

5 种秦艽组植物近红外判别模型的建立

李佩佩^{1,2}, 李 朵^{1,2}, 栾真杰^{1,2}, 孟晓萍¹, 周玉碧¹, 孙 菁^{1*}¹青海省青藏高原特色生物资源研究重点实验室 中国科学院西北高原生物研究所, 西宁 810008; ²中国科学院大学, 北京 100049

摘要:本文以 5 种秦艽植物为研究对象, 建立正品秦艽与非正品秦艽(模型 1)及不同秦艽物种(模型 2)的近红外判别模型, 并与聚类分析比较。以模型 1 为代表进行建模条件优化, 经优化, 利用一阶导数谱图经 discriminant analysis 建模, 两个模型预测率与识别率均达到 100%, 可对正品秦艽与非正品秦艽及样品种类进行准确判别。利用 5 种植物的吸光度进行聚类分析, 结果, 聚类分析能准确判别正品秦艽与非正品秦艽, 但无法对秦艽种类判别, 表明近红外判别模型比聚类分析具有更广泛的实用性及准确度, 可为秦艽药材临床用药的安全性和有效性提供快速鉴别的科学依据。

关键词: 麻花苳; 达乌里秦苳; 秦苳; 粗茎秦苳; 黄管秦苳; 近红外判别模型

中图分类号: 0657.33

文献标识码: A

文章编号: 1001-6880(2019) Suppl-0040-07

DOI: 10.16333/j.1001-6880.2019.S.008

Establishment of near-infrared discriminant model of five *Cruciata* GaudinLI Pei-pei^{1,2}, LI Duo^{1,2}, LUAN Zhen-jie^{1,2}, MENG Xiao-ping¹, ZHOU Yu-bi¹, SUN Jing^{1*}

¹Qinghai Key Laboratory of Qinghai-Tibet Plateau Biological Resource, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810008, China;

²University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Setting the 5 *Gentiana* species-*G. straminea*, *G. dahurica*, *G. macrophylla*, *G. crassicaulis*, *G. officinalis*-as the research objects. Near-infrared combined with chemometrics were used to build two discriminant models. Model 1 was to discriminant authentic *Radix Gentiana Macrophyllae* from falsify; model 2 was to distinguish 5 different species. And the cluster analysis was also proceeded to make a comparison. Model 1 was taken to optimize modeling approaches and spectral processing methods. Finally, two models were established by discriminant analysis, with no-smoothing first-derivative spectra. And the recognition rates and prediction rates of two models were all 100%. The NIR absorbance was analyzed by cluster analysis. The result showed that cluster analysis could discriminant authentic gentiana from adulterants exactly, but could not identify 5 different species. The analysis indicated that discriminant models have more practicability and accuracy. They can provide a rapid and scientific method on clinical medication safety and effectiveness of *Radix Gentiana Macrophyllae*.

Key words: *G. straminea*; *G. dahurica*; *G. macrophylla*; *G. crassicaulis*; *G. officinalis*; near-infrared discriminate model

秦艽为传统中药材, 2015 版《药典》规定秦艽饮片为秦艽(*Gentiana macrophylla* Pall.)、麻花苳(*G. straminea* Maxim.)、粗茎秦苳(*G. crassicaulis* Duthie ex Burk.)或达乌里秦苳(*G. dahurica* Fisch., 又名小秦苳)的干燥根^[1], 其均为龙胆科(*Gentianaceae*)龙胆属(*Gentiana*)秦艽组(*Sect. Cruciata*)多年生草

本。有祛风湿、清湿热、止痹痛等功效, 用于治疗风湿痹痛、中风、湿热黄疸等症^[1], 2015 版《药典》收录配伍有秦艽的复方有 14 种, 其资源用量较大。

黄管秦苳(*G. officinalis* H. Smith)为龙胆科龙胆属秦苳组多年生草本, 与粗茎秦苳种源关系较近, 株高 15~35 cm, 花冠黄绿色, 生于高山草甸、灌丛及河滩等地, 海拔 2 300~4 200 m^[4,5]。黄管秦苳与“秦苳”中主要化学成分均为龙胆苦苷、马钱苷酸等环烯醚萜苷类^[6,7], 在甘肃、青海等地有一定的用药历史, 研究报道表明, 黄管秦苳有治疗骨关节炎的作用^[8-10], 但未被收录入 2015 版《药典》。

收稿日期: 2019-04-19 接受日期: 2019-05-06

基金项目: 青海省重点实验室发展建设专项(2017-ZJ-Y10); 国家自然科学基金青年基金(81403051); 青海省自然科学基金面上项目(2019-ZJ-904); 青藏高原特色生物资源工程技术研究中心能力建设项目(2017-GX-G03)

* 通信作者 E-mail: sunj@nwipb.cas.cn

近红外 (near infrared, NIR) 光谱技术不需对样品进行复杂的处理,操作简单、快捷^[11-15],结合化学计量学方法使用时,通过建立定性、定量模型可快速对样品进行分析^[16-18],已广泛用于食品、药品、石油等领域^[19-21]。

因此,本研究以五种秦艽组植物为研究对象,利用近红外光谱技术结合化学计量学分析方法,建立5种秦艽组植物的近红外判别模型,对正品秦艽与非正品秦艽以及不同秦艽物种进行判别,以期为秦艽药材临床用药安全性提供科学依据。

1 材料与方法

1.1 材料

中药材秦艽的4种基源植物(麻花苳、小秦苳、秦苳、粗茎秦苳)及黄管秦苳共5种秦苳组植物的根部,均采自2018年8月底花期。其中,麻花苳采自青海省兴海县,有32份样品;小秦苳采自青海省天峻县,有9份样品;粗茎秦苳采自青海省果洛藏族自治州班玛县,有12份样品;秦苳采自青海省大通县种植基地,有19份样品;黄管秦苳采自青海省民和县,有47份样品。

样品带回实验室清洗、晾干,取根粉碎、备用。原植物标本经中国科学院西北高原生物研究所卢学峰研究员分别鉴定为麻花苳 (*Gentiana straminea* Maxim.)、小秦苳 (*Gentiana dahurica* Fisch.)、秦苳 (*Gentiana macrophylla* Pall.)、粗茎秦苳 (*Gentiana crassicaulis* Duthie ex Burk.)、黄管秦苳 (*Gentiana officinalis* H. Smith)。

1.2 仪器

傅里叶变换红外光谱仪近红外模块 (Nicolet is50, Thermo Fisher, 美国);样品筛(100目,浙江上虞市道墟

张兴纱筛厂,中国),样品杯(Thermo Fisher, 美国)。

1.3 实验方法

样品药用部位根部粉碎后过100目筛,取0.1365 g (± 0.0005)样品置于玻璃样品杯中,装样厚度约0.1 cm,以空气为参比,在10 000~4 000 cm^{-1} 范围内进行近红外全光谱扫描 ($n=3$),扫描分辨率为8 cm^{-1} ,扫描次数为64次。将得到的5个秦苳组植物 NIR 谱图数据导入 TQ Analysis 中,利用软件自带的建模方法、谱图处理方法进行建模方法的优化,利用优化的建模方法建立近红外判别模型。

所建近红外判别模型有两种,分别为,用以判别正品秦苳与非正品秦苳的模型,即模型1;用以判别不同秦苳物种的模型,即模型2。建模方法的优化以模型1为代表进行。

2 结果与讨论

2.1 5种秦苳组植物近红外光谱

5种秦苳组植物近红外一维全光谱采集结果如图1所示。由图1可知,各物种近红外一维谱图较相似,其中,麻花苳与小秦苳相似度较大,在4 300 cm^{-1} 及4 200 cm^{-1} 附近都有吸收峰;粗茎秦苳、秦苳、黄管秦苳均只在4 200 cm^{-1} 附近有一个吸收峰;各样品吸收强度有差异。依据近红外指纹图谱不能判别正品秦苳与非正品秦苳,及对各秦苳组物种的鉴别。

2.2 近红外判别模型的建立

2.2.1 建模条件的优化

以模型1为例进行模型建立条件的优化,将5种秦苳植物样品近红外均谱 ($n=3$)数据导入 TQ Analysis 软件建立判别模型1。建模时随机挑选2/3样品做训练集,1/3样品做验证集,模型1样品建模集及验证集数量见表1。

表1 模型1建模集及验证集样品数量

Table 1 Sample numbers of calibration and validation sets of discriminant model 1

| 种类 Species | 样品集(个) Calibration set | 验证集(个) Validation set | 总 Sum |
|---|---------------------------|--------------------------|----------|
| 正品秦苳 Authentic <i>Radix Gentiana Macrophyllae</i> | 48 | 24 | 72 |
| 非正品秦苳 Audeleterant <i>Radix Gentiana Macrophyllae</i> | 32 | 15 | 47 |
| 总 Sum | 80 | 39 | 119 |

2.2.1.1 建模方法的优化

在谱图为原谱图且无平滑的条件下对两种建模方法 (discriminant analysis 及 distance match) 进行优化,结果见表2。由表2知, discriminant analysis 方

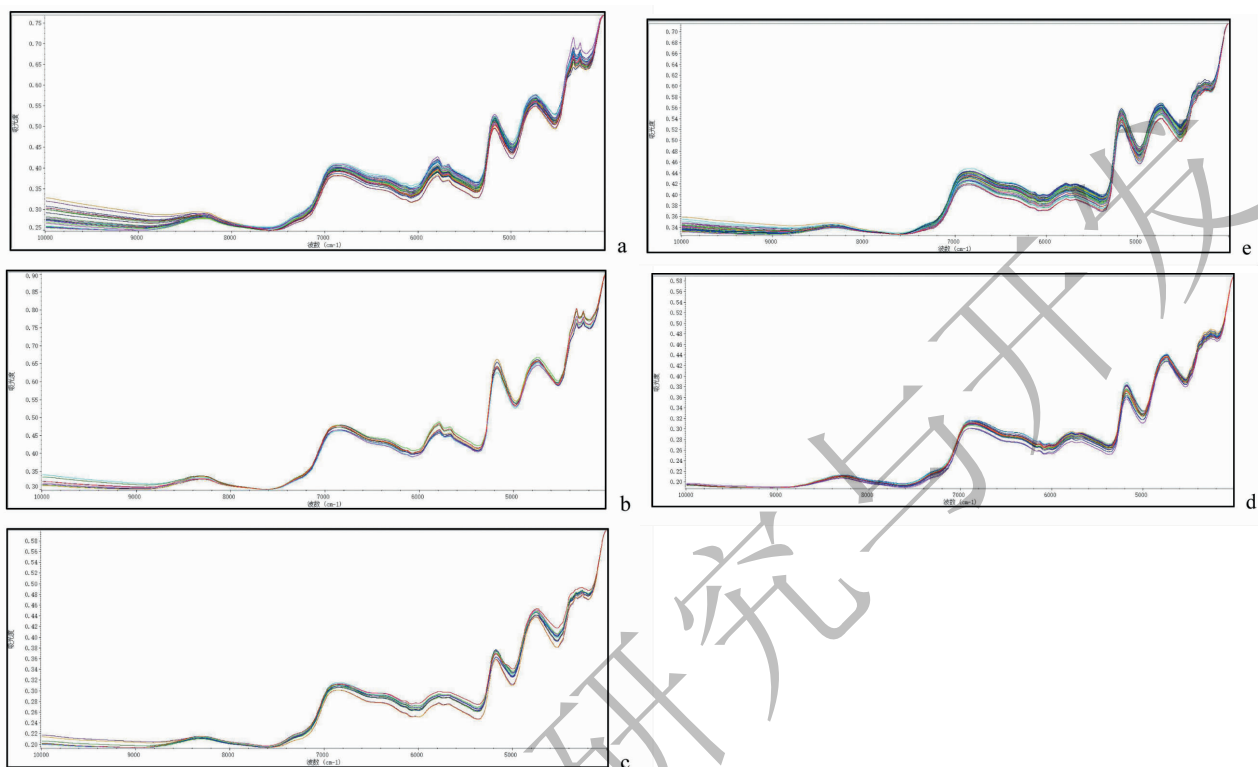
法的识别率与预测率均高于 distance match 方法,故建模方法选择 discriminant analysis 方法。

2.2.1.2 光谱处理方法的优化

在建模方法为 discriminant analysis 且谱图无平

滑的条件下对光谱处理方法(原谱图、一阶导谱图、二阶导谱图)进行优化,结果见表3。由表3知,一阶导数谱图建模时,模型识别率及预测率均达到了

100%,效果优于原谱图及二阶导数谱图建模,因此建模时谱图处理方法为一阶导数谱图。



注:a:麻花苧近红外谱图;b:小秦艽近红外谱图;c:粗茎秦艽近红外谱图;d:秦艽近红外谱图;e:黄管秦艽近红外谱图

Note:a:NIR spectra of *G. straminea*;b:NIR spectra of *G. dahurica*;c:NIR spectra of *G. crassicaulis*;

d:NIR spectra of *G. macrophylla*;e:NIR spectra of *G. officinalis*

图1 5种秦艽组植物近红外谱图

Fig. 1 NIR spectra of five cruciata gaudin

表2 不同建模方法所建判别模型效果

Table 2 Results of discriminant models by different processing methods

| 建模方法 Modeling approach | 识别率 Recognition rate (%) | 预测率 Prediction rate (%) | 总 Sum (%) |
|---------------------------|--------------------------------|-------------------------------|-----------------|
| Distance match | 77.50 | 74.36 | 76.47 |
| Discriminant analysis | 98.75 | 100.00 | 99.12 |

表3 不同光谱处理方法所建判别模型效果

Table 3 Results of discriminant models by different spectral processing methods

| 谱图处理 Spectrogram processing | 识别率 Recognition rate (%) | 预测率 Prediction rate (%) | 总 Sum (%) |
|--------------------------------|--------------------------------|-------------------------------|-----------------|
| 原谱图 Original spectra | 98.75 | 100.00 | 99.12 |
| 一阶导数 First derivative spectra | 100.00 | 100.00 | 100.00 |
| 二阶导数 Second derivative spectra | 98.75 | 97.44 | 98.32 |

2.2.1.3 平滑处理方法的优化

在建模方法为 discriminant analysis、谱图为一阶导数谱图的条件下,对平滑方法(无平滑、SG 平滑、

Norris 平滑)进行优化,结果如下(表 4)。由表 4 知,用原谱图所建模型识别率及预测率均达到了 100%。因此,建模时所用谱图不经平滑处理。

表 4 不同平滑处理方法所建判别模型效果

Table 4 Result of discriminant models by different smoothing methods

| 谱图平滑 Spectrogram smooth | 识别率 Recognition rate (%) | 预测率 Prediction rate (%) | 总 Sum (%) |
|----------------------------|--------------------------------|-------------------------------|-----------------|
| 无平滑 No smoothing | 100 | 100 | 100 |
| SG 平滑 SG | 98.75 | 100 | 99.12 |
| Norris 平滑 Norris | 98.75 | 100 | 99.12 |

2.2.2 近红外判别模型的建立

根据以上试验结果,近红外判别模型建立时,建模方法为 discriminant analysis,光谱处理方法为一阶导数谱图处理,平滑处理方法为无平滑处理,模型 1 识别率为 100%,预测率为 100%,说明所建模型效果良好,可以很好地对正品秦艽及非正品秦艽进行判别。

将 5 种秦艽组植物样品近红外均谱($n=3$)数据导入 TQ Analysis 软件建立判别模型 2,用于 5 种秦艽组植物种类的判别。建模时,随机挑选 2/3 样品做训练集,1/3 样品做验证集,5 种秦艽组植物样品建模集及验证集数量如下(表 5)。根据模型 1 建模方法,模型 2 的建模方法为 discriminant analysis,光谱处理方法为一阶导数谱图处理,谱图不经平滑处理,模型的识别率与预测率均达到了 100%。

在样品研究中,常需准确判别样品粉末的种类,

表 5 模型 2 建模集及验证集样品数量

Table 5 Sample numbers of calibration and validation sets of discriminant model 2

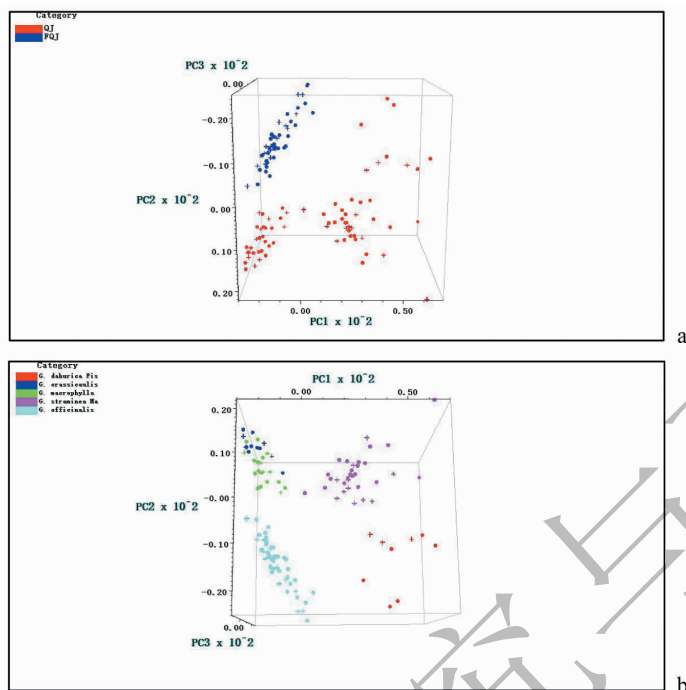
| 种类 Species | 训练集(个) Calibration rate | 验证集(个) Validation rate | 总 Sum |
|-----------------------------|----------------------------|---------------------------|----------|
| 麻花艽 <i>G. straminea</i> | 21 | 11 | 32 |
| 小秦艽 <i>G. dahurica</i> | 6 | 3 | 9 |
| 粗茎秦艽 <i>G. crassicaulis</i> | 8 | 4 | 12 |
| 秦艽 <i>G. macrophylla</i> | 13 | 6 | 19 |
| 黄管秦艽 <i>G. officinalis</i> | 32 | 15 | 47 |
| 总 Sum | 80 | 39 | 119 |

近红外判别模型 1 及模型 2 的 3D 判别结果见图 2。由图可知,近红外判别模型对正品秦艽与非正品秦艽判别效果好,类别间样品无交叉;模型对 5 种秦艽组植物判别效果较好,秦艽与粗茎秦艽样品有少量交叉,其余物种分离效果好。结合模型 1 及模型 2 的预测率、识别率(均为 100%),说明本实验所建近红外判别模型不仅能判别出正品秦艽与非正品秦艽,还能准确进行物种判别。Patrizia Firmani^[22]等利用偏最小二乘-Discriminant Analysis 方法(PLS-DA)对大吉岭黑茶及其不同程度的掺假品进行判别,所建方法对大吉岭黑茶及其掺假品的鉴别正确率分别为 100%、98.01%;Maribel Alexandra

Quelal-Vasconez^[23]等利用 PLS-DA 方法对可可粉中可可壳的掺假进行鉴别,所建模型对可可壳掺假程度 < 5% 及 5-40% 判别的准确率为 100%。本实验所用建模方法为 Discriminant Analysis(DA),所建模型 1 及模型 2 的预测率与识别率均达到了 100%,模型效果良好;对大吉岭及可可粉的研究建模所用方法为 DA 结合偏最小二乘,所建模型判别效果与本实验所建模型效果相当,表明 DA 用于判别模型建模适用范围较广。

对建模样品进行主成分分析(表 6),前三主成分对模型的累计贡献率达到了 95.717%,代表了样品的大部分信息,说明模型具有一定的可信度。其

中,第一主成分贡献率为 72.566%,第二主成分贡献率为 19.060%,第三主成分贡献率为 4.091%。



注:a:模型1结果;b:模型2结果。

Note:a:Result of model 1;b:Result of model 2.

图2 近红外判别模型

Fig. 2 Near-infrared discriminant model of gentiana

表6 近红外判别模型前三主成分

Table 6 The first three principal components of Near-infrared discriminant model

| 主成分 Principal components | 贡献率 Contribution rate (%) | 累计贡献率 Accumulating contribution rate (%) |
|-----------------------------|------------------------------|---|
| 1 | 72.566 | 72.566 |
| 2 | 19.060 | 91.625 |
| 3 | 4.091 | 95.717 |

2.3 聚类分析

将5种秦艽组植物吸光度数据导入PC-ORD软件,缺失值用随机森林处理,距离测定为correlation,组间联系方法为Ward's Method,得出样品聚类结果(图3)。由图3知,在信息保留量为75%时,样品分为两类,一类为黄管秦艽,即非正品秦艽,一类为麻花艽、秦艽、小秦艽、粗茎秦艽,即正品秦艽,聚类分析用以正品秦艽与非正品秦艽判别时正确率为100%,但不能将5种植物区分开,其判别效果不如所建判别模型。

陈叶青^[24]利用性状、显微、薄层色谱及HPLC

方法进行秦艽及混淆品牛扁的鉴别;高松^[25]等通过观察秦艽与麻布七的横切面,并比较荧光法、显色反应、HPLC中龙胆苦苷的出峰时间进行秦艽与其伪品麻布七的鉴别。现有对秦艽及其伪品的鉴别多通过显微鉴别、性状鉴别、生化鉴别,这些方法耗时、耗力,无法进行快速鉴别,本实验所建模型可进行快速、实时在线鉴别。

3 结论

以5种秦艽组植物麻花艽、小秦艽、秦艽、粗茎秦艽、黄管秦艽为研究对象,建立两种近红外判别模型,所建模型不仅能判别正品秦艽与非正品秦艽,还

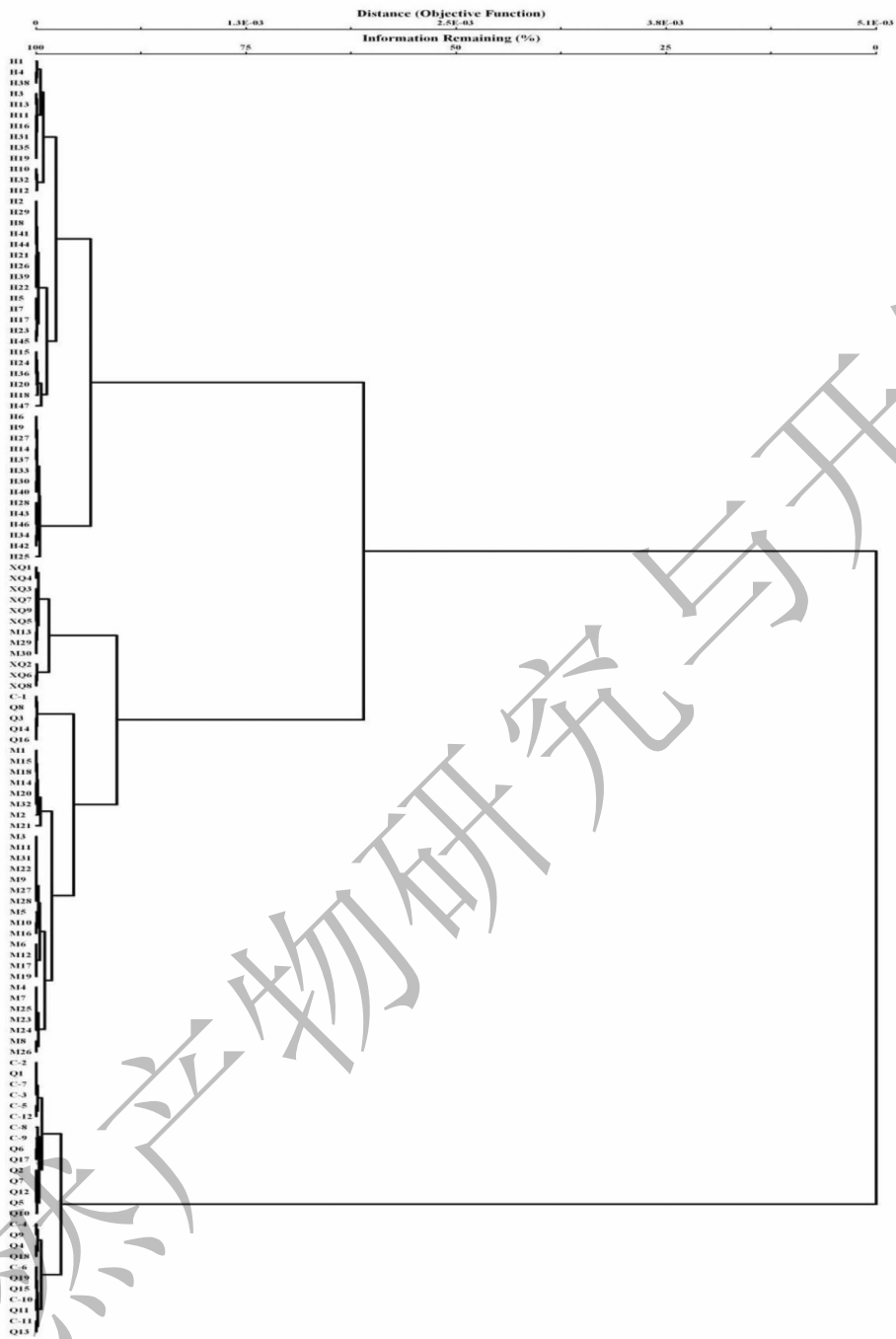


图3 样品聚类分析结果

Fig. 3 Cluster analysis result of all samples

注: M 为麻花艽, XQ 为达乌里秦艽, C 为粗茎秦艽, Q 为秦艽, H 黄管秦艽。

Note: M means *G. straminea*, XQ means *G. dahurica*, C means *G. crassicaulis*, Q means *G. macrophylla*, H means *G. officinalis*

能准确进行物种判别。两个模型均利用未经平滑的一阶导数谱图经 discriminant analysis 建模,其预测率与识别率均达到了 100%。利用样品的吸光度进行聚类分析,发现聚类分析可以将秦艽组植物与非秦艽组植物准确判别出来,但不能准确的判别不同

物种样品。说明利用近红外判别模型可以更快速、准确的对不同用途的样品进行判别。

参考文献

1 Chinese Pharmacopoeia Commission. Pharmacopoeia of the

- People's Republic of China: Vol I (中华人民共和国药典:第一部) [M]. Beijing: China Medical Science Press, 2015:270-271.
- 2 Chinese Flora Commission of Chinese Academy of Sciences. Flora reipublicae popularis sinicae; Vol 62 (中国植物志:第62卷) [M]. Beijing: Science Press Ltd, 1988:62-73.
 - 3 Northwest Institute of Plateau Biology, the Chinese Academy of Sciences. Flora Qinghaiica; Vol 3 (青海植物志:第3卷) [M]. Xining: Qinghai People's Publishing House, 1996:47-55.
 - 4 Chinese Flora Commission of Chinese Academy of Sciences. Flora reipublicae popularis sinicae; Vol 62 (中国植物志:第62卷) [M]. Beijing: Science Press Ltd, 1988:72.
 - 5 Northwest Institute of Plateau Biology, the Chinese Academy of Sciences. Flora Qinghaiica; Vol 3 (青海植物志:第3卷) [M]. Xining: Qinghai People's Publishing House, 1996:55.
 - 6 Zhou DW, Hou QZ, Si QW, et al. Concentration of the active constituents of the Tibetan Folk Medicine Qinjiao (*Gentiana sect. Cruciata*) within and between Taxonomic species across the Qinghai-Tibetan Plateau [J]. Chem & Bio, 2010, 7:2088-2094.
 - 7 Wang J, Zhu SJ, Fan Q, et al. Quality evaluation of *Gentiana officinalis* H. Smith. from different producing areas of Gansu province by grey incidence degree method [J]. Chin J Inf Tradit Chin Med (中国中医药信息杂志), 2017, 24:69-74.
 - 8 Northwest Institute of Plateau Biology, Chinese Academy of Science. Flora Tibetan (藏药志) [M]. Xining: Qinghai Ethnic Publishing House, 1991:113-114.
 - 9 Wang J. Study on the quality evaluations of *Gentiana officinalis* H. Smith in Gansu and the effect of its available substances on experimental osteoarthritis in rats [D]. Lanzhou: Gansu University of Chinese Medicine (甘肃中医药大学), 2016.
 - 10 Li XB, Ye J, Zhao L, et al. Effects of *Gentiana officinalis* from total secoiridoid glucoside on Interleukin-1 β (IL-1 β)-induced rat chondrocytes [J]. Chin Pharm J, 2014, 49:1121-1125.
 - 11 Claudia ATS, Ricardo NP, Miguel L, et al. Application of portable Near-infrared spectrometers [J]. Ency Ana Chem, 2015:1-27.
 - 12 Li Y, Zhang J, Jin H, et al. Study of the underground parts identification and saponins content prediction of panax notoginseng based on FTIR combined with chemometrics [J]. Spectrosc Spect Anal, 2019, 39:103-108.
 - 13 Nikzad-Langerodi R, Ortmann S, Pferschy-Wenzig EM, et al. Assessment of anti-inflammatory properties of extracts from Honeysuckle (*Lonicera sp. L.*, Caprifoliaceae) by ATR-FTIR spectroscopy [J]. Talanta, 2017, 175:264-272.
 - 14 Chen JB, Sun SQ, Tang XD, et al. Direct and simultaneous detection of organic and inorganic ingredients in herbal powder preparations by Fourier transform infrared microspectroscopic imaging [J]. Spectrochim Acta A: Mol Bio Spect, 2016, 165:176-182.
 - 15 Bahareh J, Ezeddin M, Hossein F, et al. Pattern recognition-based optical technique for non-destructive detection of *Ectomyelois ceratoniae* infestation in pomegranates during hidden activity of the larvae [J]. Spectrochim Acta A: Mol Bio Spect, 2019, 206:552-557.
 - 16 Li D, Wang B, Li X, et al. Identification of dyeing adulteration in Amur Cork-tree Bark by surface-enhanced Raman spectroscopy [J]. J Pharm Res (药学研究), 2015, 34:202-205.
 - 17 Wang GL, Tian JG, Nei LX, et al. Identification of Radix Ginseng Rubra by near infrared spectroscopy [J]. Chin Tradit Herbal Drugs (中草药), 2008, 39:438-440.
 - 18 Xiao R, Liu L, Zhang DJ, et al. Discrimination of organic and conventional rice by chemometric analysis of NIR spectra: a pilot study [J]. J Food Meas Charact, 2019, 13:238-249.
 - 19 Wang XX, Li JT, Sun XL, et al. Study on rapid determination of Oligosaccharides in Soybean products by near-infrared reflectance spectroscopy [J]. Spectrosc Spect Anal, 2018, 38:58-61.
 - 20 Yi ZK. Study on the high-throughput analysis of Ephedra Sinica plants with diffuse reflectance FT-NIR and chemometrics [D]. Chongqing: Chongqing Medical University (重庆医科大学), 2012.
 - 21 Shen SY, Sun WJ, Tang Q, et al. Rapid determination of total water-soluble sugars in white tea by near-infrared spectroscopy [J]. Nat Prod Res Dev (天然产物研究与开发), 2019, 31:16-23.
 - 22 Patrizia F, Silvia DL, Remo B, et al. Near infrared (NIR) spectroscopy-based classification for the authentication of Darjeeling black tea [J]. Food Control, 2019, 100:292-299.
 - 23 Maribel AQV, Maria JLG, Edgar PE, et al. Fast detection of cocoa shell in cocoa powers by near infrared spectroscopy and multivariate analysis [J]. Food Control, 2019, 99:68-72.
 - 24 Chen YQ. Identification of *Radix Gentiana Macrophyllae* and *Aconitum barbatum* Patr in ex Pers. var. puberulum Ledeb. [J]. Chin Pharm (中国药业), 2009, 18(22):66.
 - 25 Gao S, Wei J, Liu YF, et al. Identification of *Radix Gentiana Macrophyllae* and *Aconitum Sinomontanum* Nakai. by microscopic and physiochemical identification [J]. Chin Arch Tradit Chin Med (中华中医药学刊), 2007, 25:2638-2639.