

杨春艳,刘 飞,王元忠. 红外光谱结合判别分析对玛咖产地的鉴别[J]. 江苏农业科学,2017,45(5):170-173.  
doi:10.15889/j.issn.1002-1302.2017.05.049

# 红外光谱结合判别分析对玛咖产地的鉴别

杨春艳<sup>1</sup>, 刘 飞<sup>1</sup>, 王元忠<sup>2</sup>

(1. 玉溪师范学院物理系, 云南玉溪 653100; 2. 云南省农业科学院药用植物研究所, 云南昆明 650200)

**摘要:**采用傅里叶变换红外光谱技术测试 6 个产地共 82 株玛咖块根样品的红外光谱;在 Omnic 8.0 光谱软件中用一阶、二阶导数等方法对光谱进行优化处理;用 SPSS 18.5 统计分析软件对经优化的光谱数据在 950 ~ 1 800、1 200 ~ 1 800  $\text{cm}^{-1}$  范围建立 Mahalanobis distance (马氏距离)、Wilks' Lambda、Unexplained variance (未解释方差)、Smallest  $F$  ratio (最小  $F$  值) 及 Rao's  $V$  判别分析模型。结果显示,各样品的红外光谱非常相似,玛咖块根的主要成分是淀粉多糖、蛋白质、纤维素和脂类物质;以 950 ~ 1 800  $\text{cm}^{-1}$  范围的二阶导数光谱数据为判别指标进行 Mahalanobis distance 判别分析效果较好,且模型较稳定,回判正确率为 100%,平均预测正确率为 81.8%,总判别正确率为 92.7%。傅里叶变换红外光谱结合判别分析,可以有效鉴别玛咖产地来源,为玛咖的鉴别分类提供可靠依据。

**关键词:**玛咖;傅里叶变换红外光谱;逐步判别分析;产地鉴别

**中图分类号:** O657.33 **文献标志码:** A **文章编号:** 1002-1302(2017)05-0170-04

玛咖(maca)别称玛卡,属被子植物门,双子叶植物纲,罂粟目,独行菜属的一、两年生药食两用草本植物,原产于海拔 3 500 ~ 4 500 m 的南美安第斯山区,于 2002 年云南省农业科学院高山经济植物研究所在丽江海拔 3 000 m 以上区域引种植玛咖,现已在丽江玉龙雪山一带、昆明市轿子雪山、香格里拉巴雪山等地实现规模种植。玛咖块根的主要成分有蛋白质、脂肪酸、氨基酸、碳水化合物及微量元素等,还含有玛咖酰胺、玛咖烯、生物碱、芥子油苷、甾醇及多酚等具有生物活性的次生代谢物质<sup>[1-2]</sup>,其中玛咖酰胺和玛咖烯具有改善性功能的功效<sup>[3]</sup>,生物碱能调节荷尔蒙,缓解更年期综合征,芥子油苷和异硫氰酸苄酯对治疗胃癌、食道癌、肺癌、白血病等癌细胞均有一定效果。

中药材的有效成分由遗传和生态条件(包括栽培和加工技术)2 个因素共同决定,同一遗传条件的中药材,受产地因素影响,即使加工技术一致,药材的药用成分及药效均可能存在差异<sup>[4]</sup>。对于玛咖真伪的鉴别和品质评估已有学者进行了研究<sup>[5]</sup>,但对不同产地来源玛咖的鉴别研究,迄今尚未见报道。

对中药材真伪和品质优劣的鉴别,传统方法主要通过性状进行经验鉴别,但此法很难对药材产地进行鉴别。现代技术方法主要有高效液相色谱法(high performance liquid chromatography,简称 HPLC)、气相层析法(gas chromatography,简称 GC)和傅氏转换红外光谱分析(fourier transform infrared spectroscopy,简称 FTIR)技术等,但 HPLC 和 GC 须对药材进行分离提取,专业性强、技术难度大,不利于推广应用,FTIR

技术既不能破坏样品的物质组分,又能全面反映物质内部分子结构信息,且兼具样品用量少、技术难度小等优点,已在药材产地鉴别方面成功应用。判别分析是根据已知类别的若干样本的某些指标信息,寻找分类函数,对未知类别的样本进行分类的方法<sup>[6]</sup>。本试验测试了秘鲁和云南 5 个产地的 82 株玛咖植株块根样品的红外光谱,用一阶、二阶导数对光谱进行预处理,采用判别分析中的逐步判别分析法对样品的产地进行鉴别研究,以期对鉴别玛咖的产地提供简便、快捷的方法。

## 1 仪器与方法

### 1.1 仪器设备与测试条件

红外光谱仪为珀金埃尔默股份有限公司生产的 Frontier 型傅里叶变换红外光谱仪,装备 DTGS 检测器,光谱扫描范围为 4 000 ~ 400  $\text{cm}^{-1}$ ,分辨率为 4  $\text{cm}^{-1}$ ,累加扫描次数为 16 次。光谱数据处理使用 Omnic 8.0 光谱专业软件,判别分析使用 SPSS 18.5 统计分析软件。

### 1.2 样品制备与光谱预处理

试验所用 82 株玛咖植株块根样品均由云南省农业科学院药用植物研究所提供,分别取自秘鲁和我国云南的丽江、会泽、东川、昭通、香格里拉 6 个产地,其编号依次为 A1 ~ A8、B1 ~ B10、C1 ~ C9、D1 ~ D8、E1 ~ E17、F1 ~ F30,共涉及黑色、紫色、黄色和白色 4 种色型。样品在 45  $^{\circ}\text{C}$  恒温下整根热烘干至恒质量,用中药粉碎机粉碎待测。测试时取玛咖块根粉末于玛瑙研钵中磨为细粉后按样品与溴化钾 1 : 45 的质量比加入溴化钾搅磨均匀,然后压片测试光谱。每株玛咖块根样品压制、测试 1 张光谱,所有光谱均已扣除背景,在 Omnic 8.0 软件中进行自动平滑、自动基线校正及归一化处理,同时计算各样品红外光谱的一阶导数光谱和二阶导数光谱。

## 2 结果与分析

### 2.1 玛咖植株块根的红外光谱和二阶导数光谱分析

从图 1 和表 1 可以看出,不同产地玛咖块根的红外光谱

收稿日期:2016-01-11

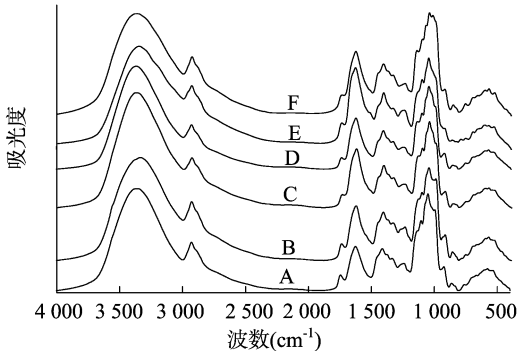
基金项目:国家自然科学基金(编号:31460538);云南省自然科学基金重大项目(编号:2013FC006);玉溪师范学院青年教师资助计划。

作者简介:杨春艳(1979—),女,云南祥云人,硕士,副教授,主要从事光谱学研究。E-mail:ychyky@163.com。

通信作者:王元忠,硕士,副研究员,主要从事药用植物资源研究。E-mail:boletus@126.com。

峰形相似、峰位相近。除香格里拉样品光谱的最强峰出现在多糖吸收区外,其他样品的最强峰均位于 O—H 和 N—H 伸缩振动吸收区。为便于比较,将光谱分为 6 个区,K1 区(3 700~3 000 cm<sup>-1</sup>)中 3 365 cm<sup>-1</sup>附近吸收峰主要来自脂类、蛋白质、淀粉中羟基 O—H 伸缩振动及蛋白质酰胺 I 带中 N—H 反对称伸缩振动;K2 区(3 000~2 800 cm<sup>-1</sup>)中 2 929 cm<sup>-1</sup>附近吸收峰,主要反映亚甲基 C—H 反对称伸缩振动,是挥发油的特征吸收峰<sup>[7]</sup>;K3 区(1 750~1 500 cm<sup>-1</sup>)是酰胺和羰基振动区,其中 1 741 cm<sup>-1</sup>附近吸收峰来自脂类及纤维素、半纤维素中羰基 C=O 伸缩振动吸收<sup>[8]</sup>,1 632 cm<sup>-1</sup>附近中等强度吸收峰,主要来自蛋白质酰胺 I 带中 C=O 伸缩振动、N—H 弯曲振动吸收,以及苷类、糖类物质中 O—H 弯曲振动<sup>[9]</sup>,以及甾体皂苷中羰基的伸缩振动吸收<sup>[10]</sup>;K4 区(1 500~1 200 cm<sup>-1</sup>)主要为蛋白质、脂肪酸和多糖混合振动吸收区,其中 1 413、1 337 cm<sup>-1</sup>附近吸收峰来自亚甲基 C—H 剪式弯曲振动,1 241 cm<sup>-1</sup>附近的吸收峰为脂类化合物中 C—O—C 的伸缩振动和蛋白质酰胺 III 带中 C—N 伸缩振动及 N—H 弯曲振动吸收的叠加峰;K5 区(1 200~950 cm<sup>-1</sup>)主要为多糖和糖苷吸收区,其中 1 104 cm<sup>-1</sup>附近的吸收峰为多糖及纤维素、半纤维素中 C—C 伸缩振动吸收峰<sup>[11-12]</sup>,1 048 cm<sup>-1</sup>附近的强吸收峰为多糖、纤维素、半纤维素中 C—O 伸缩振动吸收峰和糖苷类物质中—OH 伸缩振动吸收的叠加峰<sup>[13]</sup>;K6 区(950~700 cm<sup>-1</sup>)主要为碳水化合物中糖环伸缩振动吸收区,其中 929、861、761 cm<sup>-1</sup>附近的弱吸收峰表明玛咖块根中含有 α 型糖苷键<sup>[14]</sup>。在 700 cm<sup>-1</sup>以下,主要呈现了 577 cm<sup>-1</sup>附近吸收峰。上述光谱特征反映了玛咖块根的主要物质成分为多糖、蛋白质、纤维素和脂类物质。各样品光谱中,多糖类物质的特征吸收峰相对较强,蛋白质类

物质的次之,脂类物质的较弱,说明样品中多糖类物质的相对含量较多,脂类物质的较少。



A—秘鲁; B—丽江; C—会泽; D—东川; E—昭通; F—香格里拉, 下图同。

图1 玛咖主根的红外光谱

从图 1 和表 1 还可看出,不同产地玛咖的红外光谱间的差异主要表现为:除香格里拉产的玛咖样品光谱最强峰和第二强峰依次出现在 1 048、3 365 cm<sup>-1</sup>附近,其他产地光谱的最强峰和第二强峰位置则与之相反;在蛋白质、脂肪酸和多糖混合振动区,只有秘鲁样品在 1 374 cm<sup>-1</sup>处显示了 1 个弱吸收峰;多糖和糖苷吸收区中,在 1 140 cm<sup>-1</sup>附近只有昭通样品未显示吸收峰,在 1 075 cm<sup>-1</sup>处只有香格里拉样品显示了 1 个极强的吸收峰,在 1 000 cm<sup>-1</sup>附近,昭通样品和东川样品均未显示吸收峰,其他样品则显示了强的吸收峰;糖类异构体吸收区中,在 796 cm<sup>-1</sup>处,只有昭通样品显示了吸收峰,在 700 cm<sup>-1</sup>附近,只有秘鲁样品未显示吸收峰。这些光谱差异,为初步鉴别昭通、香格里拉、秘鲁和东川种植样品提供了可能,但无法鉴别会泽和丽江 2 个产地的样品。

表 1 不同产地玛咖块根红外光谱的主要吸收峰

产地	峰位( cm <sup>-1</sup> )																				
			K3		K4				K5					K6				<700 cm <sup>-1</sup> 区			
	K1	K2	吸收带 1	吸收带 2	吸收带 1	吸收带 2	吸收带 3	吸收带 4	吸收带 1	吸收带 2	吸收带 3	吸收带 4	吸收带 5	吸收带 1	吸收带 2	吸收带 3	吸收带 4	吸收带 1	吸收带 2	吸收带 3	吸收带 4
香格里拉	3 365	2 929	1 741	1 632	1 413		1 337	1 241	1 147	1 104	1 075	1 048	1 022	929	861	761	706	667	608	577	
昭通	3 347	2 929	1 741	1 633	1 410		1 337	1 247		1 106		1 053		929	867	796	761	702		608	578
东川	3 369	2 930	1 741	1 628	1 411		1 337	1 253	1 139	1 104		1 050		928	861		759	702			578
会泽	3 371	2 932	1 741	1 628	1 412		1 337	1 243	1 137	1 106		1 050	1 000	927	861		757	700	665	602	579
丽江	3 338	2 931	1 743	1 628	1 413		1 334	1 241	1 131	1 104		1 051	998	924	865		761	702	667	602	579
秘鲁	3 356	2 930	1 745	1 627	1 413	1 374	1 335	1 242	1 135	1 104		1 051	997	925	863		761		667		579

为放大光谱差异,采用 Omnic 8.0 软件计算所有样品红外光谱的一阶导数光谱和二阶导数光谱,各产地第 1 份黄色型玛咖样品的二阶导数光谱图见图 2。在红外二阶导数光谱中,1 461、1 205、1 159、1 108、1 080、1 052、1 020、990、922、864、764、708、612、576、530 cm<sup>-1</sup>附近 15 个吸收峰与淀粉的红外二阶导数光谱在相应位置附近吸收峰的峰形和相对强度相似,说明玛咖主根中的多糖主要为淀粉多糖。

2.2 基于 FTIR 数据的玛咖产地的判别分析

各产地玛咖样品的光谱非常相似,为能客观、全面地区别产地来源,以光谱数据作为判别指标,采用 SPSS 18.5 软件逐步判别分析模块中 Wilks' Lambda、Unexplained variance、Mahalanobis distance、Smallest F ratio、Rao's V 共 5 种挑选判别

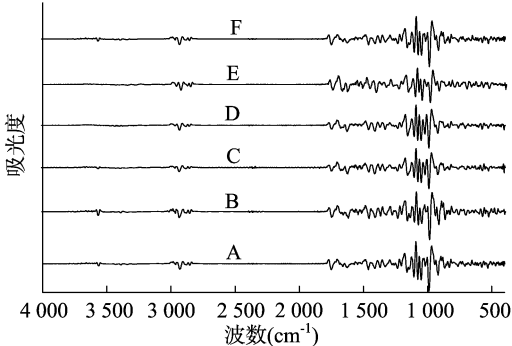


图2 玛咖块根的二阶导数红外光谱

指标的方法,依据 Fisher 线性判别准则建立分类模型对 6 个

产地共 82 份样品的红外光谱、一阶导数光谱和二阶导数光谱进行判别。研究中对 6 个产地、4 种色型的 82 份玛咖样品,按各产地不同色型样品数的 60% 作为训练样本共 49 个,其他 33 个作为测试样本。

2.2.1 不同级别的红外光谱对玛咖产地的判别分析 分别以各样品在 1 200 ~ 1 800  $\text{cm}^{-1}$  范围的红外光谱数据、一阶导数光谱数据和二阶导数光谱数据为判别指标,对 6 个产地共 82 份样品按照逐步判别法,选用 Wilks' Lambda 方法挑选判别变量,依据 Fisher 线性判别准则建立判别模型,对样品产地进行鉴别研究。从表 2 可以看出,基于二阶导数光谱数据建立的判别函数对训练样本的回判正确率和对测试样本的预测正确率及总的判别正确率最高,一阶导数光谱的次之,红外光谱的最低,这与导数光谱具有放大光谱特征,增强光谱的分辨

率,且光谱级别越高,分辨能力越强的特征相一致。因此,二阶导数光谱数据作为判别指标所建立的函数,判别效果较好,说明二阶导数光谱数据更适合于用作判别指标鉴别玛咖的产地。

2.2.2 不同光谱范围的二阶导数光谱数据对玛咖产地的判别分析 对测试样品的二阶导数光谱数据,分别选择 950 ~ 1 800、1 200 ~ 1 800  $\text{cm}^{-1}$  范围的数据作为判别指标,采用逐步判别法中的 Wilks' Lambda 法挑选变量,依据 Fisher 线性判别准则建立判别模型,对样品进行鉴别。由表 3 可知,950 ~ 1 800  $\text{cm}^{-1}$  光谱范围的数据所建立的判别函数对训练样本的回判和测试样本的预测效果均较好,训练样本的回判正确率为 100.0%,测试样本的预测正确率为 90.9%,总的判别正确率为 96.3%。

表 2 不同类型的红外光谱对玛咖产地的判别分析结果

数据类型	训练样本回判			测试样本预测			总体情况	
	数目	误判数	正确率 (%)	数目	误判数	正确率 (%)	误判数	正确率 (%)
红外光谱	49	16	67.3	33	17	48.5	33	59.8
一阶导数光谱	49	6	87.8	33	9	72.7	15	81.7
二阶导数光谱	49	1	98.0	33	7	78.8	8	90.2

表 3 不同光谱范围的二阶导数光谱对玛咖产地的判别分析结果

光谱范围 ( $\text{cm}^{-1}$ )	训练样本回判			测试样本预测			总体情况	
	数目	误判数	正确率 (%)	数目	误判数	正确率 (%)	误判数	正确率 (%)
950 ~ 1 800	49	0	100.0	33	3	90.9	3	96.3
1 200 ~ 1 800	49	1	98.0	33	7	78.8	8	90.2

2.2.3 不同挑选判别指标对玛咖产地鉴别的影响 以 950 ~ 1 800  $\text{cm}^{-1}$  范围的二阶导数光谱数据为判别指标,分别按照 SPSS 18.5 软件逐步判别分析模块中 Wilks' Lambda、Unexplained variance、Mahalanobis distance、Smallest  $F$  ratio、Rao's  $V$  共 5 种判别指标选择方法,依据 Fisher 线性判别准则建立判别模型,对样品产地进行鉴别研究;同时采用十折五次交叉验证法检验挑选判别指标进入判别函数时不同方法所建立判别模型的稳定性<sup>[15]</sup>。由表 4 可知,挑选判别指标进入判别函数的各种方法中,选择 Mahalanobis distance 和 Smallest  $F$  ratio 方法时,样品的平均总正确率、回判正确率和预测正确率相同且最高,依次为 92.7%、100.0% 和 81.8%,选择 Unexplained variance 方法时,平均总正确率、回判正确率和预测正确率均次高,依次为 91.2%、99.6% 和 78.8%。同种方法在十折五次交叉验证的不同训练样本情况下,训练样本和测试样本的回判正确率和预测正确率的差异,反映了用这种方法所建立模型的稳定性。在 5 种训练样本情况下,选择 Mahalanobis distance 和 Smallest  $F$  ratio 方法时,对训练样本的回判正确率差异均为 0,选择 Unexplained variance 方法时,其差异为 2.0%;选择 Smallest  $F$  ratio 方法时,对测试样本的预测正确率差异为 21.2%,选择 Mahalanobis distance、Unexplained variance 方法时,差异依次为 18.2%、15.1%,说明 Mahalanobis distance、Unexplained variance 方法所建立模型的稳定性较好。综合不同挑选判别指标方法的平均总正确率,以及在 5 种训练样本情况下对训练样本的回判正确率和对测试样本的预测正确率,以及最高正确率和最低正确率,基于 950 ~

1 800  $\text{cm}^{-1}$  范围的二阶导数光谱数据的判别分析对玛咖产地的鉴别,选用 Mahalanobis distance 逐步判别法最合适,其次为 Unexplained variance 逐步判别法。

3 结论

用 FTIR 技术测试 6 个产地、4 种色型的 82 株玛咖块根样品的红外光谱,通过光谱软件 Omnic 8.0 计算各样品光谱的一阶导数光谱和二阶导数光谱。以不同级别红外光谱数据为判别指标,按照逐步判别法,依据 Fisher 线性判别准则对玛咖的产地进行鉴别研究,比较 3 种级别光谱、2 个光谱范围、5 种挑选判别指标进入判别函数的方法所建立的判别模型的鉴别效果,同时按十折五次交叉验证法检验 5 种挑选判别指标方法建立的模型的稳定性。结果表明,玛咖植株块根在 950 ~ 1 800  $\text{cm}^{-1}$  范围的红外二阶导数光谱数据结合 Mahalanobis distance 挑选判别指标方法的逐步判别法能较好地鉴别样品的产地,正确率达 92.7%。FTIR 技术结合判别分析方法,可为玛咖产地的鉴别提供一种简便易行的方法。

本试验的判别分析结果都是按最大可能判别结果计算的,若将最大可能结果和次大可能结果相结合,能帮助提高样品判别结果的相对准确性,缩小样品分类所属范围。如在本研究中,以 950 ~ 1 800  $\text{cm}^{-1}$  范围二阶导数光谱数据作为判别指标,按照 Mahalanobis distance 逐步判别法,依据 Fisher 线性判别准则建立判别函数对样品进行鉴别,并按十折五次交叉验证,对测试样本的平均预测正确率达 94.5%,平均总正确率达 97.8%。

表 4 不同挑选判别指标方法对玛咖产地的判别分析结果

判别指标挑选方法	训练样本情况	训练样本回判			测试样本预测		
		数目	误判数	正确率(%)	数目	误判数	正确率(%)
Mahalanobis distance	1	49	0	100.0	33	5	84.8
	2	49	0	100.0	33	10	69.7
	3	49	0	100.0	33	6	81.8
	4	49	0	100.0	33	5	84.8
	5	49	0	100.0	33	4	87.9
	平均			100.0			81.8
Wilks' Lambda	1	49	0	100.0	33	3	90.9
	2	49	0	100.0	33	7	78.8
	3	49	1	98.0	33	7	78.8
	4	49	0	100.0	33	6	81.8
	5	49	8	83.7	33	12	63.6
	平均			96.3			78.8
Rao's V(R=0)	1	49	2	95.9	33	14	57.6
	2	49	1	98.0	33	10	69.7
	3	49	0	100.0	33	5	84.8
	4	49	1	98.0	33	10	69.7
	5	49	0	100.0	33	7	78.8
	平均			98.4			72.1
Unexplained variance	1	49	1	98.0	33	10	69.7
	2	49	0	100.0	33	6	81.8
	3	49	0	100.0	33	9	72.7
	4	49	0	100.0	33	5	84.8
	5	49	0	100.0	33	5	84.8
	平均			99.6			78.8
Smallest F ratio	1	49	0	100.0	33	2	93.9
	2	49	0	100.0	33	9	72.7
	3	49	0	100.0	33	5	84.8
	4	49	0	100.0	33	7	78.8
	5	49	0	100.0	33	7	78.8
	平均			100.0			81.8

注: Mahalanobis distance、Wilks' Lambda、Rao's V、Unexplained variance、Smallest F ratio 的平均总正确率为 92.7%、89.3%、87.8%、91.2%、92.7%。

参考文献:

[1] Ganzera M,Zhao J,Muhammad I. Chemical profiling and standardization of *Lepidium meyenii* (maca) by reversed phase high performance liquid chromatograph[J]. Chemical and Pharmaceutical Bulletin, 2002,50(7):988-991.

[2] Dini I,Tenore G C,Dini A. Glucosinolates from maca (*Lepidium meyenii*)[J]. Biochemical Systematics and Ecology,2002,30(11):1087-1090.

[3] Zheng B L,He K,Kim C H,et al. Effect of a lipidic extract from *Lepidium meyenii* on sexual behavior in mice and rats[J]. Urology,2000,55(4):598-602.

[4] 何春娥,魏建和,陈士林,等. 四个产地丹参种质根中微量元素含量的比较分析[J]. 光谱学与光谱分析,2010,30(3):801-803

[5] Jin W W,Xiong Y,Yu L J. Identification and quality evaluation of *Lepidium meyenii* (maca) based on gas chromatographic analysis of its essential oils from roots[J]. Agricultural Scientific and Technology Paper,2006,7(3):2-9.

[6] 宇传华. SPSS 与统计分析[M]. 2 版. 北京:电子工业出版社,2014.

[7] 杨兴仓,司民真,刘仁明,等. 野三七的红外光谱分析[J]. 光散射学报,2011,23(2):162-167.

[8] Gorgulu S T,Dogan M,Severcan F. The characterization and differentiation of higher plants by fourier transform infrared spectroscopy[J]. Applied Spectroscopy,2007,61(3):300-308.

[9] 孙素琴,周群,陈建波,等. 中药红外光谱分析与鉴别[M]. 北京:化学工业出版社,2010.

[10] 颜茜. 中药重楼及其伪品开口箭的 FTIR 分析与鉴别[J]. 光散射学报,2013,25(1):85-91.

[11] 李志永,刘刚,李伦,等. 蚕豆病害叶的 FTIR 研究[J]. 光谱学与光谱分析,2012,32(5):1217-1220.

[12] 史虞花,司民真. 鼠曲草的红外光谱分析[J]. 光散射学报,2009,21(4):366-370.

[13] 张黎,刘刚,李伦,等. FTIR 结合主成分分析对棕榈科植物鉴别研究[J]. 光散射学报,2013,25(1):98-102.

[14] 肖波,毛文华,梁小红,等. 基于高光谱图像和判别分析的草地早熟禾品种识别研究[J]. 光谱学与光谱分析,2012,32(6):1620-1623.

[15] 张新新,李雨,纪玉佳,等. 主成分-线性判别分析在中药药性识别中的应用[J]. 山东大学学报(医药版),2012,50(1):143-146.