

基于数据挖掘的道地产区当归不同炮制品质量评价方法的建立

郭延生, 华永丽, 杜天玺, 杨洪申, 曲亚玲, 魏彦明*

(甘肃农业大学动物医学院, 甘肃 兰州 730070)

摘要: 采用 SPSS Clementine11.0 数据挖掘软件对当归不同炮制品研究所得数据进行分析, 建立当归不同炮制品分类模型, 并确定其分类参数, 从而建立当归不同炮制品分类和质量评价方法。判别分析从 59 个指标中选择了 8 个指标为预测变量确立的当归不同炮制品的 Fish's 判别函数, 回代判别率为 96.7%。本研究结果表明, 采用数据挖掘软件能够准确、可靠的对当归不同炮制品进行识别和验证, 为当归炮制品的分类和质量评价提供了一定的科学依据。

关键词: 当归; 炮制品; 数据挖掘; 判别分析

中图分类号: R917

文献标识码: A

文章编号: 0513-4870 (2010) 09-1155-05

Establishment of a quality evaluation method for *Angelica* different processed products from genuine producing areas based on data mining

GUO Yan-sheng, HUA Yong-li, DU Tian-xi, YANG Hong-shen, QU Ya-ling, WEI Yan-ming*

(Faculty of Veterinary Medicine, Gansu Agriculture University, Lanzhou 730070, China)

Abstract: The paper reports the development of a quality evaluation method for *Angelica* different processed products. The data of high-performance liquid chromatography, water, total ash and extract were analyzed with SPSS Clementine11.0 software. Discriminant analysis (DA) established the classification model and parameter for *Angelica* different processed products. Fish's discriminant functions of *Angelica* different processed products were generated using 8 predictor variables selected from 59 indexes. The correct rate of discriminating back substitution is 96.7%. *Angelica* different processed products can be accurately and reliably recognized and validated with DA of SPSS Clementine11.0 software.

Key words: *Angelica*; processed product; data mining; discriminant analysis

当归为伞形科当归属植物当归 (*Angelica sinensis* (Oliv.) Diels) 的干燥根^[1], 主产于甘肃岷县、漳县、渭源等地, 是甘肃的道地药材之一。道地药材是传统中医药对产于特定地域的优质中药材的称谓, 被视为评价中药质量的综合标准, 是公认的优质药材^[2]。中医药临床根据实际需要, 对当归采用不同的炮制处理, 使其发挥不同的疗效。当归常用炮制品有油当归、酒当归、土当归、当归炭等^[3], 在临床实践中

有必要对其炮制品进行正确的鉴别和评价, 但其同一中药不同炮制品在外观性状、理化鉴别、化学成分上比较相似, 对同一中药不同炮制品进行质量评价存在一定的困难。Clementine 软件 (SPSS 公司) 在中药相关研究中的数据挖掘应用尚处于起步阶段, 仅有对 GC-MS 数据进行分析的报道^[4]。本研究应用 SPSS Clementine11.0 数据挖掘软件对当归不同炮制品研究所得到的数据进行预处理后, 应用判别分析 (discriminant analysis, DA) 进行数据探索, 以达到全面认识当归不同炮制品的特点, 为当归及其不同炮制品的质量评价和炮制原理研究提供科学依据。

收稿日期: 2010-01-09.

基金项目: 国家自然科学基金资助项目 (30972210).

*通讯作者 Tel: 86-931-7632110, Fax: 86-931-7631229,

E-mail: weiyym@gsau.edu.cn

材料与方法

仪器 美国 Agilent 1100 型高效液相色谱仪 (包括 G1322A 型真空脱气机、G1315B 型 DAD 检测器、G1312A 型二元梯度泵、G1316A 型柱温箱)、Agilent 1100 液质联用仪。

试剂和药品 阿魏酸对照品购自中国药品生物制品检定所 (批号 0773-9910)。当归药材 (全根) 于 2008 年 9 月采自甘肃岷县、漳县、渭源、宕昌等当归主产区, 由甘肃农业大学动物医学院魏彦明教授鉴定为 *A.sinensis* (表 1)。

当归不同炮制品的制备^[1, 3] 酒当归: 取当归片, 加黄酒 (每 100 kg 当归, 用黄酒 10 kg) 拌匀, 稍闷, 待酒被吸尽后, 用文火炒至深黄色, 微具酒香气时, 取出放凉。土当归: 将土粉 (每 100 kg 当归, 用土粉 30 kg) 置锅内炒热, 倒入当归片, 微炒至当归片粘满细土时 (称为挂土) 取出, 筛去土粉, 放凉即可。当归炭: 取当归片, 置锅内, 用中火炒至外表焦黑色、内部棕褐色, 略具焦气, 质枯脆时, 取出放凉。油当归: 取净当归片, 加入 3% 的香油拌匀, 闷润, 文火炒至色泽加深, 微显焦斑, 待显油润有香气时出锅, 晾凉。样品编号见表 1。

当归及不同炮制品的测定

水分、总灰分、酸不溶性灰分及浸出物含量的测定 按《中国药典》2005 年版一部 (附录), 分别测定当归样品中的水分、总灰分、酸不溶性灰分、醇溶性浸出物。

HPLC 方法的建立 ODS-C₁₈ (25 cm × 4.6 mm ID, 5 μm) 分析柱; 流动相: 1% 乙酸水 (A 相)–乙腈

(B 相); 梯度洗脱条件: B, 0~10 min (20%~30%); 10~20 min (30%~49%); 20~40 min (49%); 40~50 min (49%~100%); 50~60 min (100%~20%)。流速: 1 mL·min⁻¹; 柱温: 25 °C; 检测波长: 280 nm。

阿魏酸对照品溶液制备 称取阿魏酸对照品 4.8 mg, 加甲醇至 100 mL 量瓶中, 得 4.8 μg·mL⁻¹ 对照品溶液。

供试品溶液制备 精密称取当归粉末 (40 目) 0.5 g, 置 50 mL 具塞三角瓶中, 精密量取甲醇-甲酸 (95:5) 溶液 25 mL 加入三角瓶中, 密封, 称重, 超声提取 60 min, 放置至室温, 称重, 补足失重, 用 0.45 μm 滤膜过滤, 取续滤液进样 10 μL。

方法学考察

精密度 取同一份当归样品的供试液, 重复进样 5 次, 计算各主要色谱峰的相对保留时间 RSD 为 0.03%~0.14%, 峰面积比值 RSD 为 0.54%~2.06%, 符合特征图谱的检测要求。

稳定性 取同一份当归样品的供试液, 分别于 0、3、6、9、12 和 15 h 检测, 计算各主要色谱峰的相对保留时间 RSD 为 0.03%~0.12%, 峰面积比值 RSD 为 0.29%~2.66%, 表明供试液在 15 h 内基本稳定。

重复性 取当归样品 6 份, 分别精密称定, 按供试品溶液制备方法平行制备供试液, 再分别按上述液相色谱条件测定, 计算各主要色谱峰的相对保留时间 RSD 为 0.28%~0.46%, 峰面积比值 RSD 为 0.75%~1.82%, 符合特征图谱的检测要求。

波长选择 根据文献^[5-9]报道及实验结果选择

Table 1 Origins of *Angelica* and sample code of its processed products. Samples of s5, s9, s11 and j1 were lost and not included in data analysis

Unprocessed <i>Angelica</i> No.	Source of Gansu	Code of processed products			
		Parching with wine	Charred	Parching with soil	Parching with oil
s1	Jinzhong, Zhangxian	j1	t1	tu1	y1
s2	Zhongzhai, Minxian	j2	t2	tu2	y2
s3	Xizhaizhe, Minxian	j3	t3	tu3	y3
s4	Mazichuan, Minxian	j4	t4	tu4	y4
s5	Tianjianhe, Weiyuan	j5	t5	tu5	y5
s6	Wuzhu, Weiyuan	j6	t6	tu6	y6
s7	Huichuan, Weiyuan	j7	t7	tu7	y7
s8	Hetuo, Minxian	j8	t8	tu8	y8
s9	Awu, Tangchang	j9	t9	tu9	y9
s10	Dachaotan, Zhangxian	j10	t10	tu10	y10
s11	Hadapu, Tangchang	j11	t11	tu11	y11
s12	Shichuan, Zhangxian	j12	t12	tu12	y12
s13	Shilidian, Minxian	j13	t13	tu13	y13
s14	Minxian herb market	j14	t14	tu14	y14

280 nm 为检测波长, 在这个波长时其主色谱峰分离效果最好。

当归样品中特征峰的定量 将 66 批当归药材的供试品溶液和对照品溶液按上述的色谱条件进样 10 μL , 通过液相色谱工作站处理后得到 66 个当归不同炮制品 HPLC 图谱和对照品 HPLC 图, 并对其 HPLC 图谱所有峰进行标定, 按下式计算特征峰含量 (样品中无此峰的面积按照 0 计算): 特征峰含量 = (指纹峰峰面积/基准峰峰面积) $\times$ 基准峰含量。

当归及不同炮制品数据预处理 将上述试验所得到的所有数据导入到同一 Excel 表格中, 组成多特征因子构成的多维数据, 但是这些数据中包含噪声、缺失值 (0) 以及少数大峰, 导致特征指标的筛选受到干扰。因此, 在判别函数建立之前对所得数据进行预处理。将 HPLC 数据变量划分为 3 个部分: 大峰, 小峰和中等强度的峰。响应大的离子常常在判断分析中占优势, 从而屏蔽一些确实在各组之间差异明显但响应中等的离子^[10]。响应太小的离子由于分析方法的局限性通常测量误差比较大, 欠缺准确性, 即便这些离子能够经过筛选成为特征指标, 也要经过改进后续的分析方法来确认, 因此在后续判断分析当中只关注中等强度的峰。

当归及不同炮制品数据分析 为研究 HPLC 特征图谱中各色谱峰的 HPLC 数量化特征及药典规定检查的水分、灰分等指标对区分不同炮制品所起的作用, 对 66 个经鉴定明确分类的样品, 采用 SPSS Clementine11.0 数据挖掘软件, 将预处理后的 HPLC 色谱峰数量化特征及水分、灰分、酸不溶性灰分、浸出物作为训练集来建立判别函数, 分析检验样品的属性, 并与鉴定的属性相比较, 以此来验证判别函数的符合率。

当归不同炮制品特征指标的分析 采用高效液相色谱-质谱联用的方法对当归不同炮制品的特征指标进行, 样品处理和色谱条件同上述液相色谱条件。质谱条件: Agilent 1100 液质联用仪, 离子源为 ESI 源, 采用正负离子两种模式, 扫描范围: m/z 50~2 000。扫描干燥温度为 350 $^{\circ}\text{C}$, 雾化气 (N_2), 压力为 240 kPa。气体流速为 8 $\text{L}\cdot\text{min}^{-1}$ 。

结果

1 当归及不同炮制品数据获得及预处理

将当归及不同炮制品水分、总灰分、酸不溶性灰分、浸出物和 HPLC 指纹峰的数据导入到同一 Excel

表格中, 得到 59 个指标, 66 个样本的数据矩阵, 每个样品总峰面积的平均值为 24 866.46, 平均每个样品可以检测到 54 个峰, 所以认为平均峰面积是 460.49 (24 866.46/54)。去掉在 5 组中平均峰面积都大于 4 604.9 的 1 个变量和小于 46.05 的 11 个变量, 得到 66 样品 $\times$ 47 变量的数据矩阵, 进行判别分析。

2 当归及不同炮制品数据分析

将预处理后的 66 样品 $\times$ 47 变量的数据矩阵, 用 SPSS Clementine11.0 进行判别分析, 得到当归及不同炮制品饮片的 Fish's 线性判别函数, 分类函数系数见表 2。用此判别函数进行回代分类, 判别结果大多数与实际相符 (表 3), 66 个样品中有 2 个样品被判断错误, 准确率达到 96.97%, 说明此判别函数可信, 可以对其未知的样本进行分类。

Table 2 Discriminating coefficient of *Angelica* processed products. Constant quantities of j, s, t, tu, and y are -65.098, -36.310, -111.801, -58.273 and -50.565, respectively

Index	Code of processed products				
	j	s	t	tu	y
1	-14.153	3.826	106.234	-24.960	-17.333
10	-22.061	-35.362	-124.081	-10.848	-12.655
12	76.817	47.856	123.403	61.369	52.722
25	-0.631	12.915	86.952	0.649	1.458
31	135.548	165.332	31.591	86.807	147.192
47	18.290	1.870	5.957	4.132	4.893
49	68.734	33.870	-3.816	119.298	86.166
50	188.577	91.286	104.617	250.401	183.566

Table 3 Results of discriminating back substitution for *Angelica* processed products

Code of processed products	Original count	Predicted count	Mis-predicted count	Correct rate/%
j	13	12	1	92.3
s	11	11	0	100
t	14	14	0	100
tu	14	14	0	100
y	14	13	1	92.86

3 当归及不同炮制品特征指标分析

由表 2、表 3 中的判别函数系数和判断回代结果可以看出, 在所选择的 59 个指标 (色谱数据和常规检测) 中提取出 8 个特征指标, 可以区分 96.7% 的样品。这 8 个特征指标分别为峰 1、10、12、25、31、47、49、50, 其色谱峰的保留时间依次为 2.464、6.454、7.427、22.457、31.303、51.661、52.330 和 52.563 min

(图 1)。为进一步分析这些特征成分,采用高效液相色谱-质谱联用的方法对这些特征指标进行定性研究。通过对色谱峰的紫外光谱、一级质谱与文献^[5-9]比较,推测保留时间为 6.454、7.427、22.457、31.303 和 51.661 min 色谱峰分别为香草酸 (vanillic acid)、阿魏酸 (ferulic acid)、阿魏酸松柏酯 (coniferyl-ferulate)、Z-丁烯基苯酞 (Z-butylidenephthalide)、levistolide A (表 4)。其余 3 个色谱峰暂未找到相关报道。

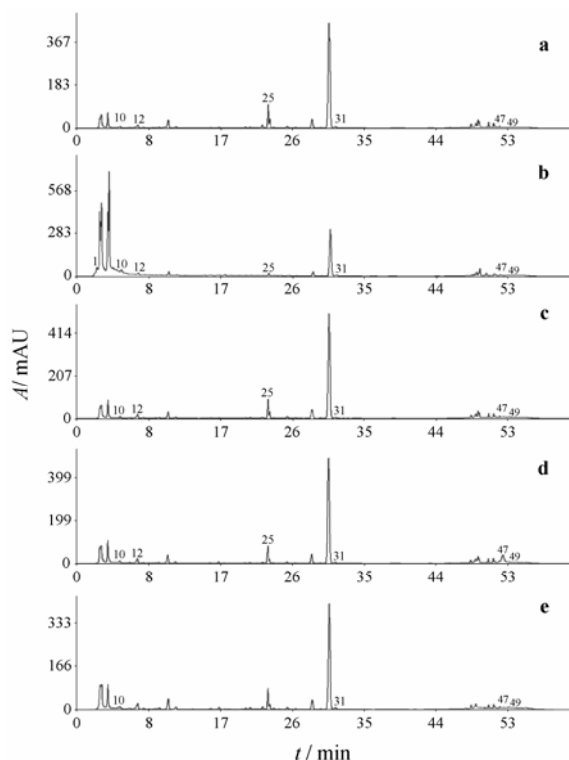


Figure 1 HPLC chromatograms of *Angelica* processed products: unprocessed *Angelica* (a), charred *Angelica* (b), parching *Angelica* with spice oil (c), parching *Angelica* with wine (d), and parching *Angelica* with soil (e). Peak 1, 10, 12, 25, 31, 47 and 49 assignments listed in Table 4

结论

目前数据挖掘和中药特征图谱相结合的模式识别方法是中药质量鉴定的重要方法,也有许多相关报道^[11-14],可实现对中药材及其炮制品的分类评价。本研究在数据挖掘和中药特征图谱结合的基础上,采用判别分析的方法建立了一种高效液相色谱峰数量化特征指标和其他指标水分、灰分、浸出物结合的用来评价当归不同炮制品质量和分类的方法。这样可以更加全面的反映当归不同炮制品的特征,可以全面的认识当归不同炮制品的特点。通过对 66 个当归样品的特征数据进行判别分析,建立了判别函数,用此判别函数回代分类,判别准确率达到 96.97%,表明该模型能够对未知样品进行正确分类,8 个变量经过筛选被确认为潜在的当归不同炮制品分类的标志指标,而这些化合物有望成为解释当归炮制的物质基础及判断当归炮制品质量的标准。其中保留时间为 51.661 的峰在酒当归中明显升高,可视为酒当归的特征成分,而保留时间为 2.464 的峰在当归炭中含量非常高,在其他炮制品中未检出此峰,因此此峰可视为当归炭的特征成分,推测该峰可能是当归在炒炭过程中可水解的鞣质苷键或酯键断裂分解产生的游离酚酸类成分,其具体的结构还有待进一步研究。

炮制品的质量评价是饮片理化特征变化和化学变化特性上的综合反映,在自然条件下很难将化学成分变化和理化特征变化对不同炮制品的影响截然分开。本研究采用化学特征和其他特征相结合的数据分析方法,对当归的不同炮制品进行综合评价,从而减少了一些未知因素的影响,更好地反映了当归不同炮制品的质量特性。

Table 4 Retention time, relative contents and chemical names of feature index for distinguishing *Angelica* processed products

Peak	t_R /min	Relative contents grounded on the peak of ferulic acid					Compound
		Unprocessed	Parching with wine	Parching with oil	Parching with soil	Charred	
1	2.464	0	0	0	0	1.04 ± 0.30	Unknown
10	6.454	0.09 ± 0.03	0.14 ± 0.04	0.11 ± 0.03	0.13 ± 0.06	0.29 ± 0.17	Vanillic acid
12	7.427	0.32 ± 0.06	0.47 ± 0.07	0.38 ± 0.08	0.43 ± 0.12	0.69 ± 0.13	Ferulic acid
25	22.457	0.19 ± 0.05	0.09 ± 0.03	0.09 ± 0.03	0.15 ± 0.05	0.14 ± 0.16	Coniferyl-Ferulate
31	31.303	0.16 ± 0.03	0.14 ± 0.02	0.13 ± 0.03	0.17 ± 0.07	0.22 ± 0.08	Z-Butylidenephthalide
47	51.661	0.01 ± 0.02	1.21 ± 0.58	0.26 ± 0.11	0.22 ± 0.08	0.11 ± 0.01	Levistolide A
49	52.330	0.06 ± 0.01	0.10 ± 0.04	0.14 ± 0.08	0.20 ± 0.05	0.10 ± 0.02	Unknown
50	52.563	0.09 ± 0.01	0.17 ± 0.04	0.16 ± 0.05	0.19 ± 0.02	0.18 ± 0.04	Unknown

References

- [1] Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China (中华人民共和国药典) [S]. 2005 ed. Part I. Beijing: Chemical Industry Press, 2005: 89.
- [2] Xiao XH, Xia WJ, Chen SY. Abstracts of original articles a review on the research of genuine traditional Chinese drugs [J]. China J Chin Mater Med (中国中药杂志), 1995, 20: 323.
- [3] Gansu Province Health Department. Gansu Processing Standard of Traditional Chinese Medicine (甘肃中药炮制规范) [M]. Lanzhou: Gansu People Publishers, 1999: 53.
- [4] Cindy QH, Andy LW, Silberman J, et al. Label-free semiquantitative peptide feature profiling of human breast cancer and breast disease sera via two-dimensional liquid chromatography-mass spectrometry [J]. Mol Cell Proteomics, 2006, 5: 1095–1104.
- [5] Lu GH, Kelvin C, Liang YZ, et al. Development of high performance liquid chromatographic fingerprints for distinguishing Chinese *Angelica* from related Umbelliferae herbs [J]. J Chromatogr A, 2005, 1073: 383–392.
- [6] Yi T, Zhang H, Xie J, et al. A new procedure for the preparative separation and isolation of Z- ligustilide from the roots of *Angelica sinensis* [J]. J Sep Sci, 2007, 30: 1973–1978.
- [7] Yi T, Kelvin L, Lu GH, et al. Comparative analysis of *Ligusticum chuangxiong* and related umbelliferous medicinal plants by high performance liquid chromatography electrospray ionization mass spectrometry [J]. Planta Med, 2007, 73: 392–398.
- [8] Wu YY, Shang MY, Cai SQ. HPLC fingerprint of the components of *Radix Angelicae Sinensis* [J]. Acta Pharm Sin (药学报), 2008, 43: 728–732.
- [9] Yang F, Xiao YS, Zhang FF. High performance liquid chromatography-mass spectrometry analysis of *Radix Angelica Sinensis* [J]. Acta Pharm Sin (药学报), 2006, 41: 1078–1083.
- [10] Yin PY, Zhao XJ, Liq RW, et al. Metabonomics study of intestinal fistulas based on ultraperformance liquid chromatography couple with Q-TOF mass spectrometry [J]. J Proteome Res, 2006, 5: 2135–2143.
- [11] Xu YQ, Sun SQ, Zhou Q, et al. Recognition of three classes of skullcaps by FTIR spectroscopy combined with artificial neural networks [J]. Spectrosc Spectral Anal (光谱学与光谱分析), 2002, 22: 945–948.
- [12] Zhang L, Lan YW, Han Y, et al. Classification and recognition of *Tripterygium wilfordii* and *T. hypoglaucom* by artificial neural networks [J]. Acta Pharm Sin (药学报), 1995, 30: 127–132.
- [13] Cai YD, Gong JW, Cheng ZN, et al. Artificial neural network method for quality estimation of traditional Chinese medicine [J]. Chin Tradit Herb Drugs (中草药), 1994, 25: 187–189.
- [14] Cai YD, Gong JW, Can JR, et al. Quality assessment of the Chinese traditional drug Weilingxian by artificial neural network [J]. China J Chin Mater Med (中国中药杂志), 1993, 18: 518.