

基于 FTIR 的红花主成分聚类分析

唐 军¹, 王 青², 王 强^{1*}

1. 新疆大学 理化测试中心, 新疆 乌鲁木齐 830046

2. 新疆大学物理科学技术学院, 新疆 乌鲁木齐 830046

摘 要: **目的** 建立红花药材的化学模式识别方法。**方法** 采用傅里叶红外光谱法, 对 5 种不同产地红花样品的红外谱图, 以 $900\sim 1\,550\text{ cm}^{-1}$ 内的吸收峰为指标, 应用主成分和聚类方法进行数据分析。**结果** 红花样品可分为 4 个产区, 同一产区内红花的化学组分相似, 可做为红花药材质量评价的依据。**结论** 基于 FTIR 谱的主成分和聚类分析能够在一定程度表征出红花在不同地理位置和气候条件的多样性分化, 能对红花的合理利用提供依据。

关键词: 红花; 傅里叶红外光谱; 主成分分析; 聚类分析; 化学模式识别

中图分类号: R282.2 文献标志码: A 文章编号: 0253-2670(2011)06-1213-04

Clustering on safflower principal component analysis based on FTIR

TANG Jun¹, WANG Qing², WANG Qiang¹

1. Physics and Chemistry Detection Center, Xinjiang University, Urumqi 830046, China

2. College of Physics Science and Technology, Xinjiang University, Urumqi 830046, China.

Key words: *Carthamus tinctorius* L.; FTIR; principal component analysis; cluster analysis; chemical pattern recognition

傅里叶转换红外光谱 (FTIR) 技术具有分析速度快, 对样品无损害, 能进行快速定量和定性分析等特点^[1], 目前已用于中药材^[2-5]、微生物^[6]等的分类鉴定中, 其在中药材分析和质量监控等方面也受到人们的普遍重视。红花为菊科红花 *Carthamus tinctorius* L. 的干燥花, 别名草红花、刺红花、杜红花, 在我国新疆、四川、山东、云南、河南等地均有栽培。具有活血通经、散瘀止痛之功效。目前主要依据它的种类以及产地评价红花的优劣, 红花鉴定仍以形态学特征和理化分析为主^[7-9], 红花质量的研究多采用单一或几种主要化学成分测定为主。刘月庆等^[10]对羟基红花黄色素 A 进行测定, 赵明波等^[11]以羟基红花黄色素 A 为基准物, 对不同产地的红花药材进行指纹图谱分析, 王若菁等^[12]用色谱法测定红花药材, 对其中色谱峰进行了指认研究, 并从色素、腺苷的量和黄酮类化合物的指纹图谱来对红花药材进行质量评价。但采用少数几种化学成分评价中药材质量的方法并不能客观地体现中药化学成分的复杂性, 而 FTIR 能对中药中的所有化学成分

产生响应吸收峰, 可以较全面地体现出中药中的化学成分变化, 而且检测方法便捷、快速。对于不同产地的同种药材, 红外谱图差异较小, 不能采用直接观测分析谱图的形式进行鉴别, 而采用主成分和聚类分析的模式识别方法能对相似红外谱图加以区分和鉴别^[13], 因为采用的是数据降维分析法, 可以实现红外光谱数据的线性降维投影显示, 从而使人们能直观地从二维或三维图中观察到光谱的聚类情况。本实验利用 FTIR 法以不同产地红花的红外谱图为对象, 对红花的产地进行聚类分析, 并以图形方式表述, 主成分和聚类分析的结果与产地都存在相关性, 结果理想。

1 材料和方法

1.1 仪器与试样

Vertex70 FT-IR 光谱仪 (德国布鲁克公司); 光谱分辨率 4.0 cm^{-1} , 波数在 $600\sim 4\,000\text{ cm}^{-1}$, 信号累加 32 次。新疆塔城红花 (编号 T) 和新疆博乐红花 (编号 B) 均在产地购买, 伊朗红花 (编号 Y)、山东红花 (编号 S)、四川红花 (编号 C) 在乌鲁木齐

收稿日期: 2010-09-12

作者简介: 唐 军 (1971—), 男, 新疆大学高级工程师, 主要从事仪器分析工作。

*通讯作者 王 强 Tel: 18909924168 E-mail: tangjun_wq@sina.com

药材公司购买,每个产地购买 4 个样品。

所用的红花药材经乌鲁木齐市友谊医院主管中药师沈莉红鉴定为菊科植物红花 *Carthamus tinctorius* L. 的干燥花。

1.2 样品的制备

不同产地红花样品用玛瑙研钵粉碎后过 200 目筛,供 FTIR 测试用。

1.3 数据分析

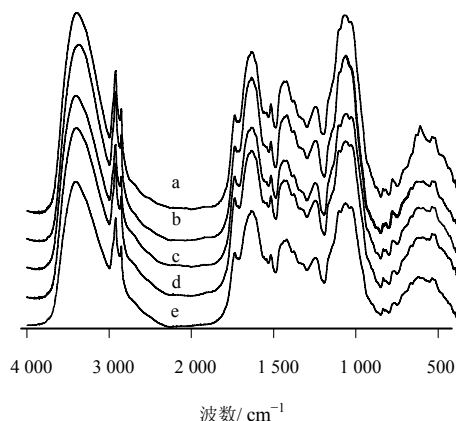
通过测定得到 5 个不同产地共 20 个红花样品的 FTIR 谱图,先将红外吸收谱图矢量归一化处理,通过对吸收峰的吸光度值进行方差分析,以波数 $900\sim 1\,550\text{ cm}^{-1}$ 的红外吸光度值为指标,以不同产地的红花为对象,构建原始数据矩阵。最后对测量结果的主成分分析 (principal component analysis, PCA) 采用 Matlab 7.5 软件自编程序进行,聚类分析 (cluster analysis) 采用德国布鲁克公司 OPUS 光谱软件中的 Cluster analysis 程序进行,得到表征 5 个不同地域红花的相似性关系的主成分三维、二维排序图和聚类图。

2 结果与讨论

2.1 光谱分析

红花作为天然产物,成分复杂,不同产地、品种的红花有着基本一致的化学组成,但也存在着差异^[8-9],从不同地域红花样品红外光谱图 (图 1) 中可以看出,不同产地红花样品的红外光谱存在一定差异,主要表现为各特征峰的相对吸收强度有所变化和指纹区小峰的微小差异,通过方差分析,能明显看出红外光谱图中差异的波数位置和大小,这可能与不同的生境条件差异有关。但其主要化学物质在红外光谱吸收峰中的峰强和出峰位置上都很相似,在 $3\,400$ 、 $1\,042\text{ cm}^{-1}$ 处有强的羟基 (O-H) 伸缩振动和弯曲振动吸收峰,在 $2\,942\text{ cm}^{-1}$ 有强的饱和 C-H 伸缩振动吸收峰,在 $1\,737\text{ cm}^{-1}$ 处的吸收峰是脂肪酸、皂苷和黄酮类化合物中的脂羰基 C=O 伸缩振动吸收,在 $1\,609$ 、 $1\,546\text{ cm}^{-1}$ 左右的肩峰为酰胺化合物的吸收 I 和 II 带,在 $1\,452\text{ cm}^{-1}$ 处的肩峰为 C-N 键的伸缩振动吸收吸收峰,在 $1\,419$ 、 $1\,246\text{ cm}^{-1}$ 处各有一个 C-O 伸缩振动峰,应为红花苷类物质的吸收峰。通过红外光谱分析完全能反映出红花的化学特征,因此以 FTIR 为指纹图谱,通过模式识别技术来描述红花样品的差异性并用于红花的质量鉴别是一种可行的方法。

2.2 光谱特征的选择



a-博乐 (B) b-塔城 (T) c-伊朗 (Y) d-山东 (S) e-四川 (C)
a-Bole (B) b-Tacheng (T) c-Yilan (Y) d-Shandong (S) e-Sichuan (C)

图 1 不同产地红花样品红外光谱图

Fig. 1 FTIR spectra of safflower from different areas

方差的意义在于能描述随机变量波动性大小,为找出合适的光谱数据分析范围,本课题组以所有红花样品的红外光谱吸光度为样本,以波长变化为变量,求出每一波长下的吸光度方差值 $\Sigma = D(X)$,并以方差 Σ 对波长进行做图,得到方差曲线图 (图 2),方差越大表明光谱数据的变异性越大,在此波数下不同产地红外谱图的差别也越大,越能体现出不同产地红花化学物质的不同。从图 2 可以看出不同产地红花的红外光谱在 $900\sim 1\,550\text{ cm}^{-1}$ 方差有较明显的差异。因此最后以特征性较强的 $900\sim 1\,550\text{ cm}^{-1}$ 波段的吸光度值作为红花主成分和聚类分析的原始数据,对此波段采集的原始数据进行 9 点平滑和矢量归一化处理后做为分析数据矩阵。

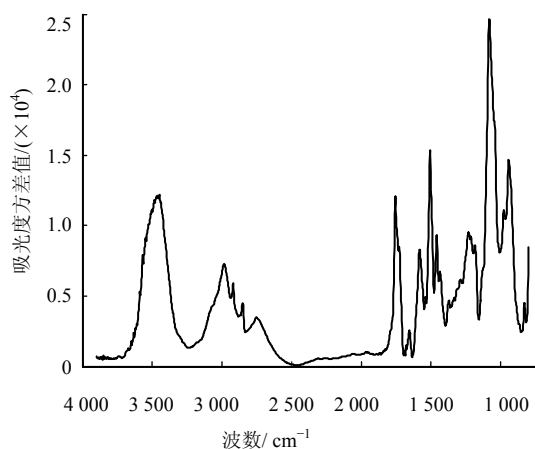


图 2 方差分析图

Fig. 2 Plot of covariance analysis

2.3 主成分分析

对来自 5 个不同地区的 20 个已知产地的红花样本进行了主成分分析。前 3 个主成分的特征值及累积可信度见表 1，第 1 主成分的方差贡献率为 45.18%，是最重要的成分，前 3 个主成分的累积方差贡献率为 86.52%，表示前 3 个主成分可表达全部信息的 86.52%，而有 13.48% 的信息丢失，所以前 3 个主成分基本可以表示原红外光谱的主要信息，并能显示出各产地红花之间的相似性关系。

表 1 特征值和可信度表

Table 1 Eigenvalues and reliabilities

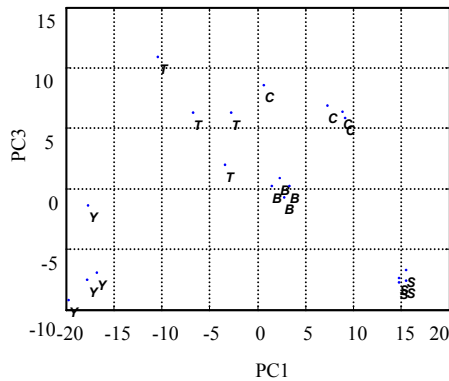
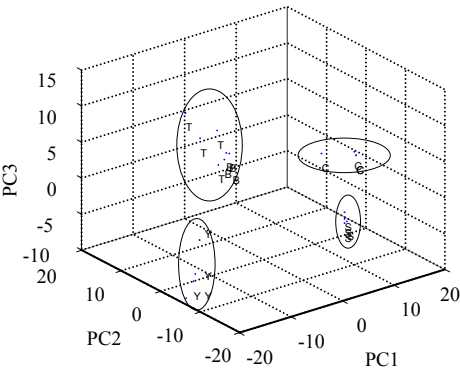
主成分	特征值	累积可信度/%
PC1	138.26	45.18
PC2	84.10	72.63
PC3	42.51	86.52

运用主成分分析程序对红花样品的光谱数据进行运算，得到 5 个不同分布地域红花的主成分分析排序三维散点图和第一、第三主成分的平面投影图(图 3)。从 3D 排序图上可以看出，不同地区红花的红外光谱在排序三维图上的位置关系与它们间的地理距离和相对地理位置有一定的对应性。其中地理

位置接近的新疆两产地的红花样品在三维图上的位置关系最近，并与其他产地有明显差异，四川、山东、伊朗红花的红外光谱也各有特点，在主成分三维图中也都存在明显差异，多样性分化明显。将 3D 和 2D 两个图结合起来，可看出，不同样本的红花由于受地理和人为等因素影响，其化学成分的积累也是不同的，因此在主成分聚类图中呈分布聚集状态。根据这一分布，可将其分为 4 个区域，分别对应着新疆、伊朗、四川和山东产区，在 3D 图中可以看出它们基本都聚在一起，可作为鉴别红花产地的依据，但在 2D 图中分辨率较低，使得新疆塔城(T)、新疆博乐(B)、四川(C)三产区红花样本有聚集在一起的趋势。

从主成分聚类结果和其分布分析结果可以看出，红花主成分聚类分析结果基本与产地相对应，这很可能是由于各地的地理环境和气候条件差异造成一定的片面性，并不能全面反映中药中化学成分的复杂性。不同产地的红花能很好地归类在一起，说明主成分聚类结果是有意义的，能正确判断出红花样品中化学成分的变化。

2.4 聚类分析



B-博乐 T-塔城 Y-伊朗 S-山东 C-四川
B-Bole T-Tacheng Y-Iran S-Shandong C-Sichuan

图 3 红花样品主成分聚类分析图(左边: 3D, 右边: 2D)
Fig. 3 PCA clustering plot of safflower sample (left: 3D, right: 2D)

基于欧式距离，采用 OPUS 软件中的 Cluster analysis 聚类分析程序中的沃德算法(Ward's algorithm)对红花样品的红外光谱进行分析，得到 5 个不同地区红花样品的树形图(图 4)。从结果来看 5 个地区红花之间的距离系数在 0.05~1.0 变动，在欧式距离系数为 0.32 处可划分为 4 个表征群。新疆博乐和新疆塔城红花的距离系数最小(0.16)，分

化不明显，首先聚在一起，这与两者之间的地理距离较近，环境条件相似，居群多样性分化相对较小有关，可做为同一产群来对待。伊朗红花、新疆红花与四川和山东红花分化较大，推测由于伊朗和新疆与另外 2 个产地空间距离较大，造成地理空间间隔，导致红花化学成分的不同。

红花聚类分析结果完全达到指纹图谱的技术要

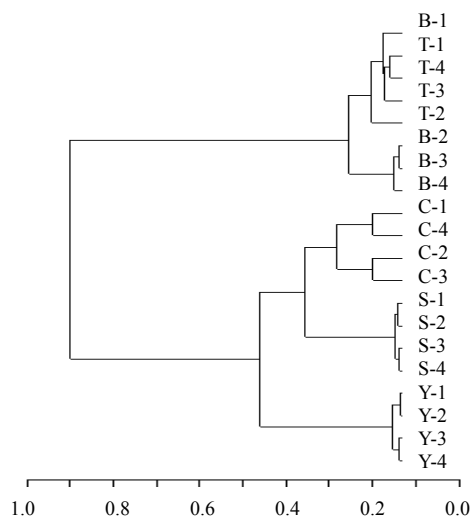


图 4 红花样品聚类分析图

Fig. 4 Clustering plot of safflower sample

求, 在距离系数为 0.32 时, 结果与主成分分析结果基本一致, 这也说明红花 FTIR 谱图能够反映出它们各自化学组成的差异性, 因此, 用聚类分析的方法能较全面、综合地反映出不同产地红外谱图之间的多样性分化, 能够对红花样品混合成分的 FTIR 指纹图谱进行鉴别, 可为保证红花产品质量的一致性提供依据。

3 结论

对不同产地红花样品的主成分分析和聚类分析均表明, 分布地不同会造成红花在化学组成上存在着差异, 同时也说明地理位置对红花质量是有影响的, 只有采取对其复杂化学组分进行全面综合的评价方法, 才能保障红花产品质量的一致性。本实验采用 FTIR 方法测定了不同产地红花的红外光谱图, 使用主成分分析和聚类分析对不同地理区域红花样品的 FTIR 谱图进行了分析, 其结果能够很好地反映出红花中的整体化学组分变化, 同一产地红花具有相似的化学组分, 能够很好地归在同一类别中。

该方法具有客观全面的优点, 是一种简便易行, 快速准确, 且具有一定实用性的鉴别方法。

参考文献

- [1] 吴瑾光. 近代傅立叶变换红外光谱技术及应用 [M]. 北京: 科学技术文献出版社, 1994.
- [2] 徐永群, 孙素琴, 袁子民, 等. 红外光谱结合主成分分析鉴别道地山药 [J]. 分析化学, 2002, 30(10): 1231-1233.
- [3] 徐永群, 黄 吴, 周 群, 等. 红外特征图谱和聚类分析法在赤芍产域分类鉴别种的应用 [J]. 分析化学, 2003, 31(1): 5-9.
- [4] 程存归, 郭水良, 陈宗良. FTIR 直接测定法对川贝母和珠贝、小东贝的区别鉴定 [J]. 中草药, 2002, 33(3): 262-264.
- [5] 毛建江, 张玉杰, 徐经纬, 等. 傅立叶近红外光谱用于鹿茸的定性分析 [J]. 理化检验: 化学分册, 2007, 43(1): 33-36.
- [6] Lamprell H, Mazerolles G, Kodjo A, *et al.* Discrimination of *Staphylococcus aureus* strains from different species of *Staphylococcus* using Fourier transform infrared (FTIR) spectroscopy [J]. *Int J Food Microbiol*, 2006, 108: 125-129.
- [7] 中国药典 [S]. 二部. 2010.
- [8] 王玉英, 林慧萍, 李水福. 西红柿的真伪优劣鉴定 [J]. 中草药, 2010, 41(7): 1194-1195.
- [9] 刘绍华, 黄世杰, 胡志忠, 等. 藏红花挥发油的 GC-MS 分析及其在卷烟中的应用 [J]. 中草药, 2010, 41(11): 1790-1792.
- [10] 刘月庆, 王 睿, 毕开顺. HPLC 法测定红花中红花黄色素 A 的含量 [J]. 药物分析杂志, 2004, 24(4): 356-358.
- [11] 赵明波, 邓秀兰, 王亚玲, 等. 红花 RP-HPLC 指纹图谱的建立及其质量研究 [J]. 药学报, 2004, 39(3): 212-216.
- [12] 王若菁, 杨 滨, 付梅红. 红花药材的质量评价 [J]. 中国中药杂志, 2008, 33(22): 2642-2646.
- [13] 张志祥, 刘 鹏, 康华靖, 等. 基于主成分分析和聚类分析的 FTIR 不同地理居群香果树多样性分化研究 [J]. 光谱学与光谱分析, 2008, 28(9): 2081-2085.