

# 利用近红外光谱对杂种松及其亲本的鉴别分析

栾启福, 李彦杰, 姜景民

(中国林业科学研究院亚热带林业研究所, 富阳 311400)

**摘要:** 为了在林木育种和良种推广过程中对杂种松的真实性进行简单、快速鉴别, 本研究以我国南方培育推广的 2 个杂种松及其 3 个亲本树种为研究材料, 通过近红外光谱仪扫描各树种的针叶来采集数据, 最后采用 PLS-DA (偏最小回归-判别分析) 方法建模对杂种松及其亲本进行鉴别。结果表明, PLS-DA 方法对杂种松具有较高的判别能力, 可以较好地将其亲本树种区别开来。因此, 近红外光谱结合 PLS-DA 判别分析方法可以用来分析鉴别松树中的杂种和纯种, 相对于同工酶和分子标记等种质鉴别方法, 近红外光谱技术具有简便、快捷等优点, 值得在林木种质判别分析中进一步研究利用。

**关键词:** 杂种松; 近红外光谱; 偏最小回归-判别分析

## Discriminant Analysis of Pine Hybrids and Their Parents Using Near-Infrared Spectroscopy

LUAN Qi-fu, LI Yan-jie, JIANG Jing-min

(Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang 311400)

**Abstract:** To discriminate among pine hybrids and their parents with simple rapid operation in forest tree breeding and improved pine variety promotion, the partial least square-discriminant analysis (PLS-DA) regression model was established using the data of near-infrared spectroscopy of the dried foliage from two pine hybrids and their three parents in southern China. The results showed that PLS-DA had a high accuracy rate, allowing a secure identification of two hybrids and their three parent species. It was clear that near-infrared technology could be used to identify pine hybrids and purebreds, and it was a more simple, inexpensive, and easy operation than standard isozyme and molecular techniques. The near-infrared technology had vast potential for applications in the discriminant analysis of forest tree species.

**Key words:** pine hybrids; near-infrared spectroscopy; PLS-DA

20 世纪中期以来, 随着林木野生资源的减少和对木材需求的增加, 人们开始培育松树种间杂种来增加单位面积木材产量或改良其抗逆性, 获得较大成功<sup>[1-2]</sup>。然而亲缘关系较近的几种杂种松及其亲本树种在表型形态特征上很难区别, 同工酶<sup>[3-4]</sup>、分子标记<sup>[5-6]</sup>等生化方法可以较好地鉴别种间甚至种内资源, 但对操作者的实验水平要求较高, 且检测费用昂贵, 耗时耗力。因此, 杂种松的鉴别分析迫切需要寻

找一种快速简洁且成本低廉的方法作为当前普遍利用的生化或分子技术判别分析方法的补充和发展。

近红外光谱技术 (NIR, near-infrared spectroscopy) 是 20 世纪 30 年代以来发展最快、最引人注目的光谱分析技术, 操作简单且鉴别率很高, 20 世纪后期即被迅速用于农作物品种和品质的定量或定性鉴别<sup>[7-14]</sup>。近年来近红外光谱技术逐渐渗透到林木材性等品质性状分析及物种鉴别领域<sup>[15-19]</sup>, 然而林木

收稿日期: 2013-03-21 修回日期: 2013-05-08 网络出版日期: 2013-10-22

URL: <http://www.cnki.net/kcms/detail/11.4996.S.20131022.1537.006.html>

基金项目: 浙江省农业新品种选育重大科技专项 (2012C12908-13); 浙江省自然科学基金 (LY12C16009); 国家科技部支撑计划项目 (2012BAD01B02)

第一作者研究方向为林木遗传育种。E-mail: luanqifu915@126.com

通信作者: 姜景民, 研究方向为林木种质资源。E-mail: exotic-pine@hotmail.com

物种鉴别技术尚不成熟,特别是对杂种松的鉴别还停留在模拟阶段<sup>[17]</sup>,缺少实例研究,鉴别方法也比较单一。PLS-DA(偏最小回归-判别分析)是基于PLS回归的一种判别分析方法,因为在构造因素时考虑到了辅助矩阵以代码形式提供的类成员信息,所以表现出高效的鉴别能力,也使假阳性鉴别的概率降低<sup>[20-21]</sup>。本研究以我国南方培育推广的2个杂种松及其3个亲本树种为研究材料,通过近红外光谱仪扫描各树种的针叶来采集数据,最后采用PLS-DA方法进行建模,期望能够使用近红外光谱快捷鉴别分析杂种松及其亲本,为杂种松的育种及良种推广提供技术支持,为相关林木种质鉴别提供借鉴。

## 1 材料与方法

### 1.1 试验材料

试验材料取自杂种松试验林及其亲本树种无性系收集区(表1)。其中亲本树种包括母本火炬松(*Pinus taeda* L.,缩写为PTA)、湿地松(*P. elliottii* Engelman var. *elliottii*,缩写为PEE)和父本洪都拉斯加勒比松(*P. caribaea* var. *hondurensis* Barrett et Golfari,缩写为PCH)。杂种松即PTA × PCH(简称为PTC)和PEE × PCH(简称为PEC)。

表1 试验材料

Table 1 Experimental materials

简称 Abbreviation	编号 Code	松树名称 及拉丁名 Pine species	营建时间 (年) Established years	取样数量 No. of sampling
PEE	1	湿地松 <i>Pinus elliottii</i> Engelman var. <i>elliottii</i>	1979	90
PTA	2	火炬松 <i>P. taeda</i> L.	1994	90
PCH	3	洪都拉斯加勒比松 <i>P. caribaea</i> var. <i>hondurensis</i> Barrett et Golfari	2005	90
PEC	4	湿加松(PEE × PCH)	2008	90
PTC	5	火加松(PTA × PCH)	2008	90

PTA和PEE无性系保存在浙江省杭州市余杭区国营长乐林场(30.33°N,119.86°E),PCH无性系保存在广东省湛江市林木良种场(21.38°N,110.26°E)。2003-2007年选取16个PTA和16个PEE优良无性系作为杂交育种的母本,9个PCH优良无性系为父本进行杂交<sup>[22]</sup>,每年杂交产生的种子

除当年的试验用种以外,余下部分进行低温干燥保存。2007年春季选取部分杂交种在长乐林场苗圃进行育苗,2008年春季在长乐林场西山林区采用随机区组设计营建杂种松试验林,单行6株小区,6次重复,株行距为2 m × 3 m。

### 1.2 样品采集与处理

2012年10月从杂种松试验林中随机选取6个湿加松(PEC)和6个火加松(PTC)杂种松组合,每个组合随机抽取15个单株,共计90个湿加松和90个火加松样本;以上12个杂交组合的亲本无性系包括6个湿地松(PEE)、6个火炬松(PTA)和6个洪都拉斯加勒比松(PCH)无性系,每个无性系随机选取15个分株,即每类亲本树种抽取90个单株样本(表1)。

每个样本针叶的采集都选取一年生健康的针叶,放入中号牛皮纸袋中,记录采集时间与样本编号,放入烘箱恒温70℃烘干至恒重,然后用旋风磨片机将针叶磨成粉末,并用26目标准筛过筛,保证粉末颗粒的均匀一致性,然后把粉末样品重新放入烘箱恒温70℃烘24h后进行近红外扫描。180个杂种松样本中有20个由于在烘干、磨粉过程中损失较大,导致数据采集时样本太少,无法获得NIR数据,所以最后用于分析的样本数为160个。损失的样本中PEC为12个,PTC为8个。

### 1.3 近红外光谱采集

烘干后的样品放入可旋转的小玻璃杯中,通过近红外扫描仪(Model N-500, NIRFlex, BUCHI, Switzerland)在近红外波段(4000~10000/cm)进行扫描,每个样品扫描8次光谱平均为1个光谱值,光谱范围为每4/cm记录1次数据,共1500个数据点,室内温度始终维持在24℃,样品从采集到扫描数据完成耗时7d。

### 1.4 PLS-DA 分析方法

PLS-DA是一种基于PLS回归分析对不同级别之间的模型差异进行分类的一种分类方法。首先需要建立分类变量集将不同类别进行定义,然后建立分类变量与光谱数据间的PLS模型,模型建立以后,将测试集中的样品代入模型进行计算未知样品的分类变量值<sup>[20]</sup>,并根据分类变量值对未知样品进行分类,验证方法采用交叉验证法,光谱范围选择4000~10000/cm。分析时随机取2/3的样本用于建立模型,其余1/3作为测试样品。

将试验材料分别编号(表1),根据PLS-DA分析方法,对5种待鉴别的树种材料设定变幅相同(即5种材料的预测变量变幅均为1)的预测变量

$Y_p$ ,其中各类别材料  $Y_p$  分类如下:(1)当  $0.5 < Y_p < 1.5$ ,且偏差  $< 0.5$ ,判定样品属于分类 1(即湿地松 PEE,下同);(2)当  $1.5 < Y_p < 2.5$ ,且偏差  $< 0.5$ ,判定样品属于分类 2;(3)当  $2.5 < Y_p < 3.5$ ,且偏差  $< 0.5$ ,判定样品属于分类 3;(4)当  $3.5 < Y_p < 4.5$ ,且偏差  $< 0.5$ ,判定样品属于分类 4;(5)当  $4.5 < Y_p < 5.5$ ,且偏差  $< 0.5$ ,判定样品属于分类 5;(6)当  $Y_p$  的偏差  $> 0.5$  时,判定该模型不稳定。

### 1.5 光谱预处理

光谱实际测量过程中不仅包括了松针粉末样品相关的各种信息,还包括仪器噪声、外界干扰等,需要对原始光谱进行适当的预处理,将松针粉末信号和噪声分离,最大程度地提取松针粉末中成分和结构的有效信息。本研究选取了 5 种不同的光谱预处理方法(表 2),包括平滑处理(smoothing)、多元散射校正(MSC, multiplicative signal correction)、标准归一化(SNV, standard normal variate)、一阶导数(1<sup>st</sup> STR, the first derivative transformation)和二阶导数(2<sup>nd</sup> STR, the second derivative transformation)与平滑处理的组合方法。

### 1.6 统计分析

数据处理与预测模型的建立均采用 Unscrambler<sup>®</sup>软件(CAMO Software, Inc., NJ, USA)进行。

## 2 结果与分析

### 2.1 不同光谱预处理对模型准确性的影响

从图 1 可以看出,不同松树品种在光谱响应波峰间存在较小的差异,波峰总体的趋势大致相同,但原始光谱存在很明显的基线漂移现象,这对模型预测准确性有很大的影响,因此,为了最大程度放大样品光谱的有效信息,去除基线偏移以及其他噪声的干扰,需要对光谱预处理,本研究比较了不同的光谱预处理方法对预测模型准确性的影响(表 2),可以看出经过平滑加二阶导数处理的光谱所需主成分数少,且  $R^2$  最大,为 0.99,  $RMSE$  值仅为 0.18。图 2 为经过二阶导数处理后的 5 种树种的光谱,由图 2 可知平滑加二阶导数可以有效去处基线漂移,更好地反映光谱的有效信息,从而提高预测准确度。因此光谱预处理方法选定为平滑加二阶导数处理。

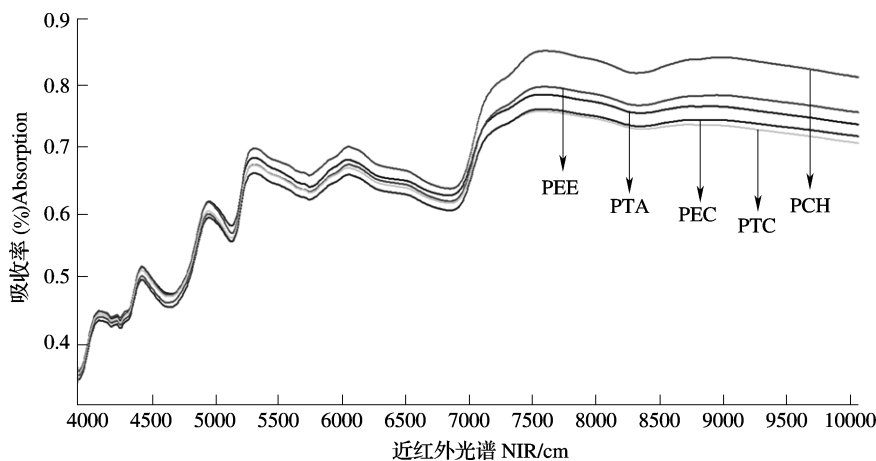


图 1 PEE、PTA、PCH、PEC 和 PTC 针叶烘干打磨后的近红外光谱

Fig. 1 NIR spectra for dried conifer foliage from PEE, PTA, PCH, PEC, and PTC

表 2 PEE、PTA、PCH、PEC 和 PTC 针叶的近红外光谱 PLS-DA 校准模型统计

Table 2 Statistics for the PLS-DA regression calibrations of NIR spectra for dried conifer foliage from PEE, PTA, PCH, PEC, and PTC

光谱 Spectrum	因子数 Factors	均方根误差 $RMSE$	$R^2$	标准误 $SE$
原始光谱 Original spectrum	12	0.68	0.75	0.68
多元散射校正 MSC	12	0.60	0.80	0.60
标准归一化 SNV	12	0.66	0.76	0.66
平滑处理 Smoothing	12	0.68	0.74	0.68
平滑处理 + 一阶导数 Smoothing + 1 <sup>st</sup> STR	12	0.22	0.94	0.22
平滑处理 + 二阶导数 Smoothing + 2 <sup>nd</sup> STR	8	0.18	0.97	0.18

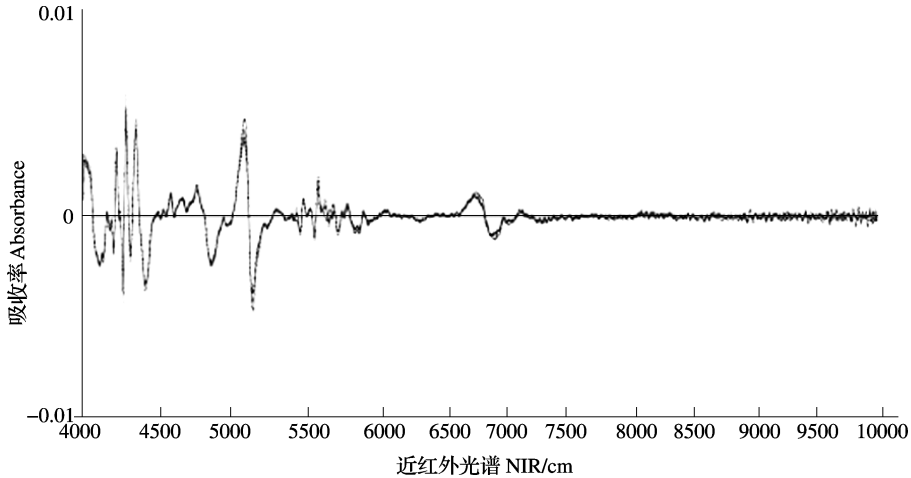


图2 PEE、PTA、PCH、PEC 和 PTC 松针烘干打磨后经平滑 + 二阶导数处理的平均 NIR 光谱

Fig. 2 Smoothing + 2<sup>nd</sup> STR NIR spectra for dried conifer foliage from PEE, PTA, PCH, PEC, and PTC

## 2.2 PLS-DA 判别模型的建立与验证

基于 PLS 方法建立了样本分类变量与 NIR 光谱特征间的回归模型,然后利用 PLS 回归方法对校正集样本的 NIR 光谱与分类变量进行回归分析,并建立 NIR 光谱特征与分类变量间的 PLS 模型,验证方法采用交叉验证法,光谱预处理方式选择平滑 + 二阶导数处理,光谱范围选择 4000 ~ 10000/cm。

图 3 是 PLS-DA 中 5 个分类变量的前 3 个主因子得分的三维图,由图 3 可以看出,每种松树都有各自的分布区域,2 个杂种松的分布区域重叠很大,可能因为其具有共同的亲本加勒比松; 3 种松树亲本的分布区域较容易区分,显示其进化为独立物种后

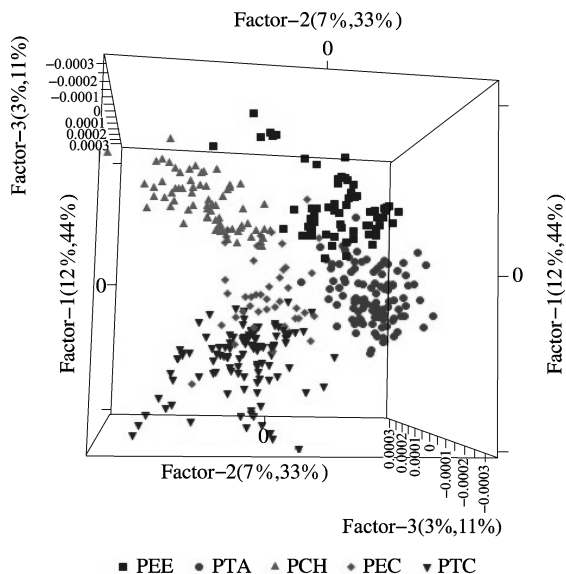


图 3 PEE、PTA、PCH、PEC 和 PTC 近红外光谱的 PLS-DA 因子得分的三维图

Fig. 3 Results from the PLS-DA factors scores of all the individual NIR spectra of PEE, PTA, PCH, PEC and PTC

在 NIR 光谱下的差异,尤其是加勒比松(正三角图标)与另外 2 个松树(方块和圆形)之间在三维图中几乎没有重合区域,表明之间的物种差异性较大,这与 D. S. Gernandt 等<sup>[23]</sup>利用分子标记研究的松树分类结果相同。

基于 PLS-DA 方法的 NIR 分析对 5 种松树的分析结果不但能够较好地判别各类树种,从一定程度上还阐述了各树种间的亲缘关系,显示出 NIR 对杂种松及其亲本具有较强的识别能力,并且具有与分子标记相类似的物种亲缘关系辨别潜力,值得在相关物种鉴别方面进一步研究。

## 2.3 PLS-DA 判别模型的检验

为了检验 PLS-DA 模型的准确性,将未参加模型建立的测试集,代入模型中进行准确性检验,结果见图 4 和图 5。图 4 中纵坐标 1~5 分别代表分类变量 PEE、PTA、PCH、PEC 以及 PTC,图中 5 个物种都准确地被判别出来。统计数据显示,在偏差值 < 0.5 的情况下,判别准确率达 100%,因此 PLS-DA 结合近红外光谱建立的判别模型非常有效,可以用来判别不同杂种松及其亲本。图 5 以聚集点的形式给出了判别分析结果,该结果与图 3 结果较类似,但是更清晰地将 5 类松树区分开来,其相对位置显示的各树种间的亲缘关系更清晰。

## 3 结论与讨论

遗传学家很久以来便认识到物种间的杂交能够在树种的遗传改良和驯化上发挥积极的作用。杂交能够促进基因重组从而获得更好的经济性状。杂种的许多性状都表现为双亲中间性状或偏向一个亲本,或出现超过亲本的性状,而林木种间或种内种源

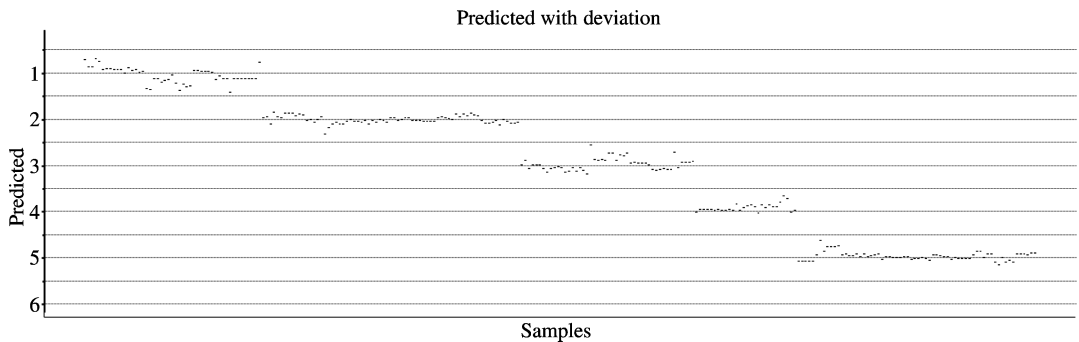


图4 PEE、PTA、PCH、PEC 和 PTC 的 PLS-DA 判别结果

Fig. 4 Discriminant results for PEE, PTA, PCH, PEC, and PTC samples by PLS-DA model

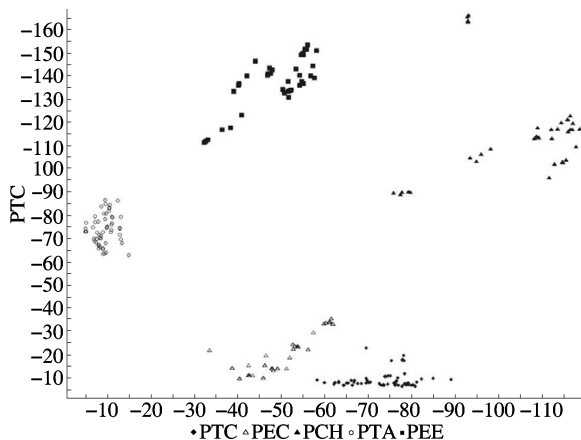


图5 PEE、PTA、PCH、PEC 和 PTC 近红外验证样品判别结果

Fig. 5 Discrimination plot of PEE, PTA, PCH, PEC, and PTC validation data sets

间的杂种还会综合双亲的优良性状<sup>[1]</sup>。因此,广泛开展林木的杂交育种研究是林木遗传改良的重要途径之一。火炬松、湿地松以及加勒比松作为我国南方重要的外来引进针叶树种,规模的杂交育种已经进行了近 20 年<sup>[22]</sup>,在杂交育种和育成杂种推广造林过程中经常需要对杂种松的真实性进行鉴别。由于松树属于长期异交授粉林木,单株基因型高度杂合,种间杂种表现型往往和亲本类似,极难从形态学上鉴别,而分子标记或生化标记鉴别过程较复杂,成本较高<sup>[3-6]</sup>。近红外光谱技术是一项新的无损分析技术,在分析化学领域得到了迅猛发展,被誉为分析“巨人”,在材料科学领域也不断得到应用。由于近红外光谱是 4000 ~ 10000/cm 范围的电磁波,对植物材料十分敏感<sup>[20]</sup>。因此,利用近红外光谱技术对杂种松及其亲本进行识别研究是一条鉴别林木种质的快捷途径。本研究利用近红外技术进行判别分析取得了与分子标记或生化标记相同的判别效果,本试验没有使用生化试剂,过程更简单,是一种更理想的物种判别分析方法。

然而,正是因为近红外光谱对植物材料十分敏感,在利用该技术对林木种质进行鉴别时特别要考虑取样材料以及光谱数据的处理。对于林木来说,叶片无疑是最容易取样、携带树种信息最丰富的器官<sup>[17-18,23]</sup>。树种鉴别时对叶片进行光谱扫描前的处理,目前采用烘烤、制粉的方法,使扫描对象具有统一湿度和统一颗粒大小的特征,该方法较大程度上减少了测量误差。对于针叶树种来说,针叶是较好的取样器官,干燥、粉碎处理也是扫描前减少测量误差的手段。

同时与样本类型及处理方法相对应的近红外光谱数据处理、分析方法也影响着鉴别的效率和结果,几乎所有的相关文献在分析过程中都采用了不同的统计方法加以选择和对比。在光谱数据预处理时经过对比选择了平滑加二阶导数处理的方法,取得较好的效果。然而求导对山参的鉴别并不可用<sup>[25]</sup>。因此在进行光谱数据预处理时需要综合考虑,对几种常用的方法进行对比,选择最适合的方法以减少误判率。近红外光谱的吸收弱、谱带较宽且重叠严重,需用化学计量学中的多变量数据分析方法进行分析,偏最小二乘法(PLS)是近红外光谱分析中应用很广的一种方法,研究中采用近红外光谱结合 PLS 判别法进行分析<sup>[20]</sup>,大大提高了判别的正确率。但是主成分(PCA)-聚类分析<sup>[25]</sup>、主成分-人工神经网络<sup>[26-27]</sup>等判别分析方法也具有各自的优越性,在实际判别分析中可以通过预分析加以选择。

杂种松及其亲本均具有差异性的近红外光谱,初步研究显示,通过合理的统计分析方法(如 PLS-DA)可以较好地利用这些特征光谱对几种松树进行鉴别,而合适的光谱数据处理方法可以更好地分析各树种的光谱特征。因此,近红外光谱结合 PLS-DA 判别分析方法可以用来分析鉴别松树中的杂种和纯

种,相对于同工酶和分子标记这些种质鉴别方法,近红外光谱技术具有简便、快捷等优点,值得在林木种质判别分析中进一步研究利用。

#### 参考文献

- [1] Dungey H S. Pine hybrids-a review of their use performance and genetic[J]. Forest Ecol Manag, 2001, 148: 243-258
- [2] Nikles D G. Experience with some pinus hybrids in Queensland, Australia[C]//Dungey H S, Dieters M J, Nikles D G. Hybrid breeding and genetics of forest trees. Noosa, Queensland, Australia; Proceedings of QFRI/CRC-SPF Symposium, 2000: 27-43
- [3] Conkle M T. Isozyme variation and linkage in six conifer species [C]//Berkeley C A. Proceedings, symposium on Isozymes of North American Forest Trees and Forest Insect. USA; USDA Forest Service, 1979: 11-17
- [4] Chen J W, Tauer C G, Bai G H, et al. Bidirectional introgression between *Pinus taeda* and *Pinus echinata*: evidence from morphological and molecular data[J]. Can J Forest Res, 2004, 34(12): 2508-2516
- [5] Plomion C, Bahrman N, Durel C E, et al. Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers [J]. Heredity, 1995, 74(8): 661-668
- [6] Furman B J, Grattapaglia D, Dvorak W S, et al. Analysis of genetic relationships of Central American and Mexican pines using RAPD markers that distinguish species [J]. Mol Ecol, 1997, 6(4): 321-331
- [7] 李晓丽, 胡兴越, 何勇, 等. 基于主成分和多类判别分析的可见-红外光谱水蜜桃品种鉴别新方法[J]. 红外与毫米波学报, 2006, 25(6): 417-420
- [8] 章海亮, 孙旭东, 郝勇, 等. 近红外漫反射无损检测梨果糖度及 pH 值的研究[J]. 西北农林科技大学学报: 自然科学版, 2010, 38(4): 128-132
- [9] 何勇, 李晓丽. 近红外光谱杨梅品种鉴别方法的研究[J]. 红外与毫米波学报, 2006, 25(3): 192-194
- [10] 赵杰文, 陈全胜, 张海东, 等. 近红外光谱分析技术在茶叶鉴别中的应用研究[J]. 光谱学与光谱分析, 2006, 26(9): 1601-1604
- [11] 陈苗, 金松恒, 黄坚钦, 等. 近红外光谱技术在植物组分与特性研究中的应用 [J]. 科技通报 2011, 27(1): 38-44
- [12] 吴建国, 石春海. 近红外反射光谱分析技术在植物育种与种质资源研究中的应用 [J]. 植物遗传资源学报, 2003, 4(1): 68-72
- [13] 李楠, 许韵华, 宋雯雯, 等. 近红外光谱技术快速预测大豆氨基酸[J]. 植物遗传资源学报, 2012, 13(6): 1037-1044
- [14] 石平, 白琪林, 陈稳良. 玉米自交系秸秆品质性状鉴定与评价[J]. 植物遗传资源学报, 2012, 13(1): 29-34
- [15] Schimleck L R, Monteiro de Matos J L, Silva Oliveirac J T, et al. Non-destructive estimation of pernambuco (*Caesalpinia echinata*) clear wood properties using near infrared spectroscopy[J]. J Near Infrared Spec, 2011, 19(5): 411-419
- [16] Schimleck L R, Higa A R, Jorge L M. Utilisation of near infrared spectroscopy in *Pinus taeda* progeny tests located in southern Brazil[J]. J Near Infrared Spec, 2010, 18(6): 389-396
- [17] Espinoza J A, Hodge G R, Dvorak W S. The potential use of near infrared spectroscopy to discriminate between different pine species and their hybrids[J]. J Near Infrared Spec, 2012, 20(4): 437-447
- [18] Humphreys J R, Reilly-Wapstra J M, Harbard J L, et al. Discrimination between seedlings of *Eucalyptus globulus*, *E. nitens* and their F<sub>1</sub> hybrid using near-infrared reflectance spectroscopy and foliar oil content[J]. Silvae Genet, 2008, 57(4-5): 262-269
- [19] Atkinson M D, Jervis A P, Sangha R S. Discrimination between *Betula pendula*, *Betula pubescens* and their hybrids using near-infrared reflectance spectroscopy [J]. Can J Forest Res, 1997, 27(11): 1896-1900
- [20] 杨忠, 吕斌, 黄安民, 等. 近红外光谱技术快速识别针叶材和阔叶材的研究 [J]. 光谱学与光谱分析, 2012, 32(7): 1785-1789
- [21] 杨忠, 任海青, 江泽慧, 等. PLS-DA 法判别分析木材生物腐朽的研究 [J]. 光谱学与光谱分析, 2008, 28(4): 793-796
- [22] 栾启福, 姜景民, 张建忠, 等. 国外松种间杂交育种及其 F<sub>1</sub> 代早期生长评价 [J]. 林业科学研究, 2008, 21(3): 314-319
- [23] Gernandt D S, López G G, García S O, et al. Phylogeny and classification of *Pinus* [J]. Taxon, 2005, 54(1): 29-42
- [24] Wessman C A, Aber J D, Peterson D L, et al. Foliar analysis using near-infrared reflectance spectroscopy [J]. Can J Forest Res, 1998, 18(1): 6-11
- [25] 卜海博, 聂黎行, 王丹. 近红外光谱法无损识别林下山参及其生长年限 [J]. 光谱学与光谱分析, 2012, 32(7): 1801-1805
- [26] 何勇, 李晓丽, 邵咏妮. 基于主成分分析和神经网络的近红外光谱苹果品种鉴别方法研究 [J]. 光谱学与光谱分析, 2006, 26(5): 850-853
- [27] 牛晓颖, 邵利敏, 赵志磊, 等. 基于 BP\_ANN 的草莓品种近红外光谱无损鉴别方法研究 [J]. 光谱学与光谱分析, 2012, 32(8): 2095-2099