



## 技术主题

## Technology Feature

# 冷杉基因组 DNA 遗传标记芯片的构建与分析

饶龙兵<sup>✉</sup>

中国林业科学研究院亚热带林业研究所, 富阳, 311400

✉ 通讯作者: raolb@163.com   ✉ 作者

分子植物育种, 2011 年, 第 9 卷, 第 101 篇   doi: 10.5376/mpb.cn.2011.09.0101

收稿日期: 2011 年 08 月 14 日

接受日期: 2011 年 09 月 07 日

发表日期: 2011 年 09 月 15 日

这是一篇采用 Creative Commons Attribution License 进行授权的开放取阅论文。只要对本原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

引用格式(中文):

饶龙兵, 2011, 冷杉基因组 DNA 遗传标记芯片的构建与分析, 分子植物育种(online) Vol.9 No.101 pp.1726-1734 (doi: 10.5376/mpb. cn.2011.09.0101)

引用格式(英文):

Rao, 2011, Creating and Analyzing of Biochip of Genetic Markers on Genomic DNA of *Abies*, Fenzi Zhiwu Yuzhong (online) (Molecular Plant Breeding) Vol.9 No.101 pp.1647-1653 (doi: 10.5376/mpb.cn.2011.09.0101)

**摘要** 分子遗传标记技术正在从以测量 DNA 片段大小的凝胶电泳技术向以杂交为基础的高通量芯片技术过渡。本文主要研究了冷杉基因组 DNA 提取方法、基因组 DNA 的复杂性和复杂性降低方法、芯片的文库构建、芯片质量控制和有效标记参数选取、基因组 DNA 遗传标记芯片的制作方法, 并将冷杉基因组 DNA 遗传标记芯片在冷杉遗传多样性检测中做了初步应用。研究表明: 用改良的 CTAB 法提取冷杉基因组 DNA 质量较好, 满足生物芯片对 DNA 质量要求; 利用 *Pst* I + *Taq* I + *Mse* I 酶切组合和 *Pst* I 引物扩增得到克隆片段数量与质量均较好; 试验共挑选 4 608 克隆点(标记点)制作芯片, 得到 821 个有效标记点, 得率约 17.8% ; 用该芯片试分析 8 个供试样品遗传多样性水平, 结果显示 8 个样品多态位点百分率为 100%, 多态信息指数为 0.40, 利用得到的标记位点可以有效开展 8 个样品的遗传关系分析, 样品遗传距离位置和大小在重复样品中支持率为 100%, 表明结果可靠。本试验将生物芯片技术和遗传标记技术结合起来, 成功制作了能用于遗传分类和遗传多样性检测的冷杉基因组 DNA 遗传标记芯片, 该技术还可以广泛应用于遗传作图、分子辅助育种、物种分类和进化等研究。

**关键词** 冷杉; 基因组 DNA; 遗传多样性; 遗传标记芯片技术

## Creating and Analyzing of Biochip of Genetic Markers on Genomic DNA of *Abies*

Rao Longbing<sup>✉</sup>

The Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang, 311400, P.R. China

✉ Corresponding author, raolb@163.com;   ✉ Authors

**Abstract** Molecular genetic marker technologies are undergoing a transition from gel-based assays measuring DNA fragment sizes to hybridization-based microarray assays with high throughput. The methods of extracting and enzyme digestion of genomic DNA, creating microarray library, the quality and parameters of biochip were discussed in this paper, and eight samples also were tested by this biochip. The results showed that the quality of extracted DNA were good for test, the microarrays library creating by combined enzyme of *Pst* I + *Taq* I + *Mse* I and primer *Pst* I were also fit for *Abies*. 4 608 clones were selected and 821 markers were effective, the productivity of effective markers was about 17.8%. The genetic diversity, genetics distance of eight samples were also analyzed, repeatability of samples in genetic distance had 100% consistency. This parameters above showed that genetic markers biochip were made successfully, and this technology could be used in the field of researching genetic mapping, taxonomy and evolution, assistant breeding, genetic diversity analyzing, and so on.

**Keywords** *Abies*; Genomic DNA; Genetic diversity; DArT

## 研究背景

遗传标记(Genetic markers)是指与目标性状紧密连锁, 同该性状共同分离且易于识别的可遗传的等位基因变异。目前, 应用较为广泛的遗传标记有形态标记、细胞标记、生化标记和分子标记。DNA 分子标记具有标记数量多、分布广、遍及整个基因组、多态性高、遗传稳定、不受环境限制的特点而

成为一种重要的遗传标记技术, 广泛应用于遗传育种、基因组作图、基因定位与克隆、遗传多样性检测、系统进化、分子分类与物种亲缘关系鉴别等方面(Patrick et al., 1997; Sasaki and Burr, 2000; Qian et al., 2001; Jeung et al., 2007; 王云生等, 2007)。目前常用几种标记如 RFLP、AFLP、RAPD、SSR、ISSR 等(周延清等, 2008)。这些技术应用有部分优点, 也

有一定局限性。如 RFLP 技术需要利用放射性同位素, 实验成本较高; RAPD 技术稳定性不太理想; SSR 技术开发需要一定的 DNA 序列背景信息, 引物开发难度大; ISSR 由于是随机引物筛选, 费事费力; AFLP 由于得到的是酶切混合片段, 不能将不同来源的相同长度大小 DNA 片段分开, 技术相对比较粗放等。这些技术离理想的遗传标记技术还有一段差距, 而且这些技术都基于毛细管电泳或凝胶电泳技术, 整个检测过程自动化程度低, 通量低, 产出率低, 精度低。

基因芯片技术是近年发展起来的一项新的生物技术(陈忠斌等, 2005)。芯片技术是将得到的探针 DNA 用芯片点样仪点于芯片载体上, 然后将待检样品和芯片杂交, 最后利用检测仪判断杂交信号大小。芯片样品点直径通常在 25~500  $\mu\text{m}$ , 通常 1  $\text{cm}^2$  可点样 2 500 个点阵, 一般 1 张载玻片可容纳 30 000 左右 DNA 探针, 芯片技术具有明显的高通量的特点, 而且芯片技术是基于分子杂交的检测技术, 与常规凝胶电泳技术相比具有高精度的特点, 另外检测样品需求微量。基因芯片技术已经在生命科学的许多领域得到广泛应用, 如在医药、动植物疾病诊断、遗传代谢分析、食品安全检测和侦测等方面都有应用(陈忠斌等, 2005)。目前制作的基因芯片绝大部分属表达谱芯片, 芯片点样的 DNA 主要是 cDNA 或合人工成的 DNA, 芯片检测的目的主要集中在新基因发现、基因功能研究、基因诊断、DNA 序列分析、SNP 检测等研究方面(梁秀彬等, 2002)。如果能将芯片技术的自动化、高通量、高精度特点和遗传标记技术结合起来, 将基因组 DNA 制作成遗传标记的检测芯片, 用于遗传分类、进化、遗传作图、辅助育种和遗传多样性检测等方面, 必将极大促进遗传标记技术的发展。但基因芯片技术与遗传标记技术如何结合, 如何应用, 应用效果如何等问题, 目前国内外在此方面应用研究报道较少, 仅限于位于澳大利亚的 Diversity Arrays Technology (DArT) 技术实验室开展相关研究(Jaccoud et al., 2001; Wenzl et al., 2004; Xia et al., 2005; Wittenberg et al., 2005; Akbari et al., 2006; Yang et al., 2006)。

分布在我国南方的 5 种冷杉植物百山祖冷杉(*A. besganzuensis*) (吴鸣翔定名发表)、元宝山冷杉(*A. yuanpaoshanensis*) (吕庸浚和傅立国定名发表)、资源冷杉(*A. ziyuanensis*) (傅立国和莫新礼定名发表)、梵净山冷杉(*A. fanjinshanensis*) (黄威廉定名发表)和

大院冷杉(*A. dayuanensis*) (刘起銜定名发表)均为我国的一级重点保护植物(《国家重点保护野生植物名录》)和濒危植物(《中国植物红皮书》), 其中百山祖冷杉目前野生植株仅剩 3 棵, 被世界保护联盟维护物种生存委员会(SSCT)列为世界最濒危的 12 种植物之一。这 5 种冷杉植物的发现否定了以往亚热带的低、中山无冷杉分布的观点, 被认为是植物地理学和植物学上的一大发现。由于冷杉属植物的分布受第四纪冰川时期的影响, 因此上述冷杉植物是研究我国第四纪冰川时期植物区系和气候变迁的“活化石”, 对研究古气候、古地理和植物区系具有重要意义。目前这 5 种冷杉分类存在争议, 现有分类的种内和种间遗传分化、遗传多样性不明, 群体保护地位重要性不清。要开展有关研究, 必须采用分子标记技术分析多态位点。多态位点数量具有重要作用, 位点少, 信息可靠性低, 位点多可靠性高。多态位点的获得和检测样本大小及检测技术有关。在有关冷杉研究中遇到特别是百山祖冷杉, 现有植株太少, 满足不了一般群体取样样本数, 常规凝胶电泳技术位点检测能力低, 通量低, 基于目前状况作者拟建立遗传标记芯片体系用于冷杉保护遗传学研究。

该技术原理和基本步骤为: 首先提取高质量的基因组 DNA, 然后将基因组 DNA 通过限制性内切酶酶切, 产生不同大小的酶切片段, 接着将酶切片段通过添加选择性碱基引物扩增, 降低酶切片段总数量, 再将这些选择性扩增后的酶切片段通过 TOPO 载体克隆, 很多个克隆点构成一个芯片文库。此部分操作为芯片文库构建, 不同的材料采用不同的酶切和引物扩增方法可以得到不同的芯片文库。克隆产物通过培养, PCR 扩增, 扩增后的 DNA 再经过纯化, 最后将该纯化的克隆 DNA 点样在生物芯片上, 制成一张用于遗传标记检测的基因组 DNA 芯片。当要利用该芯片对待测样品检测时, 先将被检材料酶切, 酶切后的 PCR 扩增物采用荧光标记, 然后将标记物和芯片杂交, 最后通过杂交信号检测判断检测样品与标记点同源性差异, 杂交信号强, 则同源性高, 否则就低。

由于不同材料的基因组 DNA 复杂性不同, 以及限制性酶切位点也存在很大差异, 因此用不同的限制性酶切, 以及采用不同添加选择性碱基引物扩增时产生的 DNA 片段大小、数量、以及片段在基因组上分布位置的均匀性都有差异, 最后得

到的有效标记数也不同。要得到高质量的基因组 DNA 芯片和得到符合一定质量参数的有效标记数, 就需要研究合适 DNA 提取方法、合适的酶切和引物组合试验, 笔者根据此思路开展了基因组 DNA 提取、多种不同酶切组合和不同引物扩增实验, 成功制作了用于冷杉遗传多样性检测的芯片, 并做初步应用分析。

## 1 结果与分析

### 1.1 冷杉基因组 DNA 提取

利用改良的 CTAB 法从冷杉叶片中提取冷杉基因组 DNA, 经检测样品  $D_{260\text{ nm}}$  和  $D_{280\text{ nm}}$  光度值, 其比值都在 1.9 左右; 风干后 DNA 溶液清亮, 无黏稠感, 实验结果说明提取的 DNA 纯度较好, 样品中基本无多糖、蛋白质、和酚类等杂质。

将提取的冷杉叶片 DNA 与生物芯片 DNA 工作缓冲液混合, 室温静置 3 h, 利用琼脂糖凝胶电泳检测分析。如果 DNA 提取质量不好, 或不适合 DNA 芯片工作缓冲液环境, 则凝胶电泳后 DNA 可能检测不出或裂解。本实验检测结果如图 1 所示。由图可以看出, 电泳样品孔均不发亮, 各泳道 DNA 主带清晰, 基本无弥散带, 说明提取的 DNA 完整性较好, 纯度也较高。电泳图谱显示有少许 RNA 存在, 但 RNA 在本实验系统中不影响实验结果。

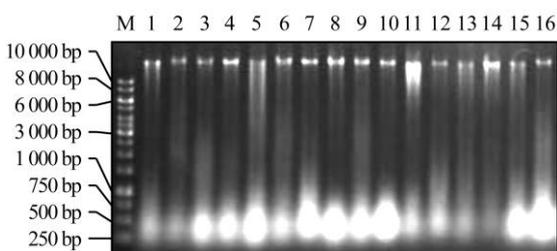


图 1 冷杉基因组 DNA 电泳检测图

注: M: GeneRuler 1 kb DNA ladder marker; 1~16: 16 个冷杉样品

Figure 1 The electrophoresis pictures of genomic DNA  
Note: M: GeneRuler 1 kb DNA ladder marker; 1~16: Sixteen different fir samples

### 1.2 冷杉基因组 DNA 不同组合酶切和不同引物扩增效果

采用 7 种不同的限制性内切酶组合酶切冷杉基因组 DNA 后, 用无选择性碱基 *Pst* I 引物扩增, 扩增电泳图如图 2, 用添加选择性碱基 CAG 的 *Pst* I 引物扩增图如图 3。从图 2 中可看出, 酶切组合 *Pst* I + *Taq* I + *Mse* I (图 2C, 泳道 17~24) 酶切片段大小分布在 750 bp 左右, 而且分布比较均匀集中, 分

子量比较大的片段和比较小的片段相对较少。从该酶切组合的 *Pst* I 引物添加选择性碱基 CAG 扩增图 3 分析, 所在泳道 13、泳道 14 没有明显的条带, 说明该酶切组合能较均匀切开冷杉基因组 DNA。综合图 2 和图 3 分析认为 *Pst* I + *Taq* I + *Mse* I 酶切组合效果最好。 *Pst* I + *Bst*NI、 *Pst* I + *Taq* I 的酶切效果次之, 其它几种酶切组合不甚理想, 表现为片段过大或均匀度差, 出现明显条带。

### 1.3 冷杉芯片文库盘质量分析

基因组 DNA 经过酶切扩增后采用 TOPO 载体克隆, 多个克隆构成芯片文库。在文库构建克隆挑选过程中, 理论应该在文库盘每格(孔)位置上有一个克隆点, 实际操作过程中有可能漏放或重放或挑选了空克隆, 这样可能会影响芯片质量。为了保证文库盘质量, 挑选的克隆点需要先检测, 一般文库盘中错误克隆点必须要小于 5%, 最好控制在 2% 以内水平。本实验文库盘克隆点检测部分电泳图样如图 4 示, 从左致右 1~33 泳道, 第 1~16 泳道为 16 个克隆点, 第 17 泳道为间隔, 第 18~33 为另外 16 个克隆点。从此电泳图可看出, 挑选出的克隆片段大小主要约在 500~1 000 bp 左右, 少数克隆在 1 000~1 500 bp, 而且每检测孔道均有 1 个克隆, 满足芯片文库盘的质量要求。

### 1.4 冷杉芯片质量分析

当芯片文库盘作好后, 就可以将文库盘上克隆点点样于芯片上, 制作成遗传标记检测芯片。芯片上样点的均匀程度是衡量芯片质量的重要指标。制作出一批芯片后, 将采用部分样品和芯片杂交, 通过杂交信号及参数可以判断芯片质量。

从表 1 可知, 各张芯片的样点基本比率、中值、均值比率相关系数为 0.866~0.976, 说明样点点样及杂交后的均匀性很好, 质量可靠。信噪比是判断检测系统性能和芯片杂交稳定性的重要参数, 理想检测仪器的信噪比都在一定范围。从表 1 中可看出参考通道信噪比为 79.017~92.024, 杂交后检测通道信噪比为 20.786~32.827, 符合检测系统的理想信噪比值。综合样点的 3 种比率相关系数及信噪比参数可以判断, 采用该方法得到的芯片质量比较好, 可以用于有关冷杉检测分析工作。

### 1.5 有效标记得率分析

本试验测试中共随机挑选有效克隆 4 608 个点样制作遗传标记检测芯片, 经 8 样品杂交检测数据

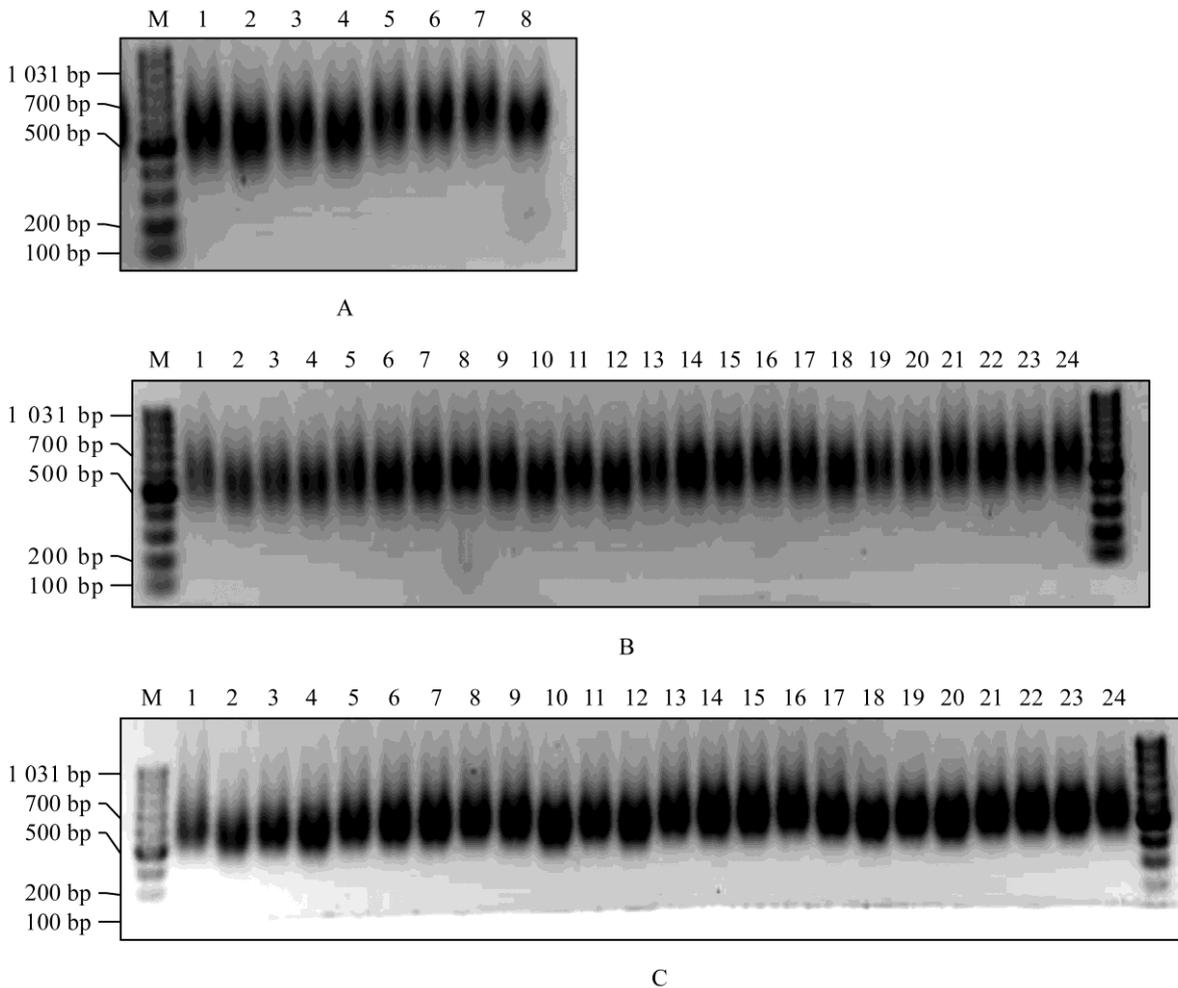


图 2 *Pst* I 未添加选择性碱基引物扩增图

注: M: GeneRuler 100 bp DNA ladder marker; 采用 7 种不同的限制性内切酶组合酶切冷杉基因组 DNA 后, 用未添加选择性碱基的 *Pst* I 引物扩增图, 每种酶切处理 8 个冷杉样品; A: 1~8: *Pst* I +*Taq* I; B: 1~8: *Pst* I +*Bst*NI; 9~16: *Pst* I +*Alu* I; 17~24: *Pst* I +*Hae*3; C: 1~8: *Pst* I +*Hha* I; 9~16: *Pst* I +*Msp* I; 17~24: *Pst* I +*Taq* I +*Mse* I

Figure 2 The picture after PCR using primer *Pst* I

Note: M: GeneRuler 100 bp DNA ladder marker; Using 7 different restriction endonucleases *Hind*III fir combination of genomic DNA was added, selectively base *Pst* I primers for amplification of graph, each enzyme processing 8 fir samples; A: 1~8: *Pst* I +*Taq* I; B: 1~8: *Pst* I +*Bst*NI; 9~16: *Pst* I +*Alu* I; 17~24: *Pst* I +*Hae*3; C: 1~8: *Pst* I +*Hha* I; 9~16: *Pst* I +*Msp* I; 17~24: *Pst* I +*Taq* I +*Mse* I

分析后得到重复性 100%、召回率 85%、P 值大于 75%、Q 值大于 75%的标记点共 821 个, 有效标记点约 17.8%。

### 1.6 8 样品 821 个标记位点遗传多样性及遗传聚类分析

将 8 个不同的样品, 按上述方法经杂交检测后取标记参数值均为 P 值、Q 值均大于 75%, 重复性为 100%, 召回率大于 85%的标记点统计分析。分析得到标记点的多态性位点百分比率 P 为 100%, 8 样品多态性信息含量指数为 0.4。采用 DARwin5 软件对 8 样品聚类分析, 如图 5 所示。用该方法可

以有效分析样品间的遗传距离或亲缘关系, 而且用不同荧光标记的同一重复样品亲缘关系聚类支持率为 100%, 说明该方法切实可行, 而且重复性、可靠性均较理想。

## 2 讨论

遗传标记技术广泛应用于分子育种、图谱构建、基因定位与克隆、物种分类与进化、多样性检测等。目前常用的检测技术还都基于凝胶电泳技术, 具有产出率低、精度低、通量低的共性。基因芯片技术具有检测系统微型化、样品需求的微量化、检测的高效化及高通量化的优势和特点, 从本实验结果看,

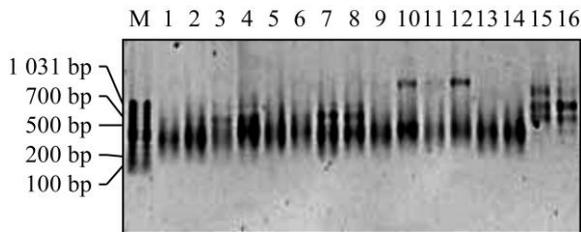


图3 *Pst* I 添加选择性碱基 CAG 引物扩增图  
 注: M: GeneRuler 100 bp DNA ladder marker; 采用 7 种不同的限制性内切酶组合酶切冷杉基因组 DNA 后, 用添加选择性碱基 CAG 的 *Pst* I 引物扩增图, 每种处理 2 个冷杉样品; 1, 2: *Pst* I +*Taq* I; 3, 4: *Pst* I +*Bst*NI; 5, 6: *Pst* I +*Alu* I; 7, 8: *Pst* I +*Hae*3; 9, 10: *Pst* I +*Hha* I; 11, 12: *Pst* I +*Msp* I; 13, 14: *Pst* I +*Taq* I +*Mse* I; 15, 16: 对照样品  
 Figure 3 The picture after PCR using primer *Pst* I +CAG and seven different combination of incision enzyme  
 Note: M: GeneRuler 100 bp DNA ladder marker; Using 7 different restriction endonucleases *Hind*III fir genomic DNA combination, with added selective base CAG *Pst* I primers for amplification of graph, each kind of processing 2 fir samples; 1, 2: *Pst* I +*Taq* I; 3, 4: *Pst* I +*Bst*NI; 5, 6: *Pst* I +*Alu* I; 7, 8: *Pst* I +*Hae*3; 9, 10: *Pst* I +*Hha* I; 11, 12: *Pst* I +*Msp* I; 13, 14: *Pst* I +*Taq* I +*Mse* I; 15, 16: Control sample

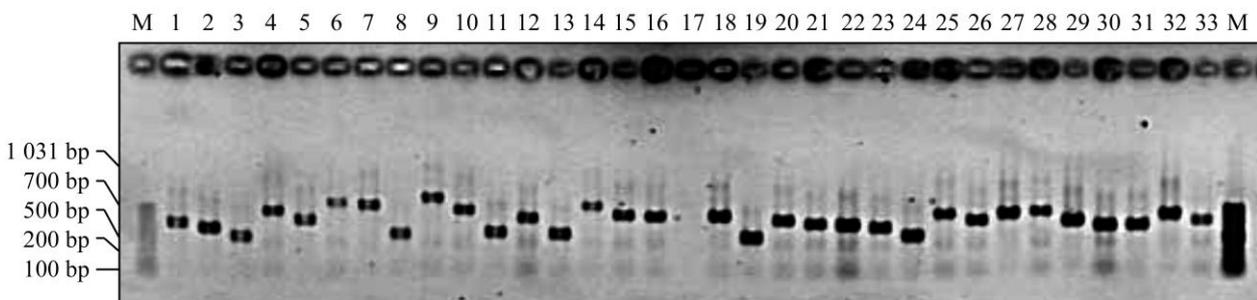


图4 文库构建时克隆检测图  
 注: M: GeneRuler 100 bp DNA ladder marker; 1~16, 18~33 为 32 个不同克隆, 17 泳道为间隔  
 Figure 4 Testing clone for creating library  
 Note: M: GeneRuler 100 bp DNA ladder marker; 1~16, 18~33 showed the picture of 32 different clones, lane 17 was empty

本试验开展了冷杉遗传标记芯片技术研究, 从各项芯片质量、检测参数和位点参数数据分析, 得到的芯片遗传标记芯片满足质量要求, 有效标记位点约 17.8%, 利用通过此方法得到的标记分析标记点的多态性位点百分比率为 P 为 100%, 8 个样品 *Nei* 基因遗传多样性指数为 0.4, 而且样品间遗传距离的支持率以及重复样品的位置关系都比较理想, 说明该方法切实可行, 重复性, 可靠性均较理想, 可以应用于冷杉遗传分类、遗传多样性检测中。

本文进行的 8 个冷杉样品有关分析只是为了供遗传标记芯片质量检测和分析试用, 8 个样品并不能较好地代表某冷杉群体遗传多样性水平, 冷杉群体

芯片技术可以很好的和标记技术结合起来, 充分发挥芯片技术高通量、高精度特点以及标记技术应用广泛的特点。利用此技术具有明显的优点: (1)标记源于基因组 DNA 某部分未知片段, 不依赖 DNA 背景信息, 即使是一个未知物种也可开展工作, 该优势在许多遗传负荷大, 前期遗传背景研究较少的濒危木本植物多样性研究中尤其重要。(2)高通量, 标记数多、检测效率高。通常一张芯片可容纳  $3 \times 10^4 \sim 4 \times 10^4$  左右标记, 而且样品处理也可采用批量化, 检测效率很高, 常规凝交电泳标记无法达到的。(3)利用该标记技术产生的标记数几乎可无限量扩展, 直至覆盖整个基因组, 而且要想增加标记数也比较容易; 标记数越多, 遗传多样性检测精度越高。(4)得到的文库具有一定通用性, 一旦建成一个比较好的文库后, 相近物种的样品检测将变得非常容易。该技术非常适合并未完成全基因组测序的物种开展分类进化、辅助育种、图谱构建、种质真假鉴定等研究, 应用潜力广泛。

间分类和遗传分化问题探讨另文待发(饶龙兵, 2009)。

DArT 技术实验室已成功将 DArT 技术运用在大麦(*Hordeum*)、小麦(*Triticum*)、水稻(*Oryza*)、苹果(*Malus*)、木薯(*Manihot*)、拟南芥(*Arabidopsis thaliana*)、木豆(*Cajanus*)、高粱(*Sorghum*)、西红柿(*Lycopersicon esculentum*)、甘蔗(*Saccharum*)等的分子育种、亲源分析与种质鉴定、图谱构建、遗传多样性研究。

### 3 材料与方法

#### 3.1 冷杉基因组 DNA 提取

材料: 新鲜冷杉叶片, 材料包含有元宝山冷杉

表 1 冷杉遗传标记芯片质量参数表

Table 1 The table of qualitative parameters of microarray slides

冷杉样品名称	2 种颜色标记	芯片片号与检测通道	样点基本比率相关系数	样点均值比率相关系数	样点中值比率相关系数	检测通道平均信噪比	参考通道平均信噪比
Sample	Target	Slide	The basic rate of sample correlation coefficient	Sample mean correlation coefficient	Sample median correlation coefficient	Sample median ratio correlation coefficient	Reference channel average SNR
fa16	fa16	1-543	0.941	0.946	0.931	27.666	89.497
	fa16	1-633	0.940	0.946	0.930	32.288	89.497
fa9	fa9	2-543	0.970	0.975	0.958	24.432	84.494
	fa9	2-633	0.968	0.974	0.957	30.165	84.494
fa6	fa6	3-543	0.976	0.984	0.965	24.888	84.166
	fa6	3-633	0.976	0.983	0.969	32.827	84.166
fa5	fa5	4-633	0.964	0.973	0.951	27.832	77.783
	fa5	4-543	0.968	0.975	0.956	26.567	77.783
fa26	fa26	5-543	0.963	0.969	0.949	26.978	92.024
	fa26	5-633	0.949	0.958	0.932	22.225	92.024
fa3	fa3	6-633	0.968	0.976	0.957	24.379	79.017
	fa3	6-543	0.972	0.978	0.958	24.152	79.017
fa2	fa2	7-543	0.889	0.895	0.876	27.079	87.871
	fa2	7-633	0.884	0.888	0.866	20.786	87.871
fa4	fa4	8-543	0.953	0.961	0.942	23.638	79.875
	fa4	8-633	0.950	0.960	0.941	24.199	79.875

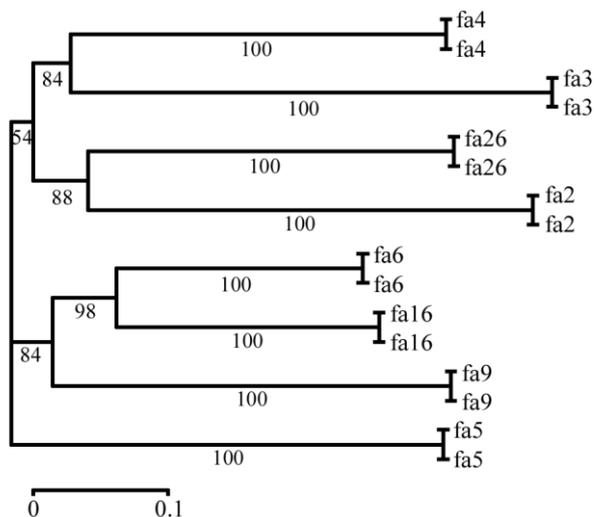


图 5 8 试测样品遗传亲源关系图

Figure 5 The genetic distance of eight samples

(*A. yuanpaoshanensis*), 取自广西元宝山; 梵净山冷杉(*A. fanjinshanensis*), 取自贵州梵净山; 百山祖冷杉(*A. besganzuensis*), 取自浙江庆元百山祖; 资源冷杉(*A. ziyuanensis*), 取自广西资源县、和大院冷杉(*A. dayuanensis*), 取自湖南炎陵; 以及日本冷杉(*A. firma S.et*), 取自江西庐山等。除日本冷杉来源人工引种林外, 其它几种冷杉均来自该种冷杉天然野生群体。

提取方法: ①取粉末 0.3 g; ②在 10 mL 离心管中加 3 mL EB, 100  $\mu$ L  $\beta$ -巯基乙醇, 涡旋混匀后, 在 5 000 g 离心力下离心 10 min(室温), 去上清液; ③再加入 1 mL EB, 2 mL LB 和 200  $\mu$ L 5% Sarkosyl, 颠倒离心管, 混匀; ④ 60 $^{\circ}$ C 恒温水浴 1.5 h, 水浴期间摇动离心管 3~4 次; ⑤加入 24: 1 的氯仿: 异戊醇 3 mL, 混匀静置 10 min; 再 5 000 g 室温离心 10 min; ⑥将上清液转于新的离心管中, 重复操作步骤⑤、⑥一次; ⑦向上清液中加入 2 倍体积的 -20 $^{\circ}$ C 无水乙醇, 然后轻轻颠倒离心管, 混匀, 静置 15 min, 自然沉淀, 去上清。⑧向沉淀中加入 2 mL -20 $^{\circ}$ C 70%乙醇, 轻轻震荡, 使 DNA 重悬于乙醇中。接着将重悬液转入 2.5 mL 离心管, 500 rpm 离心 10 s, 去上清。⑨最后将离心管斜放于离心架上, 自然风干后加 TE 保存; 实验中 EB 配方为(500 mL): ph 8.0 100 mmol/L Tris-Hcl; ph 8.0 50 mmol/L EDTA; 灭菌, 使用前加 2.0 g 亚硫酸氢钠, 及 10 g PVP, 31.88 g 山梨醇; LB 配方为(500 mL): ph 8.0 200 mmol/L Tris-Hcl、ph 8.0 50mmol/L EDTA、58.45 g NaCl、15 g CTAB; TE 配方为: ph 8.0 10 mmol/L Tris-Hcl、ph 8.0 1 mmol/L EDTA。

DNA 纯度与完整性检测方法: 基因组 DNA 纯度检测: DNA 样品按 30 倍稀释后, 在 260 nm 和

280 nm 处测定其紫外吸收值, 计算出  $D_{260\text{ nm}}/D_{280\text{ nm}}$  的比值。基因组 DNA 的完整性检测: 取 2  $\mu\text{L}$  DNA 提取液, 加 3  $\mu\text{L}$  上样缓冲液(内含 0.25% 溴酚蓝, 40% 蔗糖), 用 0.7% 琼脂糖凝胶(内含 0.5 mg/L 的 EB(溴化乙锭))电泳, 电泳缓冲液为 1 $\times$ TAE, 电压 110 V, 电泳 40 min 后, 紫外凝胶成像拍照(图 1)。

### 3.2 冷杉基因组 DNA 不同组合酶切和不同引物扩增效果

材料: 上述几种冷杉基因组 DNA。

方法: 采用 7 种不同内酶切组合酶切基因组 DNA。扩增引物 1: 采用 *Pst* I 酶切对应引物, 未添加选择性碱基扩增, 每种不同酶切组合处理 8 个冷杉样品, 见电泳图 2; 扩增引物 2: 采用 *Pst* I 酶切对应引物再添加+CAG 3 个选择性碱基, 每种不同酶切组合处理 2 个冷杉样品, 电泳图 3。操作步骤: 取冷杉基因组 DNA 1  $\mu\text{L}$  加入酶切反应混合体系进行酶切和接头连接。(酶切反应混合体系包括: *Pst* I 及不同的酶切组合, *Pst* I 人工接头, BSA, 酶切缓冲液, T<sub>4</sub>DNA 连接酶, 去离子水)。酶切时间如下: 1: *Pst* I +*Bst*NI 先 37 $^{\circ}\text{C}$  2 h 后 60 $^{\circ}\text{C}$  2 h; 2: *Pst* I +*Taq* I 先 37 $^{\circ}\text{C}$  2 h 后 60 $^{\circ}\text{C}$  2 h; 3: *Pst* I +*Hae*3 37 $^{\circ}\text{C}$  3 h; 4: *Pst* I +*Alu* I 37 $^{\circ}\text{C}$  3 h; 5: *Pst* I +*Msp* I 先 37 $^{\circ}\text{C}$  2 h 后 80 $^{\circ}\text{C}$  20 min; 6: *Pst* I +*Hha* I 先 37 $^{\circ}\text{C}$  2 h 后 80 $^{\circ}\text{C}$  20 min; 7: *Pst* I +*Taq* I +*Mse* I 先 37 $^{\circ}\text{C}$  3 h 再 60 $^{\circ}\text{C}$  2 h 后 65 $^{\circ}\text{C}$  20min。酶切后产物添加 dNTP、DNA 聚合酶及 *Pst* I 对应的引物、PCR 缓冲液, 进行 PCR 扩增, 扩增条件为 94 $^{\circ}\text{C}$  1 min、60 $^{\circ}\text{C}$  40 s、72 $^{\circ}\text{C}$  50 s 共 30 个循环。扩增后的产物用 1.2% 凝胶电泳检测。

### 3.3 基因组 DNA 芯片文库构建

芯片文库构建材料: 芯片文库材料来源和库容大小可以根据不同研究目的构建不同大小的文库。如果建库材料来源比较广, 而且库容比较大则可应用于多种材料的样品检测。本试验建库材料包含有元宝山冷杉(*A. yuanpaoshanensis*)、梵净山冷杉(*A. fanjinshanensis*)、百山祖冷杉(*A. besganzuensis*)、资源冷杉(*A. ziyuanensis*)、大院冷杉(*A. dayuanensis*)、以及日本冷杉(*A. firma* S.et), 主要目的是为以后相关材料检测准备。

芯片文库构建方法: 基因组 DNA 提取采用前法。取上述建库材料各样品的基因组 DNA 1  $\mu\text{L}$  加入酶切反应混合体系, 60 $^{\circ}\text{C}$  温育 90 min 进行酶切和

接头连接。酶切后的产物在扩增反应混合液中(包括 PCR 缓冲液, dNTP, DNA 聚合酶及 *Pst* I 引物)进行 PCR 扩增, 扩增条件 94 $^{\circ}\text{C}$  1 min, 60 $^{\circ}\text{C}$  40 s, 72 $^{\circ}\text{C}$  50 s 共 30 个循环。将不同来源(材料, 样品)的 PCR 扩增产物各取少量等量混合, 然后利用 TOPO-TA 克隆法克隆。再将克隆转化物转到培养基中培养增殖, 增殖后的克隆菌落在扩增反应液中直接 PCR 扩增(扩增反应液包括: 扩增缓冲液, dNTP, DNA 聚合酶及 M13 前向引物和 M13 后向引物)。扩增条件 95 $^{\circ}\text{C}$  4 min, 57 $^{\circ}\text{C}$  35 s, 72 $^{\circ}\text{C}$  1 min, 共 35 个循环。PCR 扩增后的产物烘箱干燥, 然后再用乙醇洗涤两次, 最后加入芯片点样缓冲液, 利用芯片点样仪点样载玻片上, 每样品 4 次重复, 点样后的载玻片上经过 DNA 与基片高温再水合、紫外线交联固定及干燥处理, 制成一批遗传标记检测芯片, 置于芯片盒避光保存备用。

### 3.4 试测样品处理及与芯片杂交

试测材料来源及处理方法: 待检 8 份材料来源于梵净山冷杉野生群体。处理方法: 提取样品基因组 DNA, 经 *Pst* I +*Taq* I +*Mse* I 限制性内切酶酶切, 再采用和前面文库构建相同的方法 PCR 扩增、样品 DNA 沉淀、洗涤等处理。将处理过的样品 DNA 与荧光标记反应液混合标记样品 DNA (荧光标记混合液包括 dNTP, Klenow 酶, 随机引物, 缓冲液)。每一待测样品均用两种不同荧光标记, 作为标记重复。

芯片与试测样品的杂交及检测: 将标记好的试测样品 DNA 和杂交缓冲液混合, 再使样品 DNA 高温变性, 变性好的混合液用吸头滴加到芯片上, 盖上盖玻片, 置于一定湿度和温度的杂交仪中, 杂交 12~20 h。杂交后的芯片采用杂交清洗液依次洗涤。洗涤后的杂交芯片用离心机甩干, 再干燥, 最后用芯片扫描仪扫描, 进行数据分析。由于每个克隆在芯片上有 4 次重复, 且每一待测样品均采用 2 种荧光标记, 故实际上每个样品信号值有 8 次杂交重复。

芯片点样采用 BioRobotics MicroGrid 点样系统; 芯片扫描采用 Tecan LS Reloaded 扫描系统。

### 3.5 芯片数据分析

信号数据分析参数及计算方法如下: 为了检测芯片点样质量和杂交信号质量, 需要将每个杂交点样(图 6C 为多个杂交样点杂交信号实际照片), 采用分割法再分成多个信号检测点(图 6D 所示)分析, 每个信号检测点将有一个检测值, 将这些被分割的检

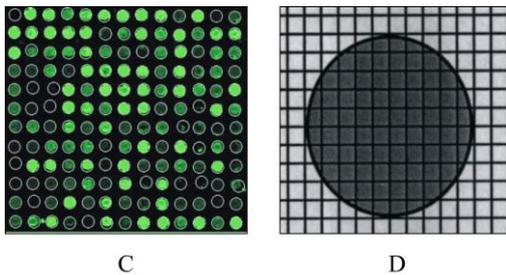


图 6 C: 多个杂交点芯片杂交图; D: 1 个杂交点信号检测分割图

Figure 6 C: Part shows many spots after hybridization; D: Part shows one spot was divided into many tested parts

测点的检测值按照一定数学统计分析后将得到某杂交样点的信号值。芯片点样的质量和杂交信号质量可由以下参数判断:

①信号强度比率: 定义为  $\log(\text{杂交信号强度} - \text{杂交背景强度} / \text{参考信号强度} - \text{参考背景强度})$ ;

②信号强度基本比率: 指某一样点被分割成许多检测点后, 所有信号切割位点的众数的中值;

③信号强度中数比率: 指某一样点被分割成许多检测点后, 所有信号切割位点的中位数比率。

$\text{RatioMed} = \log(\text{杂交信号强度中值} - \text{杂交背景强度中值} / \text{参考信号强度中值} - \text{参考背景强度中值})$ ;

④信号强度平均数比率(RatioAvg)指某一样点被分割成许多检测点后, 所有信号切割位点的平均数比,  $\text{RatioAvg} = \log(\text{杂交信号强度平均数} - \text{杂交背景强度平均数} / \text{参考信号强度平均数} - \text{参考背景强度平均数})$ ;

⑤信噪比: A 参考通道信噪比( $\text{SigToNoiseRef} = (\text{参考通道强度} - \text{参考通道背景值})$ 的中值/参考通道背景值标准差); B 检测通道信噪比( $\text{SigToNoise-Target} = (\text{检测通道强度} - \text{检测通道背景值})$ 的中值/参考通道背景值标准差)。

多态位点分析参数及计算方法如下: 在进行同一批芯片和多个样品杂交检测分析时, 芯片上的标记点和不同的样品杂交会显现不同的杂交信号, 将杂交点信号经数学分析后会出现 1、0、-(不能判断) 3 种判断结果。为了提高实验检测精度, 必须设置同一位点和同一样品的多次杂交重复, 结合召回率、P、Q 值可判断某标记位点质量。

①重复性 Reproducibility: 重复性为判读结果相同次数占总重复次数的比率, 或为每位点判读结果相同的概率。如果所有重复都出现判读结果相同, 则重复性为 100%。  $\text{Repro} = \sum_{i=0}^N (\text{N} - \text{dif}_i) / \text{Nrep}$

(Nrep 重复次数,  $\text{dif}_i$  为不同的次数);

②召回率 Call rate: 召回率指某个位点上能判读得到的赋分值(1, 0)的样本数占总样本数的百分比。其相对值为该位点不能判读为(1, 0)的样本数(即缺失值样本数)占总样本数的比率。该值的高低代表该位点判读情况好坏的程度, 可作为该位点是否选取的依据之一, 如果该位点在多个样品中都不能判读出 0、1 结果, 则该位点不被选取。  $\text{Callrate} = 100 \times (\sum_i \text{score}_i / \text{N})$  (N 总样本数,  $\text{score}_i$  能赋分的样本数);

③Q 值: Q 值为某位点 0、1 两分后该位点强度比率中最低值组(0 组)信号数值变异与强度比率最高值组(1 组)信号数值变异之和占总变异的百分数。如果 Q 值越大说明分组后的方差占总方差比重越大, 说明 0、1 分组效果越好。Q 值的强度比率值选取可以是强度基本比率、中数比率和平均数比率三种中任 1 组值, 因此 Q 值有 3 种, 取最大值作为该位点的 Q 值。  $Q = \max(i \times (x - x_{lo_i})^2 + (N - i) \times (x - x_{hi_i})^2) / \text{N} \times \text{var}(x) \times 100$  ( $0 \leq i < N$ , X 为 N 个点的平均值,  $x_{lo_i}$  为第 i 组最低值平均,  $x_{hi_i}$  为第 i 组最高值组平均,  $\text{var}(x)$  代表 N 个点的方差)。

④P 值: P 值和 Q 值类似, 不同之处在于 Q 值分别统计 3 种强度比率中 1 种, 然后取最大值。而 P 值同时考虑这 3 种强度比率, 取 3 组中中心值比率。同样如果 P 值越大说明分组后的方差占总方差比重越大, 0、1 分组效果越好。P 值是 Q 值的补充参数。  $P = (n_1 \times (g - g_1)^2 + n_2 \times (g - g_2)^2 + \sum_{i=1}^{n_1+n_2} (x_i - g)^2) / 100$  ( $n_1$  和  $n_2$  分别代表聚类到 0、1 的点的个数; g 是 3 个强度比率的中心值,  $g_1$  和  $g_2$  是  $n_1$  和  $n_2$  点中心值)。位点好坏和筛选参数的重要性程度依次为重复性、召回率、P 值和 Q 值。

⑤有效标记数: 指某标记位点满足一定域值(P 值, Q 值, 重复性, 召回率)的标记数。有效标记率指某标记位点的标记参数(P 值, Q 值, 重复性, 召回率)达到一定域值的标记数占总标记的百分数。

### 作者贡献

饶龙兵为本研究的具体完成人。

### 致谢

本研究由国家自然科学基金项目(30700644)资助。作者感谢澳大利亚 Diversity Arrays Technology 实验室 Kilian A、Xia L. 及其他几位同事在遗传标记芯片技术方面给予的支持和帮助, 感谢百山祖自然保护区、元宝山自然保护区、资源冷杉保护区、梵净山自然保护区、桃源洞自然保护区在样品

采集过程中给予的协助。

## 参考文献

- Akbari M., Wenzl P., Caig V., Carling J., Xia L., Yang S., Uaszynski G., Mohler V., Lehmsiek A., Kuchel H., Hayden M.J., Howes N., Sharp P., Vaughan P., Rathmell B., Huttner E., and Kilian A., 2006, Diversity Arrays Technology for high-throughput profiling of the hexaploid wheat genome, *Theor. Appl. Genet.*, 113(8): 1409-1420
- Chen Z.B., ed, 2005, Bio-chip Technology, Chemical Industry Press, Beijing, Chinese, pp.1-23 (陈忠斌, 2005, 生物芯片技术, 化学工业出版社, 北京, 中国, pp.1-23)
- Jaccoud D., Peng K., Feinstein D., and Kilian A., 2001, Diversity Arrays: a solid state technology for sequence information independent genotyping, *Nucleic Acids Res.*, 29(4): E25
- Jeung J.U., Kim B.R., Cho Y.C., Han S.S., Moon H.P., Lee Y.T., and Jena K.K., 2007, A novel gene, Pi40(t), linked to the DNA markers derived from NBS-LRR motifs confers broad spectrum of blast resistance in rice, *Theor. Appl. Genet.*, 115(8): 1163-1177
- Liang X.B., Zhang H., Wang H.Y., Screening for genes related to novel gene AngRem 104 by gene chip, *Zhonghua Shenzang Zazhi (China J. Nephrol.)*, 18(3): 159-165 (梁秀彬, 张宏, 王海燕, 2002, 采用基因表达谱芯片技术筛选新基因AngRem 104的功能相关基因, 中华肾脏病杂志, 18(3): 159-165)
- Patrick J.C., Susan K.B., and Norman F.W., 1997, Random amplified polymorphic DNA -based genetic linkage maps of three cultivars, *Amer. Hort. Sci.*, 122(3): 350-359
- Qian W., Ge S., and Hong D.Y., 2001, Genetic variation within and among populations of a wild rice *Oryza granulate* from China detected by RAPD and ISSR marks, *Theor. Appl. Genet.*, 102: 440-449
- Rao L.B., 2009, Study on conservation genetics of five endangered fir by microarray technology, Dissertation for doctorate, China Academy of Forestry, supervisor: Chen Y.T., pp.1-96 (饶龙兵, 2009, 基于生物芯片标记技术的5种濒危冷杉保护遗传学研究, 中国林科院博士学位论文, 导师: 陈益泰, pp.1-96)
- Sasaki T., and Burr B., 2000, International rice genome sequencing project: The effort to completely sequence the rice genome, *Curr. Opin. Plant Biol.*, 3(2): 138-141
- Wang Y.S., Huang H.W., and Wang Y., 2007, Recent progress in plant molecular population genetics, *Yichuan (Hereditas)*, 29(10): 1191-1198 (王云生, 黄宏文, 王瑛, 2007, 植物分子群体遗传学研究动态, 遗传, 29(10): 1191-1198)
- Wenzl P., Carling J., Kudrna D., Jaccoud D., Huttner E., Kleinohs A., and Kilian A., 2004, Diversity Arrays Technology (DArT) for whole-genome profiling of barley, *Proc. Natl. Acad. Sci.*, 101(26): 9915-9920
- Wittenberg A.H., van der Lee T., Cayla C., Kilian A., Visser R.G., Schouten H.J., 2005, Validation of high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*, *Mol. Genet. Genomics*, 274(1): 30-39
- Xia L., Peng K., Yang S.Y., de Vicente MC, Fregene M., and Kilian A., 2005, DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives, *Theor. Appl. Genet.*, 110(6): 1092-1098
- Yang S.Y., Pang W., Ash G., Harper J., Carling J., Wenzl P., Huttner E., Zong X.X., and Kilian A., 2006, Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology, *Theor. Appl. Genet.*, 113(4): 585-595
- Zhou Y.Q., Yang Q.X., Zhang G.N., (eds), 2008, Biologic genetic marker and application, Chemical Industry Press, Beijing China, pp.1-13 (周延清, 杨清香, 张改娜, 编著, 2008, 生物遗传标记与应用, 化学工业出版社, 中国, 北京, pp.1-13)



**5<sup>th</sup>Publisher**是一个致力于科学与文化传播的中文出版平台

在5<sup>th</sup>Publisher上发表论文, 任何人都可以免费在线阅读您的论文

- ※同行评审, 论文接受严格的高质量评审
- ※在线发表, 论文一经接受, 即刻在线发表
- ※开放取阅, 任何人都可免费取阅无限使用
- ※快捷搜索, 涵盖谷歌学术搜索与知名数据库
- ※论文版权, 作者拥有版权读者自动授权使用

在线投稿: <http://5th.sophiapublisher.com>