

近红外光谱桉树杂交种鉴别

卢万鸿, 齐杰*, 罗建中

国家林业和草原局桉树研究开发中心, 广东 湛江 524022

摘要 桉树育种和遗传分析是开展桉树世代改良及其目标性状改良等研究的前提。而常用的遗传基础研究方法专业性要求高,且费时费力。该研究旨在利用近红外光谱(NIRs)分析 NIRs 信息与桉树遗传信息间的关系,并探索 NIRs 信息用于桉树杂交种判别分析的可行性和准确性。以现有的桉树杂交种测试试验及其亲本材料为对象,用手持式近红外仪 Phazir Rx(1624)采集了 7 个桉树杂交种及其 4 个亲本树种叶片的 NIRs 信息。每个树种选择 10 个单株,每个单株选 10 片当年生健康叶片,扫描其正面叶脉中部两侧光谱各 5 次,以均值代表单个叶片的 NIRs 信息。每种基因型总共获得 100 条 NIRs 信息,其中 70 条构成训练集样本,30 条构成验证集样本。原始 NIRs 信息采用 S.G 二阶导数转换预处理,以消除基线及其他因素对光谱信息的影响,增强特征峰信息。经预处理后的 NIRs 信息用于后续分析,首先通过主成分分析(PCA)的因子得分对树种的分类判断 NIRs 信息与测试树种遗传信息间的关系。在此基础上,分别用簇类独立软模式(SIMCA)和偏最小二乘判别分析(PLS-DA)两种判别模式建立桉树杂交种的 NIRs 判别模型。经预处理后的 NIRs 信息的变异系数曲线显示,在波长 2 000 nm 后,各树种的 NIRs 信息存在丰富的特征峰,且特征峰的分布范围存在较大的差异。PCA 结果显示,不同的亲本间、杂交种间及杂交种与亲本间样本的 PC1 和 PC2 得分可以清晰地各树种进行分类,这在很大程度上表明 NIRs 信息可以正确反映桉树不同基因型的遗传信息。NIRs 模型的判别效果显示,少数遗传关系比较接近的杂交组合的 SIMCA 模式相互判别准确率较低,而多数杂交组合间的 SIMCA 判别准确率则在 73%~100%之间;桉树各杂交组合间的单独和综合模型的 PLS-DA 判别准确率均为 100%,且基于 PLS-DA 判别的综合模型能将 7 个杂交组合一一与其他组合正确区分开,判别效果明显优于 SIMCA 模式。结果表明:NIRs 信息可以正确反映桉树不同基因型的遗传信息,NIRs 判别模型可以比较准确地将各树种进行区分,因此,NIRs 信息可用于桉树杂交种和纯种的田间定性判别,从而辅助桉树育种材料遗传基础的研究。

关键词 近红外光谱;簇类独立软模式(SIMCA);偏最小二乘判别分析(PLS-DA);桉树杂交种

中图分类号: S722.34 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)05-1451-07

引言

长期有效的桉树(桉属 *Eucalyptus* 和伞房属 *Corymbia*)遗传改良是解决桉树品种单一所导致的生态问题的最有效途径。桉树育种群体的世代改良和杂交亲本选配首先需要解决的则是育种材料遗传亲缘关系的鉴别和分类。常见的物种亲缘关系鉴定手段包括形态学标记^[1-2]、细胞学标记^[3]、生物化学标记、免疫学标记^[4]和分子标记^[5]等方法,这些分析方法通常都耗时耗资,且需具备很强的专业知识,很难快速便捷地开展。随着光谱学和化学计量学的不断发展,借助近红

外光谱(NIRs)信息进行植物分类、产地鉴定及其物质成分快速预测的研究越来越多。Abasolo等^[6]用 NIRs 鉴别伞房属(*Corymbia*)桉树杂交种,模型判别准确率为 72%~100%。Espinoza等^[7]采集了 16 个纯种及人为混合松树(*Pinus*)松针的 NIRs 信息,建立的 NIRs 判别模型经检验,其对纯种和混合树种的判别准确率均达 90%以上。Sandak等^[8]对欧洲 4 个不同地理来源的挪威云杉(*Picea abies* L. Karst.)原木材与其 NIRs 信息间的关系做了研究,结果证明不同来源的木材化学成分存在差异,且 NIRs 信息能够准确地检测出这种差异。Hayes等^[9]发现伞房属桉树的杂交子代与其亲本间的萜烯含量差异很小,常规测量检测不到这种微小差异,但 NIRs

收稿日期:2018-03-20,修订日期:2018-07-25

基金项目:中央级公益性科研院所基本科研业务费专项资金项目(CAFYBB2017MA022),国家自然科学基金青年项目(31700599)资助

作者简介:卢万鸿,1982年生,国家林业和草原局桉树研究开发中心副研究员 e-mail:luwanhong@outlook.com

* 通讯联系人 e-mail:1569545149@qq.com

信息对物质化学成分的极微小差异反映更敏感。

不同来源的物种在特定条件下内在的遗传物质从根本上决定了其组织成分的差异,这是 NIRs 用于物种成分预测和分类的主要依据^[10]。本研究计划以课题组现有的遗传材料为对象,获得其 NIRs 信息,用 SIMCA 和 PLS-DA 两种判别分析方法建立桉树不同杂交种间的判别模型,探索 NIRs 技术用于桉树杂交种鉴别的可行性及方法。

1 实验部分

1.1 材料

杂交种叶样取自课题组 2013 年和 2014 年控制授粉的杂交子代测定林,共计 21 个杂交种家系(组合),按亲本树种将其分为 7 类(表 1)。四个亲本树种包括粗皮桉(*E. pellita*) (P2604, P4963, P5261, P6614, P6713, P6730, P9060, P9070)、韦塔桉(*E. wetarensis*) (W04, W2209, W3226, W0644)、尾叶桉(*E. urophylla*) (EC18)和细叶桉(*E. tereticornis*) (T0105),分别来自各树种的育种群体。根据 Pryor 等的分类系统^[11],粗皮桉、尾叶桉同属横脉组,其亲缘关系较近,细叶桉为缘脉组,与前两者关系较远。

表 1 测试桉树杂交组合及其亲本信息
Table 1 The details of the hybrids and their parents of eucalyptus

组合	分类	杂交种	母本	父本
1	粗韦桉 P×W	K3	P2604	W04
	粗韦桉 P×W	K2	P2604	W3226
	粗韦桉 P×W	K34	P2604	W2209
	粗韦桉 P×W	K4	P2604	W0644
2	粗尾桉 P×U	K50	P9060	EC18
	韦粗桉 W×P	H01	W04	P5261
3	韦粗桉 W×P	H02	W04	P9060
	韦粗桉 W×P	H07	W04	P9070
4	韦细桉 W×T	H19	W04	T0105
	尾粗桉 U×P	K11	EC18	P5261
5	尾粗桉 U×P	K13	EC18	P6730
	尾粗桉 U×P	K14	EC18	P6713
	尾粗桉 U×P	K15	EC18	P4963
6	细粗桉 T×P	H55	T0105	P9060
	细粗桉 T×P	H58	T0105	P5261
	细韦桉 T×W	H53	T0105	W3236
7	细韦桉 <i>E. tereticornis</i> × <i>E. wetarensis</i>	H63	T0105	W04

Note: P: *E. pellita*; T: *E. tereticornis*; U: *E. urophylla*;

W: *E. wetarensis*, the same below

1.2 仪器

手持式近红外仪 Phazir Rx (1624) (Polychromix, Thermo Scientific, USA)用于 NIRs 数据的采集光学分辨率为 12 nm,波长范围为 1 600~2 400 nm,内置基于 MEMS 技术的可编程微衍射光栅,自带背景校正参考片。

1.3 方法

1.3.1 光谱采集

每个树种选择 10 个单株,每株选 10 片当年生健康叶片,扫描其正面叶脉中部两侧光谱各 5 次,以均值代表单个叶片的 NIRs 信息,以提高光谱信噪比^[12-13]。每个树种最终得到 100 条 NIRs 信息,其中 70 条用作训练集,30 条构成验证集。

1.3.2 NIRs 数据的预处理和分析

NIRs 信息通常会受多种因素,如高频噪声、偏移、光的散射等影响,光谱预处理在一定程度上则可以消除这些影响,增强光谱的特征峰信息。有研究认为导数处理可以降低植物叶片的 NIRs 的基线影响^[9, 13]。通过预分析和建模效果对比等过程,本研究对 NIRs 信息采用 S. G 二阶导数转换预处理。光谱预处理和后续分析建模都在 Unscrambler v9.8 (CAMO, Oslo, Norway)中完成。

1.3.3 NIRs 定性判别模型的建立

采用簇类独立软模式(soft independent modeling of class analogy, SIMCA)和偏最小二乘判别分析(partial least squares-discriminant analysis, PLS-DA)两种有监督模式的方法分别建立桉树杂交种的判别模型。

SIMCA 判别以主成分分析(principal component analysis, PCA)为基础,先是建立每个已知类的 PCA 模型,然后用建立的模型对未知样本进行预测,根据未知样本的主成分得分与 PCA 模型中已知类样本因子得分间的距离,判定其归属。

PLS 是 NIRs 定量分析(性状预测)常用的方法,较其他方法更稳定可靠^[14]。在做 PLS-DA 分析之前,人为对各类赋予二进制响应变量(表 2),再建立 NIRs 与响应变量间的校正模型。采用 PLS1 和 PLS2 两种回归方法。PLS-DA 判别以响应变量的预测值 Y_{pred} 为准:

当 $Y_{pred} > 0.5$ 时,且偏差(Deviation)小于 0.5 时,则认为 $Y_{pred} \approx 1$,判定其属于本类;

当 $Y_{pred} \leq 0.5$ 时,且偏差小于 0.5 时,则认为 $Y_{pred} \approx 0$,判定其不属于本类。

表 2 各组合 PLS-DA 分析时的响应变量赋值

Table 2 The response values assigned for seven cross combinations in PLS-DA

组合	赋值
韦粗桉 W×P	1000000
韦细桉 W×T	0100000
细韦桉 T×W	0010000
细粗桉 T×P	0001000
尾粗桉 U×P	0000100
粗韦桉 P×W	0000010
粗尾桉 P×U	0000001

2 结果与讨论

2.1 桉树杂交种叶片的 NIRs 信息变异特征

NIRs 主要是物质有机分子的倍频与合频吸收光谱,能

够得到物质的分子结构、组成和状态信息，还能反映密度、粒度、高分子物的聚合度及纤维形态等物质的物理状态信息^[12]。图 1 是七个桉树杂交组合原始 NIRs 的平均值曲线，通过 NIRs 原始光谱的直观变化很难发现其特征峰，七种组合的 NIRs 信息在全波段变化趋势基本一致，且存在明显的重叠。在波长 1 860 nm 之前和波长 1 910 nm 之后，七个杂交组合的 NIRs 反射率在在一定程度上存在差异，但不足以据此进行树种判别。

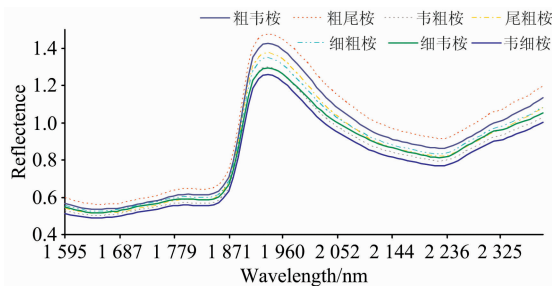


图 1 桉树杂交种的原始 NIRs 反射率光谱

Fig. 1 The raw NIRs reflectance of spectra eucalypt hybrids

由于从桉树杂交种叶片的原始 NIRs 很难看出各组合间的差异信息，为了寻找明确的特征峰，确定建立桉树杂交种判别模型的特定 NIRs 波段，对原始 NIRs 进行了二阶导数预处理，以消除外因对 NIRs 信息的影响，并突出光谱的特征峰。图 2 为经预处理后的桉树杂交种全波段 NIRs 变异系数曲线，从图中可以看出，在 2 000 nm 后，七种桉树杂交种的 NIRs 信息存在丰富的特征峰，而该波段主要为 NH, CH + CH 和 CH + CC 基团的基频峰。图 2 中粗韦桉的特征峰相对平缓，粗尾桉的特征峰分别出现在 2 000, 2 170 和 2 340 nm 左右，韦粗桉和韦细桉的特征峰集中在 2 150~2 200 nm 之间，细韦桉的特征峰在测试波长区间出现多次，尾粗桉的特征峰出现在 2 170 和 2 280 nm 左右，粗细桉的特征峰则出现在 2 250 nm 之后。

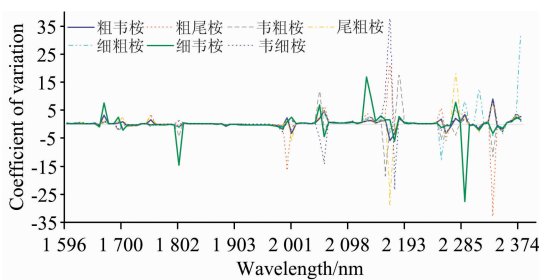


图 2 桉树杂交种的 NIRs 反射率的变异系数

Fig. 2 The coefficient of variation of NIRs reflectance of eucalypt hybrids

2.2 桉树杂交种及其亲本的 PCA 聚类

PCA 过程可以简化多维数据中大量重叠的信息，同时，PCA 的因子得分可以反映受试样本间的距离关系。图 3(a) 和 (b) 是 4 个亲本树种 NIRs 数据主因子 PC1 和 PC2 的因子得分图。从图中可以看出，5 个亲本树种各自的样本清晰地聚在一起。聚类结果与各树种间的真实遗传亲缘关系非常吻

合，表明 NIRs 信息能够直接且真实地反映了各树种内在的遗传信息。

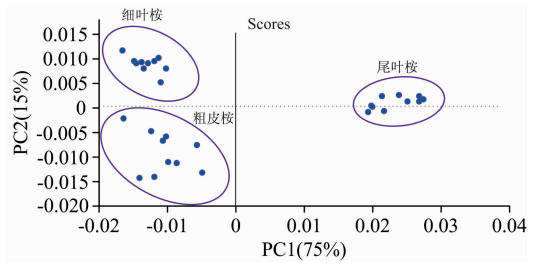


图 3(a) 杂交亲本的 PCA 因子得分图

Fig. 3(a) The scores plot for PC1 and PC2 from PCA with the NIRs of crossing parents

将尾叶桉和韦塔桉单独分析在于学界对两者的分类尚未有定论^[15]，有学者认为韦塔桉只是尾叶桉的一个种源 (Wetar Island)，应属同一个种，即尾叶桉。从本研究来看，如果把 5 个树种放在一起分析，韦塔桉和尾叶桉较难清楚地分开，而通过对两个树种 NIRs 信息的单独 PCA 分析来看，两个树种确实存在较大差异[图 3(b)]。

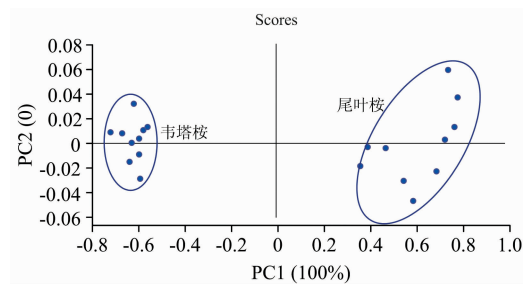


图 3(b) 杂交亲本的 PCA 因子得分图

Fig. 3(b) The scores plot for PC1 and PC2 from PCA with the NIRs of crossing parents

图 4 是两个韦粗桉组合与其亲本树种的 NIRs 数据 PCA 结果中的因子得分图。两个韦粗桉组合母本相同，都是韦塔桉 W04，父本都是粗皮桉，具体家系分别是 P9060 和 P9070。从理论上讲，这两个杂交组合的遗传关系应该很近，但因子得分图还是能将两个组合清楚地分开。同时，图中 3 个亲本树种 W04, P9060 和 P9070 也都各聚一类，而且与对应的杂交子代也清晰地分开，这进一步证明 NIRs 信息完全能够区分不同的基因型，NIRs 可以对遗传差异做出响应。

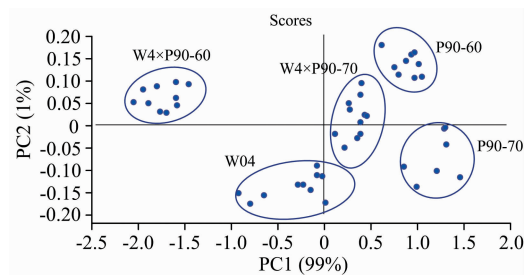


图 4 杂交组合韦粗桉与其亲本的 PCA 因子得分图

Fig. 4 The scores plot for PC1 and PC2 from PCA with the NIRs of W x P combinations and their parents

2.3 桉树杂交种 SIMCA 判别模型的建立与检验

我们用 7 个桉树杂交组合的训练集样本分别建立了各自的 PCA 模型。针对每个组合的 PCA 模型,用其他组合的预

测集样本进行验证,临界概率水平定为 0.05。表 3 统计了每个组合 PCA 模型 SIMCA 模式判别的准确率。

表 3 桉树杂交组合 PCA 模型的 SIMCA 模式判别准确率($\alpha=0.05$) (%)

Table 3 The discriminant accuracy of eucalyptus hybrids' PCA model with SIMCA pattern recognition

预测集	粗韦桉模型	粗尾桉模型	韦粗桉模型	韦细桉模型	尾粗桉模型	细粗桉模型	细韦桉模型
粗韦桉 P×W	96.0	65.3	100.0	100.0	98.0	100.0	100.0
粗尾桉 P×U	33.8	90.0	100.0	100.0	96.3	98.8	100.0
韦粗桉 W×P	84.3	97.1	100.0	100.0	80.0	62.0	88.0
韦细桉 W×T	100.0	100.0	95.0	100.0	75.0	50.0	30.0
尾粗桉 U×P	57.5	93.8	95.0	95.0	95.0	72.9	93.8
细粗桉 T×P	86.7	100	93.3	86.7	90.0	100.0	80.0
细韦桉 T×W	100	100	87.5	86.7	85.0	60.0	92.5

从表 3 可以看出,7 个杂交组合 PCA 模型的 SIMCA 判别准确率普遍都很高,多数杂交组合的模型判别准确率达到了 73%~100%。个别模型在判别某些杂交种时的准确率较低,如粗韦桉模型在判别粗尾桉和尾粗桉的准确率分别为 33.8%和 57.5%,粗尾桉模型判别粗韦桉的准确率为 65.3%。这两个模型判别准确率低的主要原因可能与这几个杂交组合亲本间的遗传亲缘关系有关。前面已经说到,尾叶桉和韦塔桉实际上被认为是同一个树种,这就从理论上决定了“粗韦桉”和“粗尾桉”的遗传亲缘关系也很近。另外,由于粗皮桉和尾叶桉同属横脉组,其亲缘关系较近,所以这几个杂交组合 NIRs 信息的相似性自然就很高。

细粗桉模型在判别韦粗桉、韦细桉和细韦桉时的准确率也比较低(50%~62%),原因可能是这几个杂交组合中都有一个相同亲本(不只是同树种),这就决定了其杂交子代的遗传在很大程度上存在相似性。细韦桉模型判别韦细桉的准确率也很低(30%),则是因为它们是正反交的关系。SIMCA 模式判别是基于样本距离的判别,可以说它更客观地反映了样本间的遗传差异。

2.4 桉树杂交种 PLS-DA 判别模型的建立与检验

PLS-DA 定性判别分析与 PLS 定量分析类似,根据所建模型预测未知样本的响应变量值,判别其类归属。本研究建立了 PLS1 和 PLS2 两种 PLS 模型。PLS1 是针对每个杂交组合分别建立其响应变量和 NIRs 间的 PLS 模型(7 个),而 PLS2 则是以 7 个组合的响应变量同时与 NIRs 信息建立一个综合 PLS 模型(1 个)。表 4 列出了 7 个桉树杂交组合两种 PLS 模型的建模效果和判别准确率。

从表 4 可以看出,7 个组合两种 PLS 模型测量值与预测值间的决定系数 R^2 为 0.79~0.95,测量值与预测值间的平均均方根误差 RMSEC 在 0.08~0.16 间变化,这表明响应变量与各杂交组合 NIRs 信息间的建模效果很好。同时,表中还列出了所建模型在预测未知样本时的表现,其决定系数 R^2 的变化范围为 0.68~0.89,平均均方根误差 RMSEV 在 0.12~0.20 间变化。对比所建模型的建模和预测效果可以发现,两者之间存在一定差距,理论上来说,这主要是由建模集样本与预测集样本的差异所致,这也是 NIRs 模型特异性

的一个体现。但 7 个杂交组合两种 PLS 模型的 PLS-DA 判别效果均显示,每个模型的判别准确率都为 100%,判别效果很理想。

表 4 桉树杂交组合 PLS 模型的 PLS-DA 模式判别准确率($\alpha=0.05$) (%)

Table 4 The discriminant accuracy of eucalyptus hybrids' PLS model with PLS-DA pattern recognition

模型	因子数	训练集		预测集		判别效果/%
		R^2	RMSEC	R^2	RMSEV	
韦粗桉 W×P PLS1	11	0.79	0.16	0.68	0.20	100
韦细桉 W×T PLS1	14	0.93	0.09	0.86	0.13	100
细韦桉 T×W PLS1	15	0.89	0.12	0.74	0.18	100
细粗桉 T×P PLS1	13	0.87	0.13	0.76	0.17	100
尾粗桉 U×P PLS1	14	0.95	0.08	0.89	0.12	100
粗韦桉 P×W PLS1	8	0.84	0.14	0.80	0.16	100
粗尾桉 P×U PLS1	11	0.89	0.12	0.82	0.15	100
7 个组合 PLS2	17	0.80	0.16	0.69	0.20	100

图 5 是 7 个桉树杂交组合 PLS2 模型对 7 个组合验证集样本的 PLS-DA 判别结果,图中更为直观地展示了 PLS-DA 判别模式的判别效果。尽管有些杂交组合的 PLS2 模型预测值偏差稍大,如细粗桉,但整体来看,PLS2 模型的 PLS-DA 判别还是准确清晰地将每一个桉树杂交组合与另外 6 个组合区分开来。必须注意的是借助任何方法的样本判别都离不开对相应领域专业背景知识的了解,纯粹的数学聚类算法一般都不能实现理想的分类效果,或者说特定的分类结果需要对应专业知识的解读。

3 结 论

不同桉树杂交组合的原始 NIRs 反射率、经导数处理后 NIRs 变异系数及其 PCA 的因子得分均说明,桉树叶片的 NIRs 信息真实地反映了不同基因型的遗传信息,NIRs 信息用于桉树杂交种的定性判别具有可靠的理论基础,甚至可以用于林木基因流的评估。

SIMCA 和 PLS-DA 两种模式的判别效果表明, NIRs 成功用于桉树杂交种的田间判别。对于辅助桉树的遗传改良来说, 同时结合两种模式对树种的鉴别分析会更有意义。

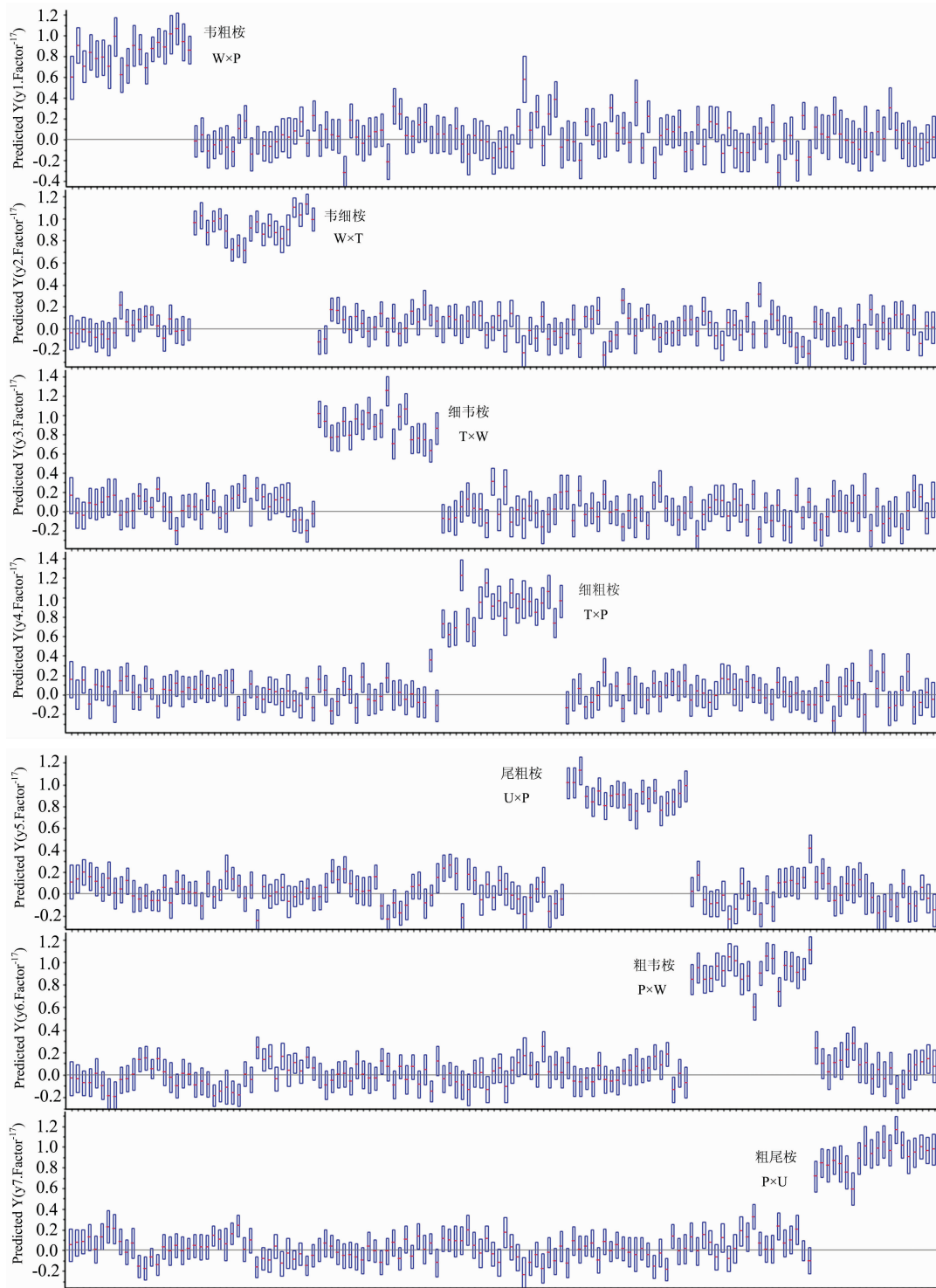


图 5 七个桉树杂交组合 PLS2 模型的 PLA-DA 判别预测值 (横坐标为预测集样本)
Fig. 5 The predicted values for 7 eucalyptus hybrids' by PLS2 model with PLS-DA pattern

References

- [1] Asamoah-Boaheng M, Sam E K. Springerplus, 2016, 5(1): 1.
- [2] SONG Ji-xuan, LI Yun, DENG Ren-ju, et al(宋吉轩, 李云, 邓仁菊, 等). Northern Horticulture(北方园艺), 2017, 18: 58.
- [3] JIA Fang-xin, ZHOU Ming-bing, CHEN Rong, et al(贾芳信, 周明兵, 陈荣, 等). Scientia Silvae Sinicae(林业科学), 2016, 52(9): 57.
- [4] Kawashima A, Kanazawa T, Goto K, et al. Cancer Immunology Immunotherapy, 2017, (1500-11): 1.
- [5] Samah S, Cruz M A S, Valadez-Moctezuma E. Plant Molecular Biology Reporter, 2015, 34(1): 1.
- [6] Abasolo M, Lee D J, Raymond C, et al. Forest Ecology & Management, 2013, 304(304): 121.
- [7] Espinoza J, Hodge G, Dvorak W. Journal of Near Infrared Spectroscopy, 2012, 20(4): 437.
- [8] Sandak A, Sandak J, Negri M. Wood Science & Technology, 2011, 45(1): 35.
- [9] Hayes R A, Nahrung H F, Lee D J. Australian Journal of Botany, 2013, 61: 52.
- [10] Yang G L, Lu W H, Lin Y, et al. Journal of Tropical Forest Science, 2017, 29(1): 121.
- [11] Pryor L D, Johnson L A S. A Classification of the Eucalypts. The Australian National University, Canberra, 1971.
- [12] YAN Yan-lu(严衍禄). Near Infrared Spectroscopy Foundations and Applications(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005.
- [13] LU Wan-hong, YANG Gui-li, LIN Yan, et al(卢万鸿, 杨桂丽, 林彦, 等). Scientia Silvae Sinicae(林业科学), 2017, 53(5): 16.
- [14] Diniz P H G D, Pistonesi M F, Araújo M C U. Analytical Methods, 2015, 7(8): 3379.
- [15] Brooker M I H. Australian Systematic Botany, 2000, 13: 79.

Application of NIRs for Discrimination of Eucalyptus Hybrids

LU Wan-hong, QI Jie*, LUO Jian-zhong

China Eucalypt Research Centre, Zhanjiang 524022, China

Abstract The analysis of genetic basis of breeding materials is the precondition for the improvement programs on populations and interesting traits in eucalypt. However, the traditional ways for that have high professional requirements and are time-consuming and laborious. The aim of present study was to study the relationship between NIRs and genetic information of eucalypt, and discuss the practicability and the accuracy of the discriminant model for the classification of eucalypt hybrids by NIRs data. The NIRs of seven eucalypt hybrids and four parental pure species were scanned with healthy leaves using handheld portable near infrared spectrometer Phazir Rx (1624). 10 individuals were selected for a genotypic species, and 10 healthy current-year leaves were chosen per individual tree. Specially five scans for NIRs from each side of the middle part of the frontal vein of the leaves were taken, and estimated the average of that as the NIRs information of a leaf. In total, 100 NIRs were gained per genotypic species, 70 of which constitute the calibration set, and the validation set consists of the rest 30 NIRs. The transformation of S. G 2nd derivative were performed for the raw NIRs data in present study so as to eliminate the effects of baseline and other factors on the NIRs information, and to strengthen the characteristic peaks of NIRs. The later analysis were conducted after the pre-treatment. Firstly, the relationship between NIRs and genetic information of eucalypt hybrids was studied by the scores plot of principal components (PCs) in principal component analysis (PCA), and on this basis, the NIRs discriminant model was developed. The soft independent modeling of class analogy (SIMCA) and partial least squares-discriminant analysis (PLS-DA) pattern recognition were used to classify eucalypt hybrids with the NIRs model calibrated. The coefficient variation curves of NIRs transformation showed that all phenotypic species studied had rich characteristic peaks, and big differences among them after the wavelength of 2000 nm. The scores plot of PC1 and PC2 in PCA demonstrated clear groups among parental species, hybrids, as well as between hybrids and their parents, suggesting NIRs was a direct response to the genetic information of different genotypes. The discriminant accuracy of SIMCA pattern recognition between some cross combinations, which shared close genetic relation of cross parents, were relatively low using NIRs model. In contrast, the discriminant accuracy of SIMCA pattern recognition among most of eucalypt combinations changed between 73% and 100%. The discriminant accuracy of PLS-DA pattern recognition using single and combined NIRs model of hybrids all were 100%, and the combined model of hybrids based on PLS-DA pattern can discriminate seven hybrids clearly. Studies showed that, the discriminant accuracy of PLS-DA pattern was much higher than that of SIMCA pattern recognition. The current study indicated that NIRs information is the correct response of dif-

ferent genotypic eucalypt species, and the NIRs calibrated model can classify different species of eucalypt accurately, so the NIRs would be used in the qualitative discrimination analysis of eucalypt hybrids and pure species in field, providing an alternative way for the analysis of genetic basis of breeding materials in eucalypt.

Keywords Near infrared spectroscopy (NIRs); Soft independent modeling of class analogy (SIMCA); Partial least squares-discriminant analysis (PLS-DA); Eucalypt hybrids

(Received Mar. 20, 2018; accepted Jul. 25, 2018)

* Corresponding author