindl. *Acta Phytotaxon Sin* (植物分类学报), 1995, **33** (2): 105~130
- Kan XZ, Wang SS, Ding X, Wang XQ. Structural evolution of nrDNA ITS in Pinaceae and its phylogenetic implications. *Mol Phylogenet Evol*, 2007, **44**: 765~777
- Xiang QP, Xiang QY, Guo YY, Zhang XC. Phylogeny of *Abies* (Pinaceae) inferred from nrITS sequence data. *Taxon*, 2009, **58** (1): 141~152
- Qiao CY, Ran JH, Li Y, Wang XQ. Phylogeny and biogeography of *Cedrus* (Pinaceae) inferred from sequences of seven paternal chloroplast and maternal mitochondrial DNA regions. *Ann Bot*, 2007, **100**: 573~580
- Gao X, Zhu YP, Wu BC, Zhao YM, Chen JQ, Hang YY. Phylogeny of *Dioscorea* sect. *Stenophora* based on chloroplast $matK$, $rbcl$ and $trnL-F$ sequences. *J Syst Evol*, 2008, **46** (3): 315~321
- Havill NP, Campbell CS, Vining TF, LePage B, Bayer RJ, Donoghue MJ. Phylogeny and biogeography of *Tsuga* (Pinaceae) inferred from nuclear ribosomal its and chloroplast DNA sequence data. *Syst Bot*, 2008, **33** (3): 478~489
- Chen KM, Abbott RJ, Milne RI, Tian XM, Liu JQ. Phylogeography of *Pinus tabulaeformis* Carr. (Pinaceae), a dominant species of coniferous forest in northern China. *Mol Ecol*, 2008, **17**: 4276~4288
- Xue XM (薛喜枚), Wang YH (王芋华), Luo JX (罗建勋), Li CY (李春阳). Application of different DNA molecular markers in genetic diversity of *Picea* Diert. *Chin J Appl Environ Biol* (应用与环境生物学报), 2004, **10** (6): 803~810
- Liston A, Robinson WA, Oliphant JM, Alvarez-Buylla ER. Length

- variation in the nuclear ribosomal DNA internal transcriber spacer region of non-flowering seed plants. *Syst Bot*, 1996, **21** (2): 109~120
- 10 Halanych KM. 5S ribosomal RNA sequences inappropriate for phylogenetic reconstruction. *Mol Biol Evol*, 1991, **8** (2): 249~253
- 11 Kan XZ (阚显照), Qiao CY (乔才元), Gao H (高卉), Zhu GP (朱国萍). Secondary structure analysis on the nrDNA interna transcribed spacers and 5.8S in *Nothotsuga longibracteata*. *Acta Laser Biol Sin* (激光生物学报), 2007, **16** (3): 327~333
- 12 Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*, 1987, **19**: 11~15
- 13 Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol Biol*, 1991, **17**: 1105~1109
- 14 Hamilton MB. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol Ecol*, 1999, **8** (2): 521~523
- 15 Hall TA. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser*, 1999, **41**: 95~98
- 16 Felsenstein J. PHYLIP version 3.66. <http://evolution.genetics.washington.edu/phylip.html> [Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA], 2006
- 17 Wang XQ, Tank DC, Sang T. Phylogeny and divergence times in Pinaceae: Evidence from three genomes. *Mol Biol Evol*, 2000, **17** (5): 773~781
- 18 Galtier N, Piganeau G, Mouchiroud D, Duret L. GC-content evolution in mammalian genomes: The biased gene conversion hypothesis. *Genetics*, 2001, **159**: 907~911
- 19 Hu QP (胡启鹏), Guo ZH (郭志华), Li CY (李春燕), Ma LY (马履一). Advance at phenotypic plasticity in plant responses to abiotic factors. *Sci Silv Sin* (林业科学), 2008, **44** (5): 135~142
- 20 郑万钧, 傅立国. 中国植物志. 北京: 科学出版社, 1978
- 21 Li N (李楠), Fu LG (傅立国), Zhu ZD (朱政德). Studies on systematic of the family Pinaceae Lindl. *Bull Bot Res* (植物研究), 1996, **16** (1): 32~45
- 22 Hu YS, Napp-Zinn K, Winne D. Comparative anatomy of seed-scales of female cones of Pinaceae. *Bot Jahrb Syst*, 1989, **111** (1): 63~85
- 23 王伏雄. 银杉生物学. 北京: 科学出版社, 1990
- 24 Zou XH (邹新慧), Ge S (葛颂). Conflicting gene trees and phylogenomics. *Acta Phytotaxon Sin* (植物分类学报), 2008, **46** (6): 795~807