色谱指纹图谱相似度方法的适应性研究

詹雪艳, 林兆洲, 段天璇, 李磊, 乔延江

北京中医药大学, 北京 100102

摘要：目的　评价 8 种相似度方法,研究各相似度方法的适应性,为相似度方法的应用提供指导。方法　以土茯苓药材色谱指纹图谱峰面积数据为载体,以峰数弹性、峰比例同态性和峰面积同态性为指标,比较了 8 种相似度方法,分析多种相似度方法在药材真伪鉴别、同一品种药材的分类评价和中药制剂质量稳定性评价 3 个中药质量评价层面中的适应性。结果　数量相似度仅对小峰增减响应灵敏,可用于药材的品种鉴别；夹角余弦、相关系数和组合相似度仅对峰比例差异响应灵敏,适合药材品种鉴别和原料药峰比例稳定性的初步评价；基于相似系统理论的多种程度相似度对峰比例差异和峰面积差异同时响应,能定量地衡量色谱峰峰面积的差异和对应成分配比的稳定性,适合中药制剂质量稳定性的精细评价。结论　数量相似度可用于中药材品种的初步鉴别,共有峰夹角余弦和相关系数相似度适合药材的品种鉴别,所有峰夹角余弦和组合相似度适合原料药峰比例稳定性的初步评价,基于相似系统理论的多种程度相似度适用于中药制剂质量稳定性的精细评价。

关键词：色谱指纹图谱；相似度方法；适应性；质量评价

DOI：10.3969/j.issn.1005-5304.2012.05.024

中图分类号：R284.1　　文献标识码：A　　文章编号：1005-5304(2012)05-0061-04

Applicability of Similarity Measures of Chromatographic Fingerprints

ZHAN Xue-yan, LIN Zhao-zhou, DUAN Tian-xuan, LI Lei, QIAO Yan-jiang *(Beijing University of TCM, Beijing 100102, China)*

Abstract：**Objective** Comparation of eight similarity measures was studied to provide some advice for applicability of these similarity measures. **Methods** Eight similarity measures were evaluated according to three indexes (elasticity of peak number, homostasis of proportion among peaks and homostasis of peak area) based on peaks area data of chromatographic fingerprints of *Smilax glabra* Roxb.. The applicability of similarity measures to the variety identification of Chinese crude drug and the quality stability evaluation of the intermediate and the product during manufacture of TCM were analyzed. **Results** The numerical similarity (i.e. Tanimoto, Peak overlap ratio) could only respond to increase or decrease of peak number, and could be used for species identification of Chinese herbal medicines. The similarity based on similarity calculation (i.e. cosine of the angle, correction coefficient and combination similarity) could only respond to the fluctuation of peak ratio, and be applicable for species identification of Chinese herbal medicines and stability evaluation of peak ratio of Chinese crude drug. Several extent similarity based on Similarity System Theory could respond to the variation of peak ratio and peak area, and quantitatively measure the difference of peak area and the stability of corresponding ingredients, which could be applicable for fine evaluation of the quality stability of Chinese drugs preparation. **Conclusion** The numerical similarity could be used for preliminary identification of Chinese herbal medicines. Cosine of the angle and correction coefficient calculating common peaks area were applicable for the variety identification of Chinese crude drug, and cosine of the angle calculating all peaks area and combination similarity were applied to the quality stability evaluation of the crude drug. Extent similarity based on Similarity System Theory was applicable for fine evaluation of the quality stability of Chinese drugs preparation.

Key words：chromatographic fingerprints；similarity measures；applicability；quality evaluation

色谱指纹图谱相似度用于中药质量稳定性的评价, 简单易行, 易于推广, 是中药质量控制的重要手段。色谱指纹图谱相似度可用于中药材的品种鉴别及中药生产中原料、中间体和产品的质量控制, 不同的质量控制环节应选择合适的相似度方法以及针对不同的中药对象选用合适的相似度方法是中药质量评价中的一个关键问题。目前应用最广泛的色谱指纹图谱相似度方法主要有 Tanimoto、峰重叠率等数量相似度[1], 向量夹角余弦[2]、相关系数[3-4]、组合相似度[5]和距离相似度, 以及基于相似系统理论的多种程度相似度等[6-8]3 类方法。但是, 对相似度

基金项目：“重大新药创制”国家科技重大专项(2010ZX09502-002)；北京中医药大学自主课题(2009JYB22-JS030)；北京中医药大学科研创新团队基金(2011-CXTD-11)

通讯作者：乔延江,E-mail：yjqiao@263.net

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

方法的比较仅局限于前2类,对以上3类相似度计算方法的比较和适应性研究很少。程氏等[9]比较了欧氏距离、绝对距离、Minkowsky 距离、夹角余弦、相关系数和峰匹配度 6 种相似度方法,王氏等[10]对比了欧氏距离、夹角余弦、相关系数、指数相似系数和相似性比 5 种相似度方法,得到一致的结果,表明峰匹配度的峰数弹性值最大,能灵敏地检测小峰的增减,夹角余弦和相关系数只在峰比例同态性上有较高计算值,两者能灵敏地反映峰比例波动,不能检测峰面积的线性波动;欧氏距离对峰比例和峰面积均能灵敏地反映。聂氏[11]、孟氏等[12]对数量相似度和定性相似度两类相似度方法的评价进一步验证了以上结论。本研究以峰数弹性、峰比例同态性和峰面积同态性为指标[9]评价数量相似度、定性相似度和定量相似度 3 类相似度方法,总结各种相似度方法在药材真伪鉴别、同一品种药材的分类评价和中药制剂质量稳定性评价 3 个不同的中药质量评价层面中的适应性,以期为相似度方法的应用提供指导。

1 相似度评价指标原理

色谱指纹图谱的多种相似度方法采用峰数弹性、峰比例同态性、峰面积同态性 3 个指标来评价。

峰数弹性衡量某相似度方法对指纹图谱间色谱峰个数波动的检测能力,定义为相似度值变化的相对量与指纹图谱间谱峰个数差异的相对量之比,如式 $E=(\triangle S/S)/(\triangle N/N)$。其中 E 为峰数弹性值,$\triangle S$、$\triangle N$分别为待测指纹图谱中色谱峰增减引起相似度值变化量和色谱峰增减的数量,S 为色谱峰增减前两指纹图谱间的相似度,N为待测指纹图谱的谱峰数。通常,E值取决于小峰增减所得到的计算值,峰数弹性可视为某相似度方法对小峰的检测能力,E 值越大,表明该相似度对小峰增减响应越灵敏。

峰比例同态性衡量某相似度方法对两指纹图谱峰比例差异的敏感程度,定义为多个样本间的峰比例差异$\triangle P(\triangle P1, \triangle P2 \cdots\cdots, \triangle Pj)$与其对应的相似度值变化量$\triangle S(\triangle S1, \triangle S2 \cdots\cdots, \triangle Sj)$两列数据间的 pearson 相关系数。峰比例同态性越大,表明该相似度方法对指纹图谱间峰比例的变化响应越灵敏。

峰面积同态性衡量某相似度方法在待测样本色谱峰个数和峰比例不变时对指纹图谱中总峰面积线性波动的检测能力,定义为待测样本的色谱峰数目和峰比例不变的前提下,样本的总峰面积变化量$\triangle Aj$ 与相似度值变化量$\triangle Sj$ 两列数据间的 pearson 相关系数。峰面积同态性越大,表明该相似度方法对样本总峰面积的线性波动响应越灵敏。

综上所述,峰数弹性、峰比例同态性和峰面积同态性 3 个指标的计算值反映了相似度方法检测色谱指纹图谱色谱峰个数的差异、色谱峰峰比例的差异以及总峰面积差异的能力。

2 8 种相似度方法评价

2.1 数据来源

采用 9 个土茯苓样本色谱指纹图谱的共有峰峰面积数据计算 8 种相似度方法(Tanimoto、峰重叠率、夹角余弦、相关系数、欧式距离、程度相似度、改良程度相似度、新改良程度相似度)的峰数弹性、峰比例同态性和峰面积同态性的 3 个评价指标值。土茯苓样本来源和数据采集方法参见文献[5]。

2.2 数据处理及软件

将 9 个土茯苓样本的色谱指纹图谱导入国家药典委员会《中药色谱指纹图谱相似度评价系统》2004A 版,时间窗宽度设为 0.05,得到 9 个土茯苓样本的共有峰峰面积数据和中位数对照指纹图谱峰面积数据。以中位数对照指纹图谱为参照,计算 8 种相似度方法的 3 个评价指标值。相似度方法的评价指标值采用 Excel 软件计算。

2.3 结果与分析

2.3.1 9 批土茯苓提取物色谱指纹图谱峰面积数据(见表 1)

表 1 9 批土茯苓样品提取物色谱指纹图谱的峰面积

s10yl	s7hy	s8cs	s9cd	S6wy	s4gz	s5gz	s2hf	s1hf
227.97	296.93	219.49	270.91	259.56	204.03	245.93	247.57	291.63
1 559.64	107.18	1 455.42	797.89	1 516.74	162.00	322.47	295.34	150.08
375.09	145.15	1 392.78	428.03	364.14	420.08	146.62	221.10	107.47
11 423.79	1 829.24	12 721.49	7 884.62	7 086.19	4 761.68	3 910.48	4 047.00	2 501.66
139.38	169.96	782.22	253.19	713.90	320.12	93.19	172.62	167.02
1 434.42	394.10	2 710.82	1 165.06	1 133.61	973.46	516.65	702.14	364.44
396.52	147.21	104.49	253.28	241.67	66.00	172.01	220.38	101.34
266.24	97.58	647.22	559.93	217.09	130.45	148.91	142.22	104.69
2 730.85	839.97	1 001.17	1 817.49	1 239.46	1 124.80	1 841.55	1 974.31	1 566.84
166.94	164.03	123.55	209.18	106.38	152.14	258.06	190.71	212.90

2.3.2 相似度方法 3 个评价指标的比较 以表 1 中土茯苓色谱指纹图谱峰面积数据为载体计算了 8 种相似度方法的峰数弹性、峰比例同态性和峰面积同态性的 3 个评价指标的计算值,结果见表 2。

表 2 显示,数量相似度中峰重叠率和 Tanimoto 相似度仅具有较大的峰数弹性值,基于土茯苓样本峰面积数据计算的峰数弹性值分别为 0.526 和 1.000,表明数量相似度对色谱峰的检测能力强,能灵敏地检测指纹图谱中小峰的增减。夹角余弦和相关系数相似度具有较高的峰比例同态性值,仅对峰比例变化响应灵敏。组合相似度是在向量夹角余弦的基础上将非共有峰

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

对相似度值的影响纳入计算,以峰比例同态性值最大为目标建立的相似度模型,因此组合相似度与向量夹角余弦相似,仅对峰比例变化响应灵敏。

峰面积同态性是在前 2 个指标不变的情况下,即图谱的峰数和峰比例一定时,相似度对总峰面积变化响应的灵敏程度,表征在峰数和峰比例不变时对成分总量线性波动的响应。表 2 表明,数量相似度、夹角余弦和相关系数定性相似度的峰面积同态性计算值为 0,对这个指标均不响应,表明在峰数和峰比例不变时总峰面积增大或减小,以上相似度的计算值不变。这与文献[9-12]对以上相似度方法的评价结果是一致的。欧氏距离和基于相似系统理论的 3 种程度相似度都有很高的峰面积同态性值,说明以上 4 种相似度对峰面积的线性波动响应灵敏。此外,欧氏距离还具有较高的峰比例同态性计算值,说明其对峰比例和峰面积的变化均能响应。基于相似系统理论的 3 种程度相似度在 3 个评价指标上均有较高的计算值,说明峰数、峰比例的变化或总峰面积的线性波动都能引起 3 种相似度计算值的变化。其中,3 种程度相似度的峰面积同态性值最高,说明该 3 种相似度对总峰面积的线性波动响应更灵敏。

表 2　8 种相似度方法的峰数弹性、峰比例同态性及峰面积同态性比较

相似度方法	峰数弹性	峰比例同态性	峰面积同态性
Tanimoto	1.000	0.000	0.000
峰重叠率	0.526	0.000	0.000
夹角余弦	0.002	0.940	0.000
相关系数	0.000	0.897	0.000
欧式距离	0.002	0.695	0.987
程度相似度	0.503	0.694	0.936
改良程度相似度	0.532	0.286	0.956
新改良程度相似度	0.492	0.529	0.961

2.3.3　相似度方法的适应性　色谱指纹图谱相似度可用于中药材的品种鉴别,中药生产过程中的原料药、中间体和成品质量的稳定性评价,不同的质量控制环节应选择合适的相似度方法。根据上述 8 种相似度方法的评价结果,本研究初步提出了选择合适相似度方法的依据,见图 1。

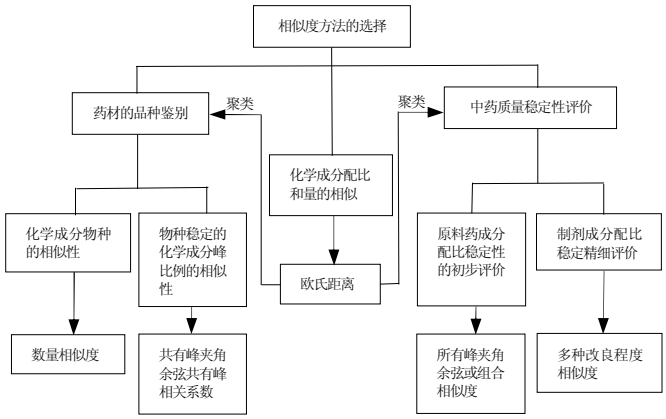

图 1　中药色谱指纹图谱相似度方法的选择

中药材内在的化学成分是次生代谢产物,受生长环境和生长年限的影响而产生个体间的差异,但同时植物的代谢具有遗传性,同一物种的个体内在化学成分和成分间配比具有相似性。药材样本通过与对照药材比较内在化学成分物种上或物种稳定的化学成分间配比的相似性可进行品种鉴别。对色谱峰增减响应灵敏的数量相似度能表征两色谱指纹图谱所对应化学成分物种上的相似性,可用于药材品种的初步鉴别。两色谱指纹图谱的共有峰代表了药材中物种稳定的化学成分,仅对峰比例变化响应灵敏的夹角余弦和相关系数相似度能表征化学成分间比例的相似性,因此,常采用色谱指纹图谱的共有峰夹角余弦或相关系数相似度衡量药材中物种稳定的化学成分间比例的相似性和药材的品种鉴别。

中药制剂过程中常采用原料药、中间体和成品的色谱指纹图谱来评价中药质量的稳定均一。对于制剂的原料药,需初步控制其内在成分的配比,以保证后续生产过程中中间体的化学成分间配比具有一定的稳定性。如果中药提取物偏离所要求的固定配比很远,通过多种中药提取物的勾兑也难以实现后续中间体的化学成分间配比符合固定配比。因此,原料药质量控制的关键是内在成分间的配比,其色谱指纹图谱相似度应选择能灵敏反映峰比例差异的相似度方法。由相似度评价结果知,欧氏距离对峰比例的变化和总峰面积的线性变化同时响应,其相似度值不仅仅只反映样本间峰比例的差异,是峰比例和峰面积差异的共同反映。基于相似系统理论的 3 种定量相似度只能衡量峰面积差异小于 100%的色谱指纹图谱的相似度,不能衡量峰面积差异大的原料药的色谱指纹图谱的相似度。此外,原料药质量均一性控制中还应控制其杂质的量,对相应的非共有峰进行控制。因此,选择所有峰夹角余弦或组合相似度衡量两图谱峰比例和对应的化学成分间比例的一致性,初步控制原料药化学成分间配比的稳定性,从而保证后续生产中中间体和产品质量的稳定。

通过混合均一化操作后,不同批次的中间体内在化学成分的配比都在某一固定配比的上下波动不大,此时维持该固定配比需要对化学成分间含量进行精细反映和控制,夹角余弦和相关系数两相似度方法对色谱指纹图谱峰面积数据间差异响应不够灵敏[13-16],不能精确地反映成分间配比的差异。基于相似系统理论的 3 种程度相似度能定量地评价峰面积差异相差不大的两色谱指纹图谱的相似程度,灵敏地反映各化学成分含量的差异,通过控制改良程度相似度值的大小,可以控制中药中对应成分含量的稳定性,适合中药中间体和产品质量稳定性的精细评价和控制。

欧氏距离衡量中药色谱指纹图谱相似度时对色谱峰峰比例的差异和总峰面积的差异同时响应,缺乏专属性,计算值受仪器重现性和响应值差异的影响较大,稳定性较差。因此,欧氏距离常通过模式识别的聚类和分类来鉴别中药材或初步评价中药质量稳定性。

3　结论

色谱指纹图谱相似度可用于中药材的品种鉴别及中药生产中原料、中间体、产品的质量控制,不同的质量控制环节应选择合适的相似度方法。本研究对 8 种相似度方法进行了比较,分析了 3 类相似度方法的适应性。数量相似度可用于中药材品种的初步鉴别,共有峰夹角余弦和相关系数相似度适合药材的

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

品种鉴别,所有峰夹角余弦和组合相似度适合原料药峰比例稳定性的初步评价,基于相似系统理论的多种程度相似度适用于中药制剂质量稳定性的精细评价。通过对中药色谱指纹图谱多种相似度方法的评价和适应性研究,对相似度方法在中药质量控制中的应用提供指导。

参考文献:

[1] 洪筱坤,王智华,郭济贤,等.柴胡属19种植物挥发油的气相色谱-相对保留值的指纹分析[J].药学学报,1988,23(11):41-47.

[2] 王龙星,肖红斌,梁鑫淼,等.一种评价中药色谱指纹图谱相似度的新方法:向量夹角法[J].药学学报,2002,37(9):713-717.

[3] Liu Yongsuo, Meng Qinghua, Chen Rong, et al. Improvement of similarity measure : pearson product-moment correlation coefficient[J]. Journal of Chinese Pharmaceutical Sciences, 2004,13(3):180-186.

[4] Liu Li, Lu Yun, Shao Qing, et al. Binary chromatographic fingerprint for quality evaluation of Radix Ophiopogonis by high-performance liquid chromatography coupled with ultraviolet and evaporative light-scattering detectors[J]. J Sep Sci,2007, 30:2628-2637.

[5] 詹雪艳,史新元,段天璇,等.色谱指纹图谱组合相似度的算法[J].色谱,2010,28(11):1071-1076.

[6] 刘永锁,孟庆华,蒋淑敏,等.相似系统理论用于中药色谱指纹图谱的相似度评价[J].色谱,2005,23(2):158-163.

[7] 刘永锁,曹敏,王义明,等.相似系统理论定量评价中药材色谱指纹图谱的相似度[J].分析化学,2006,34(3):333-337.

[8] 詹雪艳,史新元,展晓日,等.基于相似系统理论的相似度计算方法的改进[J].分析化学,2010,38(2):253-257.

[9] 程翼宇,陈闽军,吴永江.化学指纹图谱的相似性测度及其评价方法[J].化学学报,2002,60(11):2017-2021.

[10] 王玺,王文宇,张克荣,等.中药HPLC指纹图谱相似性研究的探讨[J].沈阳药科大学学报,2003,20(5):360-362,366.

[11] 聂磊,曹进,罗国安,等.中药指纹图谱相似度评价方法的比较[J].中成药,2005,27(3):249-252.

[12] 孟庆华,刘永锁,蒋淑敏,等.色谱指纹图谱综合信息指数在中药质量控制中的应用研究[J].中国天然药物,2004,2(6):359-364.

[13] Rundle HD, Jackson DA. Spatial and temporal variation in littoral-zone fish communities:a new statistical approach[J]. Can J Fish Aquat Sci,1996,53:2167-2176.

[14] Kendrick BS, Dong A, Alison SD, et al. Quantitation of the area of overlap between second-derivative amide I infrared spectra to determine the structural similarity of a protein in different states[J]. J Pharm Sci,1996,85(2):155-158.

[15] Rull V. A palynological record of a secondary succession after fire in the Gran Sabana, Venezuela[J]. Journal of Quaternary Science,1999,14(2):137-152.

[16] Torriani S, Zapparoli G, Dellaglio F. Use of PCR-based methods for rapid differentiation of Lactobacillus delbrueckii subsp. bulgaricus and L. delbrueckii subsp. Lactis[J]. Appl Eviron Microbiol,1999,65(10):4351-4356.

(收稿日期:2011-11-06,编辑:华强)

中华中医药学会发布五年规划

2012年4月14日,在广东省佛山市召开的中华中医药学会第五届常务理事会第四次会议上,中华中医药学会印发五年规划并部署落实。规划内容涉及中医药学术发展、人才培养、继续教育、科学普及、文化建设、国际交流、组织建设等各个方面,对推动学会工作科学发展、全面提升学会工作水平具有十分重要的意义。卫生部副部长、国家中医药管理局局长、中华中医药学会会长王国强,中华中医药学会副会长兼秘书长李俊德等领导出席。

规划在任务与措施中提出,坚持学术特色,发挥中医优势,继承学术特色,加快传承创新重点项目,5年内,建立10～20个中医药学术流派代表医家研究工作室,系统整理、发掘、总结各学术流派思想,培养中医药学术流派传人;全面开展以“读经典、跟名师、做临床”为主题的读书活动;5年内,依托学会现有网站,建成覆盖广泛的远程网络继续教育平台;5年内,开展300场次的“百名中医科普专家讲科普”活动,推出100名中华中医药科普专家;围绕国际重大医疗卫生事件及重大中医药前沿学术问题,组织中医药国际学术交流会,提高中医药学术的国际影响力;充分利用学会的组织优势,投身到国家创新体系建设中,通过自身力量或寻求合作,选择10家中医医院、10家中医药企业,采取提供技术支持的方法,开展合作共建,加强厂会合作,促进企业的创新体系建设。

王国强对学会2012年的工作提出4点建议:一是认清形势,明确任务,切实增强做好学会工作的责任感使命感。从深化经济体制改革、社会管理体制改革、科技体制改革、文化体制改革等方面入手。二是突出重点,紧紧围绕发挥中医药特色优势,大力推进中医药学术发展。切实把加快推进中医药学术发展和科技进步作为学会创新发展的战略重点。创新大思路,始终把握中医药的原创优势,在建设创新型国家和解决好人民群众健康的突出问题中发挥应有的作用;凝炼大项目,凝炼出有利于促进经济社会发展、解决制约中医药发展的科技大项目;实现大联合,通过机制创新构建上下联动、内外结合的科研模式;搭建大平台,搭建有利于学科交融、成果共享的可持续发展的中医药科技大平台;取得大成果,在推动中医药学术进步、改善民生、促进健康、壮大产业、促进经济社会发展方面取得标志性成果;开拓大市场,积极促进产学研合作,把科研成果转化为惠及百姓的产品和服务,开拓包括中医医疗、预防保健、文化传播等在内的大市场。三是统筹协调,真抓实干,落实好各项重点工作任务。四是内增素质,外树形象,切实抓好学会自身建设。

(本刊通讯员)

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net