

基于模式识别的唇形科中药寒热药性物质基础研究

张传耀¹, 周 晴², 王 校¹, 卜菲菲¹, 王 鹏^{3*}

(1. 安徽中医药大学中医学院, 安徽 合肥 230012; 2. 安徽中医药大学医药信息工程学院, 安徽 合肥 230012;
3. 安徽中医药大学研究生院, 安徽 合肥 230012)

摘要: 基于模式识别算法对唇形科中药寒热药性的物质基础进行研究, 并建立寒热药性识别模型。首先, 建立唇形科中药物质成分数据库。采用物质成分三级分类体系, 利用频数分析、关联规则分析、逻辑回归、特征选择等数据挖掘方法得到唇形科中药寒热药性的物质基础, 并建立识别模型。卡方检验结果显示, 在一级、二级、三级分类水平上, 寒热药性的物质成分构成比均有显著差异 ($P < 0.05$), 差异随物质分类层次而变化, 各级分类下平均变异系数分别为 42.30%、79.07%、91.51%, 占比差异一级分类 < 二级分类 < 三级分类。关联规则分析结果显示, 在三级分类条件下, 有效的关联规则较多。进一步分析得到 27 个核心类群及 34 个特异类群, 特征选择结果得到 15 个决定性类群。采用多元逻辑回归分析成功建立了寒热药性识别模型, 总体准确度可达到 89%。唇形科中药寒热药性的物质基础差异明显, 且互有交叉; 双环二萜等 27 种物质为寒热药性的核心类群, 其中, 15 种物质为决定性类群; 唇形科中药寒热药性往往由多种类别的物质成分交互作用而体现, 单一类别的物质常不能用来表征寒热药性, 多种物质类群的有机组合是寒热药性的物质基础。

关键词: 模式识别; 唇形科; 中药; 寒热药性; 物质基础; 识别模型

中图分类号: R917 文献标识码: A 文章编号: 0513-4870(2023)02-0429-10

Study on the material basis of cold and hot properties of traditional Chinese medicines in Lamiaceae based on pattern recognition

ZHANG Chuan-yao¹, ZHOU Qing², WANG Xiao¹, BU Fei-fei¹, WANG Peng^{3*}

(1. School of Traditional Chinese Medicine, Anhui University of Chinese Medicine, Hefei 230012, China; 2. School of Medical Information Engineering, Anhui University of Chinese Medicine, Hefei 230012, China; 3. Graduate School, Anhui University of Chinese Medicine, Hefei 230012, China)

Abstract: To study the material basis of cold and hot properties of traditional Chinese medicines (TCMs) in Lamiaceae and to establish a cold and hot properties identification model, a database of material components of TCMs in Lamiaceae was established. A three-level classification system of material components was used to obtain the material basis of cold and hot properties of the Lamiaceae family by using data mining methods such as frequency analysis, association rule analysis, logistic regression, and feature selection. Several identification models were established to recognize the cold and hot properties. The chi-square test results showed that the material composition ratios of cold and hot properties were significantly different at the first-level, second-level, and third-level classification ($P < 0.05$), and the differences varied as the levels of substance classification changed. The average coefficients of variation were 42.30%, 79.07%, and 91.51% at the first-level, second-level, and third-level classification levels, respectively. In other words, in terms of the percentage differences in material composition ratio, the first-level was smaller than the second-level, and the second-level was smaller than the third-

收稿日期: 2022-10-10; 修回日期: 2022-12-27.

基金项目: 国家自然科学基金项目 (81673622); 安徽省自然科学基金课题 (1508085MH202); 安徽省高校自然科学研究重点课题 (KJ2019A0464).

*通讯作者 Tel: 86-551-68129333, E-mail: anhuiwangpeng@126.com

DOI: 10.16438/j.0513-4870.2022-1100

level. The results of the association rule analysis showed that under the third-level classification, there were many effective association rules, and 27 core groups and 34 specific groups of chemical components were obtained based on these rules. 15 decisive groups were obtained from the feature selection results. Multinomial logistic regression analysis was used to successfully establish a cold and hot properties identification model with an overall accuracy of 89%. The material basis of cold and hot properties of TCMs in Lamiaceae is different and intersect with each other. Twenty-seven groups of chemical components, such as bicyclic diterpenes, are the core groups of cold and hot properties, of which 15 groups are the decisive groups. The cold and hot properties are often characterized by the interaction of multiple classes of substances, and a single class of substances often cannot be used to characterize the properties. The organic combination of multiple classes of substances is the material basis of cold and hot properties.

Key words: pattern recognition; Lamiaceae; traditional Chinese medicine; cold and hot property; material basis; identification model

中药药性理论是中药理论体系的基础,是中药区别于其他种类药物的突出标志。其内容十分丰富,可包括四气、五味、归经、升降浮沉、有毒无毒、用药禁忌、配伍规律等。中药药性理论的核心是中药四性理论,即《神农本草经》所指寒、热、温、凉四气^[1-3]。《素问·至真要大论》云“寒者热之、热者寒之”,《神农本草经》指出“疗寒以热药,疗热以寒药”。在此基本原则下,中药药性理论尽管包括四气、五味、归经、升降沉浮、毒性等内容,但寒、热二性是其根本。临床使用中,是否能够准确把握药性之寒热是能否取得疗效的关键。即《本草经集注》云“其(指药性)甘苦之味可略,有毒无毒易知,惟冷热须明”。近年来,学者不断从文献研究、思路研究、方法学研究等方面对中药药性理论进行了深入探讨。肖小河和王永炎共同提出药性热力学观^[4],匡海学提出“中药性味可拆分可组合”假说^[5-7],王振国团队提出中药四性“性-效-物质三元论”假说,并进一步发展成中药“性-构关系”研究^[8-11]。其他相关研究以植物亲缘学^[12,13]、网络药理学^[14,15]、生物信息学^[16]、细胞学^[17]、基因组学^[18]、蛋白质组学^[19,20]、代谢组学^[21,22]等为研究手段,从中药药性的属性表征、物质基础、内在作用机制等方面研究,为中药药性理论的现代诠释提供了可借鉴的方法^[23,24]。

药性研究的思路和方法虽多,药性与物质成分关系研究一直是研究热点^[25-27]。中药是复杂的分子巨系统,中药药性是其分子药性的有机集成。探讨中药药性的物质属性,应该回归中医经典理论的观察认知思路,从整体上、宏观上把握药性的本质^[28,29]。基于科属分类,立足于全成分构成特征,分类探讨中药药性与物质成分的相互关系,既能避免从单一或几种成分入手研究的局限,亦可很好地切合中药药性的整体认知模式,将成为中药药性物质基础研究领域科学可行的有效方法。

药用植物亲缘学证实药用植物的亲缘关系-物质成分-药性效应之间具有高度相关性^[12,30]。唇形科植物入药历史悠久,种类丰富、品种多样,是药用植物的重要组成部分,《神农本草经》既已记载了泽兰、荆芥、夏枯草、黄芩等中药。作者课题组^[31-35]在植物亲缘学基础上,借鉴药用植物科属分类,对唇形科中药寒热药性进行了相关研究,并取得了一定研究基础,初步证实在科属分类的视角下,唇形科中药物质成分与其寒热药性有较强的对应关系。

在此基础上,本研究继续以唇形科中药为示例,选取《中华人民共和国药典》《中华本草》所载全部唇形科中药,通过系统全面的文献整理,建立唇形科中药物质成分数据库。对物质成分进行三级分类,采用频数分析、关联规则分析、逻辑回归、特征选择等数据挖掘方法得到唇形科中药寒热药性物质基础,并建立识别模型。结果表明,在三级分类条件下,寒热药性的物质基础差异明显,关联规则分析得到27个核心类群及34个特异类群,特征选择结果得到15个决定性类群;成功建立寒热药性识别的多元逻辑回归模型,总体准确度89%。进一步分析发现,唇形科中药寒热药性往往由多种类别的物质成分交互作用而体现,单一类别的物质常不能用来表征寒热药性,多种物质的有机组合是寒热药性的物质基础。

材料与方法

建立唇形科中药物质成分数据库 选取《中华人民共和国药典》(2020版)、《中华本草》所载唇形科中药,共245味。物质成分来源于集成性药学著作和各主要文献数据库,其中集成性药学著作有《中华人民共和国药典》《中华本草》《中药大辞典》《全国中草药汇编》(第三版)、《新编中药志》《中药活性成分分析手册》《常用中药及其活性成分手册》《中药现代研究与应用》

《中药材活性成分化学结构图集》《中国植物志》。主要文献数据库包括中国生物医学文献数据库、中国知网、万方数据知识服务平台、维普中文期刊服务平台、PubMed (<https://pubmed.ncbi.nlm.nih.gov>)、Web of Science (www.webofscience.com)。

文献选择纳入条件: 涉及到该味药物的物质成分研究的文献; 明确基原植物、入药部位、实验方法、研究方法的文献。剔除条件: 侧重论述提取药物成分的技术, 对药物物质成分描述简略的文献; 描述药物的鉴别、成分含量测定、质量控制等方法的文献; 描述药物不同入药部位物质成分差异的文献; 非药用部位物质成分研究的文献; 研究采收时间、炮制、煎煮、制备等与物质成分关系的文献; 研究药物在复方中含量测定的文献; 研究药物物质成分的生物活性、药理作用、药效等相关文献; 研究药物饮片相关物质成分的文献。

此外, 有些中药的物质成分来源于物质成分相关数据库, 如中国科学院上海有机所化学专业数据库、中药与化学成分数据库、中药系统药理学数据库与分析平台 (TCMSP, <https://old.tcm-sp-e.com/tcm-sp.php>)、BATMAN-TCM (a Bioinformatics Analysis Tool for Molecular Mechanism of Traditional Chinese Medicine)、本草组鉴 (HERB, <http://herb.ac.cn>) 等。

药性处理 中药的四气药性一般指寒、热、温、凉。本草文献中, 对药性的记录, 又多有大寒、微寒、大热、大温、微温等描述, 以区分寒热程度。实际上此类描述仍属于寒热的范畴。因此, 为方便整理统计, 将大寒、寒、微寒、凉统一归为“寒”, 将大热、热、温、微温统一

归为“热”。对于平性药不做变动, 删除寒热不明的中药。

化学物质的确认及信息采集 为保证每个化学物质的唯一性, 每一个物质成分选取中文化学名、英文化学名、分子式、分子量、IUPAC (International Union of Pure and Applied Chemistry) name、SMILES (Simplified Molecular Input Line Entry System)、InChI (International Chemical Identifier)、InChIKey (International Chemical Identifier Key)、结构式、一级分类、二级分类、三级分类等 12 个特征参数, 删除因中英文名称不同而重复的物质成分。化学物质的确认及信息采集依据下列公认的化学专业数据库: 中国科学院上海有机化学研究所化学专业数据库、中国医学科学院药物研究所中国天然产物化学成分库、美国国家健康研究院的有机小分子生物活性数据库 (PubChem, <https://pubchem.ncbi.nlm.nih.gov>)、爱思唯尔 (Elsevier) 公司的 Reaxys 数据库 (www.reaxys.com)。其他商业化学商品数据库作为补充。

建立物质成分分类体系 依据《中药化学》和《天然药物化学》, 将物质成分分为醌类、苯丙素类、黄酮类、萜类、甾体类、生物碱类、鞣质类、脑苷类、糖和苷类、脂肪酸类、氨基酸和蛋白质类、其他类等 12 个大类, 作为一级分类。在一级分类基础上, 依据化合物分子结构差异, 结合物质成分生源途径进行二次分类, 得到二级分类。在二级分类基础上, 进一步依据化合物分子结构特征进行三次分类, 得到三级分类。具体分类体系见表 1。

Table 1 Three-level classification system of material components

First-level	Second-level	Third-level
Quinones	Benzoquinones	<i>p</i> -Benzoquinone; <i>o</i> -benzoquinone
	Naphthoquinones	1,2-Naphthoquinones; 1,4-naphthoquinones; 2,6-naphthoquinones
	Phenanthrenequinones	<i>o</i> -Phenanthrenequinones; <i>p</i> -phenanthrenequinones
	Anthraquinones	Anthraquinone; bianthraquinone
Phenylpropanoids	Simple phenylpropanoids	Phenylpropylenes; phenylpropanols; phenylpropyl aldehydes; phenylpropionic acids; other simple phenylpropanoids
	Coumarins	Simple coumarins; furanocoumarins; pyranocoumarins; isocoumarins; other coumarins
	Lignans	Simple lignans; monoepoxylignans; lignanolides; cyclolignans; cyclolignolides; bisepoxylignans; dibenzocyclooctenes; biphenyl-neolignans; benzoheterocyclic-neolignans; other lignans
Flavonoids	Flavones	Flavones; flavonols
	Flavanones	Flavanones; flavanonols
	Isoflavones	Isoflavones; isoflavanones
	Chalcones	Chalcones; dihydrochalcones
	Flavanols	Flavan-3-ols; flavan-3,4-diols
	Aurones	Aurones
	Anthocyanidins	Anthocyanidins
	Bisflavones	Bisflavones
Other flavanoids	Other flavanoids	
Terpenoids	Monoterpenoids	Acyclic monoterpenes; monocyclic monoterpenes; bicyclic monoterpenes; tricyclic monoterpenes; iridoides

Continued

First-level	Second-level	Third-level
Steroids	Sesquiterpenoids	Acyclic sesquiterpenes; monocyclic sesquiterpenes; bicyclic sesquiterpenes; tricyclic sesquiterpenes; tetracyclic sesquiterpenes; other sesquiterpenes
	Diterpenoids	Acyclic diterpenes; monocyclic diterpenes; bicyclic diterpenes; tricyclic diterpenes; tetracyclic diterpenes
	Sesterterpenoids	Acyclic sesterterpenes; monocyclic sesterterpenes; bicyclic sesterterpenes; tricyclic sesterterpenes; tetracyclic sesterterpenes
	Triterpenoids	Acyclic triterpenes; monocyclic triterpenes; bicyclic triterpenes; tricyclic triterpenes; tetracyclic triterpenes; pentacyclic triterpenes
	Tetraterpenoids	Carotenoids
	Cardiac glycosides	Cardenolides; scillanolides
	Steroidal saponins	Spirostanols; isospirostanols; furostanols; pseudo-spirostanols
	C-21 steroids	C-21 steroids
	Phytosterols	Phytosterols
	Bile acids	Bile acids
Alkaloids	Insect moulting hormones	Insect moulting hormones
	Other steroids	Other steroids
	Ornithine series	Pyrrolidine alkaloids; tropane alkaloids; pyrrolizidine alkaloids
	Lysine series	Piperidine alkaloids; quinolizidine alkaloids; indolizidine alkaloids
	Phenylalanine/tyrosine series	Amphetamine alkaloids; isoquinoline alkaloids; benzylphenethylamine alkaloids
	Tryptophan series	Indole alkaloids; tryptamine-indole alkaloids; hemiterpene-indole alkaloids; monoterpene-indole alkaloids
	<i>o</i> -Aminobenzoic acid series	Quinoline alkaloids; acridone alkaloids
	Histidine series	Imidazole alkaloids
	Terpene alkaloids	Monoterpene alkaloids; sesquiterpene alkaloids; diterpene alkaloids; triterpene alkaloids
	Nucleic acid bases	Nucleic acid bases
Tannins	Steroid alkaloids	Pregnane alkaloids; cholestane alkaloids
	Other alkaloids	Other alkaloids
	Hydrolyzable tannins	Gallotannins; ellagitannins; hydrolyzable tannin oligomers; <i>C</i> -glycosidic tannins; caffeetannins
	Condensed tannins	Flavan-3-ol tannins; flavan-3,4-diol tannins
Cerebrosides	Complex tannins	Complex tannins
	Cerebrosides	Cerebrosides
Saccharides & glycosides	Saccharides	Monosaccharides; disaccharides; oligosaccharides; polysaccharides
	Glycosides	<i>O</i> -Glycosides; <i>S</i> -glycosides; <i>N</i> -glycosides; <i>C</i> -Glycosides
Fatty acids	Saturated fatty acids	Saturated fatty acids
	Unsaturated fatty acids	Unsaturated fatty acids
Amino acids & proteins	Amino acids	Amino acids
	Monoamides	Monoamides
	Oligopeptides	Oligopeptides
	Cyclopeptides	Cyclopeptides
	Proteins	Proteins
Others compounds	Benzene derivatives	Phthalic derivatives; protocatechuic derivatives; salicylic derivatives; gallic derivatives; phenolic derivatives; benzyl derivatives; phenylethyl derivatives; other benzene derivatives
	O-Containing compounds	Esters; organic acids; alcohols; aldehydes; ketones; ethers
	Acyclic hydrocarbons	Alkanes; alkenes; alkynes
	Alicyclic hydrocarbons	Tricyclic hydrocarbons; cyclobutanes; cyclopentanes; cyclohexanes; cycloheptanes; cyclooctanes; bridged compounds; spiro compounds; other polycyclic hydrocarbons
	Heterocyclic compounds	Pyrans; furans; pyrazines; propylene oxides; spiro-heterocyclics; dioxetanes; dioxolanes; lactones; thiazoles
	Benzo heterocycles	Benzopyrans; benzofurans; benzothiophenes; benzodioxoles; benzothiazoles
	Naphthalenes	Naphthalenes
	Anthracenes	Anthracenes
	Phenanthrenes	Phenanthrenes
	Fluorenes	Fluorenes
	P-Containing compounds	P-Containing compounds
	S-Containing compounds	S-Containing compounds
	Si-Containing compounds	Si-Containing compounds
	N-Containing compounds	N-Containing compounds
	Halogenated hydrocarbons	Halogenated hydrocarbons

数据分析工具及方法 使用 Excel 进行建库及频数分析, 采用 IBM SPSS Statistics 26 进行一般统计分析。使用 IBM SPSS Modeler 18.0 中 Apriori 关联规则分析及特征选择研究寒热药性物质基础, 使用 IBM SPSS Modeler 18.0 中多元逻辑回归分析建立寒热识别模型并评估预测准确度。

结果

唇形科 245 味中药中共搜集到 193 味中药的物质成分信息, 其中 4 味中药寒热药性不明, 予以剔除。此外, 一级分类 < 2 的中药 17 味, 也予以剔除, 共得到 172 味中药, 10 726 个物质成分。其中, 寒性中药 97 味、物质成分 6 598 个, 热性中药 44 味、物质成分 2 764 个, 平性中药 31 味、物质成分 1 364 个。

1 频数分析

在一级分类下, 唇形科中药萜类物质最多, 达 49.73% (图 1), 其次为其他类 (17.29%)、黄酮类 (11.73%)。各类物质在寒、热、平性中药中的占比有一定差异, 其中, 脂肪酸类、甾体类、醌类物质的占比差异最大。卡方检验结果显示, 在一级分类水平上, 寒、热、平性中药物质成分构成比有显著差异 ($P < 0.05$)。

在二级分类下, 唇形科中药单萜物质最多, 达 18.94% (图 1), 其次为倍半萜 (14.16%)、二萜 (10.60%)。各类物质在寒、热、平性中药中的占比有差异, 其中, 单萜、

倍半萜、二萜、一般含氧化合物、植物甾醇等物质的占比差异最大, 占比差异明显大于一级分类。卡方检验结果显示, 在二级分类水平上, 寒、热、平性中药物质成分构成比有显著差异 ($P < 0.05$)。

在三级分类下, 唇形科中药单环单萜物质最多, 达 8.76% (图 1), 其次为黄酮 (6.82%)、双环倍半萜 (6.09%)。各类物质在寒、热、平性中药中的占比有差异, 其中, 单环单萜、双环倍半萜、双环二萜、苯丙酸、植物甾醇、环烯醚萜等物质的占比差异最大。卡方检验结果显示, 在三级分类水平上, 寒、热、平性中药物质成分构成比有显著差异 ($P < 0.05$)。

各类物质在寒性、热性、平性中药中的占比差异随着物质分类等级变化而变化, 各级分类下平均变异系数分别为 42.30%、79.07%、91.51%, 即占比差异一级分类 < 二级分类 < 三级分类 (图 1)。

2 关联规则分析

2.1 各级分类下有效关联规则数量 分别将一级、二级、三级分类的物质作为前项, 寒性、热性、平性作为后项, 选择置信度 70%、最大前项数 5, 进行关联规则分析。考虑到唇形科寒性、热性、平性中药数量差别较大, 占比不同 (寒性中药是热性中药的 2 倍多), 寒性、热性支持度分别设为 15%、8%。因平性中药仅占 18.02%, 置信度设为 60%, 增加置信率作为限制条件, 设为 70%。分别对一级、二级、三级分类进行关联规则分析, 选择增益大于 1.2 的规则。一级分类下, 规则数量为 0 条, 二级规则数为 79 条, 三级规则数为 193 条。其中, 二级分类下, 当热性作为后项时, 规则数也为 0 条。可见, 一级分类、二级分类条件下, 无法对寒热药性进行有效区分, 因此, 本研究仅对三级分类下关联规则进行进一步分析。

2.2 定义“核心类群”和“或有类群” 关联规则分析算法决定了其分析结果或含有“冗余规则”。在三级分类条件下, 对 193 条关联规则结果做进一步分析, 发现规则之间有严重冗余信息。以寒性作为后项的 77 条规则为例, 选取其中的 10 条规则来说明“冗余规则”与“核心规则”的关系。如表 2 所示, 10 条规则的支持度、置信度、增益均满足算法要求 (支持度大于 15%、置信度大于 70%、增益大于 1.2), 是 Apriori 算法分析结果所呈现出的有效规则, 前项数 3~5 个。其中, 第一条规则的前项有且仅有“苯丙酸+五环三萜+双环单萜”, 说明其前项仅包含苯丙酸、五环三萜、双环单萜 3 类物质。其余 9 条规则的前项为 4 或 5 个, 且每条规则的前项都包含苯丙酸、五环三萜、双环单萜 3 类物质, 是这 3 类物质与其他物质的组合。与第一条规则相比较, 根据所组合的物质不同, 规则的支持度下降, 而置信度提

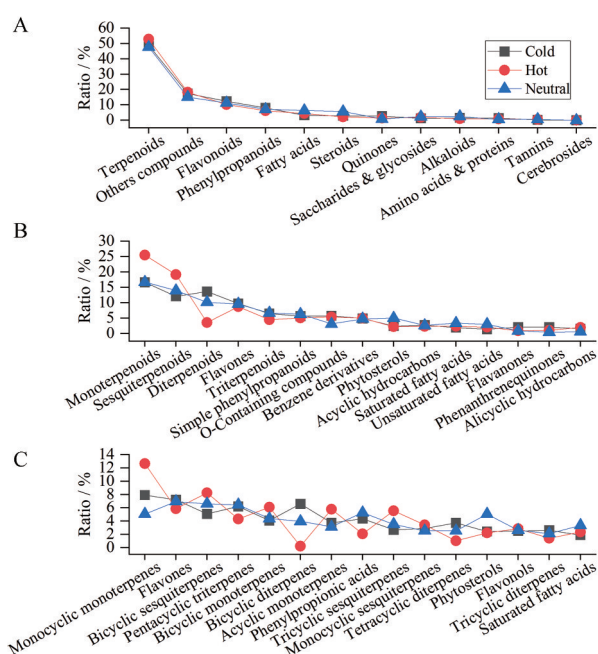


Figure 1 The material composition ratios for three levels of classification. A: The first-level classification; B: The second-level classification; C: The third-level classification. Only the top 15 items are shown in the figure

Table 2 Effective association rules with redundant rules

Antecedent	Consequent	Antecedent support/%	Confidence /%	Rule support/%	Lift
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes	Cold	22.67	71.79	16.28	1.27
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + monocyclic monoterpenes	Cold	22.09	71.05	15.70	1.26
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + bicyclic sesquiterpenes	Cold	18.60	71.88	13.37	1.27
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + bicyclic sesquiterpenes + monocyclic monoterpenes	Cold	18.02	70.97	12.79	1.26
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + bicyclic sesquiterpenes + acyclic monoterpenes	Cold	18.60	71.88	13.37	1.27
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + bicyclic sesquiterpenes + phytosterols	Cold	17.44	73.33	12.79	1.30
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + acyclic monoterpenes	Cold	19.77	70.59	13.95	1.25
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + acyclic monoterpenes + phytosterols	Cold	18.60	71.88	13.37	1.27
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + phytosterols	Cold	21.51	72.97	15.70	1.29
Phenylpropionic acids + pentacyclic triterpenes + bicyclic monoterpenes + phytosterols + monocyclic monoterpenes	Cold	20.93	72.22	15.12	1.28

升。说明,这9条规则均为第一条规则的衍生规则,视为“冗余规则”,而第一条规则视为“核心规则”。将“核心规则”的前项定义为“核心类群”,即苯丙酸、五环三萜、双环单萜定义为“核心类群”。其余9条规则中与之组合的其他物质,如单环单萜、双环倍半萜、无环单萜等定义为“或有类群”。

2.3 核心类群 将寒性作为后项的77条规则删除冗余,得到19条寒性中药核心规则、16个寒性中药核心类群,苯丙酸、五环三萜、双环单萜、植物甾醇等物质出现的频率较高。将108条以热性为后项的规则删除冗余,得到26条热性中药核心规则、15个热性中药核心类群,邻苯二甲酸类、苯丙烯、饱和脂肪酸等物质出现的频率较高。将8条以热性为后项的规则删除冗余,得到3条平性中药核心规则、5个平性中药核心类群,内酯、氧苷、其他生物碱等物质出现的频率较高。

2.4 特异类群 关联规则分析过程中,有一些特殊的规则,虽然支持度较低,但是置信度极高,甚至达到100%。这说明,有一些物质在寒热药性的识别中具有极其特殊的作用,即某中药含有该物质即可认定为拥有某种药性。为此,将支持度极低,置信度极高(100%),前项数为1的规则定义为“特异规则”,该前项定义为“特异类群”。本研究共纳入中药172味,因此将最低条件支持度设为1/172,即0.581%。在三级分类下,设置最低条件支持度0.581%、规则置信度100%、最大前项数1,得到以寒性为后项的特异规则19项,特异类群19个,包含查耳酮、其他倍半萜、四元脂环烃等;以热性为后项的特异规则12项,特异类群

12个,包含脑苷、螺环化合物、含磷化合物等;以平性为后项的特异规则3项,特异类群3个,包含单萜吡喹啉类生物碱、二氢查耳酮、二氢异黄酮。

2.5 核心类群的交互关系 综上,寒性中药核心类群16个、热性15个、平性5个,共36个。其中,双环单萜、植物甾醇、单环单萜、单环倍半萜、双环倍半萜、无环单萜、黄酮、三环倍半萜、饱和脂肪酸等9类物质,既是寒性中药的核心类群,又是热性的核心类群,寒热药性的物质基础互有交叉。结果如图2所示。其中,核心类群27个(寒热共有类群不重复计算),特异类群34个。图中数值代表核心规则中某类物质成分的最大前项支持度。

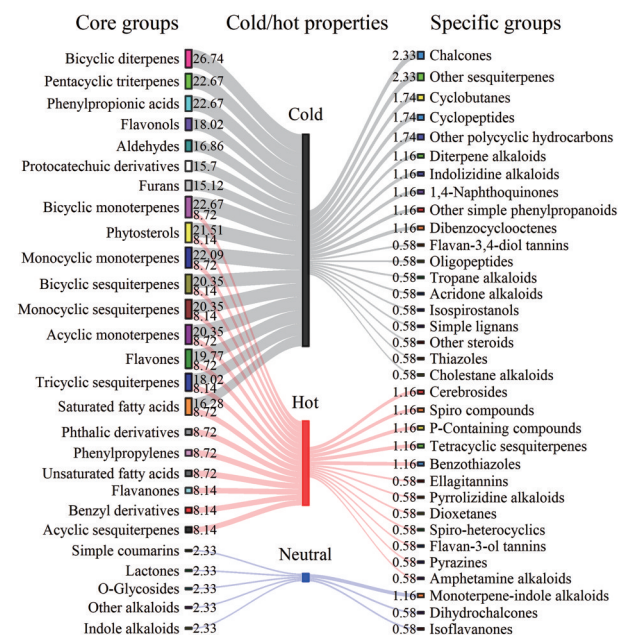
3 逻辑回归建模

采用逻辑回归建立寒热药性识别模型,以寒、热、平三性作为模型输出,平性为基准类别,以三级分类作为输入,采用进入法建立多元逻辑回归模型,并对模型进行评价。其中,方程显著性使用释然比检验,变量显著性使用Wald检验,拟合优度使用Nagelkerke R^2 及混淆矩阵的准确度来衡量。结果见表3。其中,混淆矩阵内包含寒性准确度、热性准确度、平性准确度和总体准确度。

当自变量为核心类群+特异类群时,总体准确度82%,模型效果良好(模型1)。但是,此时需要考察61个自变量,即61种物质类群,模型实用性较差。当自变量为核心类群时,自变量27个,实用型较好,但总体准确度66.9%,模型准确度较差(模型2)。进一步考察变量间的交互作用,结果显示,仅考虑核心类群的变量

Table 3 Results of logistic regression models

No.	Independent variable	No. of variables	Sig. (<i>P</i> -value)	Nagelkerke R^2 value	Percent correct/%			
					Cold	Hot	Neutral	Total
1	Core groups + specific groups	61	0.000	0.762	88.7	72.7	74.2	82
2	Core groups	27	0.000	0.507	82.5	40.9	54.8	66.9
3	Core groups (single interaction)	165	0.595	0.984	97.9	95.5	96.8	97.1
4	Core groups (all 2-way interactions)	165	0.927	0.95	96.9	97.7	96.8	97.1
5	All groups in the third-level	142	0.006	0.997	100	97.7	100	99.4
6	Core groups + intermediate variables	75	0.000	0.872	92.8	84.1	77.4	87.8
7	Intermediate variables + specific groups	82	0.005	0.722	100	47.7	35.5	75
8	Core groups + intermediate variables + specific groups	109	0.000	0.924	99	88.6	87.1	94.2

**Figure 2** Core groups and specific groups of cold and hot properties in Lamiaceae family

交互作用时,能得到97.1%的预测正确率(模型3、4)。这说明,单独考察核心类群就能很好地识别寒热药性。通过对前述关联规则分析的结果进行回顾,发现几乎所有核心前项均为物质分类的组合,极少有单一物质种类。因此,以核心前项为基础,使用核心前项构造中间变量,进行逻辑回归建模,得到含有中间变量的寒热识别模型(模型6~8)。结果表明,以核心类群+中间变量作为输入变量时,准确率达87.8%,准确性和实用型均良好。这充分证明了核心类群能够很好地识别寒热药性,核心类群可能是寒热药性的决定性物质基础。但是,单一物质种类无法决定药性,寒热药性是由多种类别的物质共同决定的,即,物质类别的不同组合方式

决定了寒热药性。

4 决定性类群分析

以寒热药性为目标,对上述核心类群中27个物质类别(寒热共有类群不重复计算,图2)做特征选择,提取“决定性类群”,利用似然比卡方检验,选择1-概率*P*值大于0.8,得到15个重要变量,结果见表4。

Table 4 Results of feature selection based on core groups

No.	Variable	<i>P</i> -value of 1-probability
1	Bicyclic diterpenes	1.00
2	Monocyclic monoterpenes	1.00
3	Bicyclic monoterpenes	0.99
4	Phthalic derivatives	0.99
5	Phenylpropylenes	0.98
6	Acyclic monoterpenes	0.97
7	Aldehydes	0.96
8	Tricyclic sesquiterpenes	0.93
9	Other alkaloids	0.92
10	Pentacyclic triterpenes	0.91
11	Phytosterols	0.91
12	Bicyclic sesquiterpenes	0.87
13	O-Glycosides	0.87
14	Monocyclic sesquiterpenes	0.83
15	Protocatechuic derivatives	0.82

选择这个15个变量,以寒热药性为目标,采用进入法建立多元逻辑回归模型,并对模型进行评价。其中,方程显著性使用释然比检验,变量显著性使用Wald检验,拟合优度使用Nagelkerke R^2 及混淆矩阵的准确性来衡量。结果见表5。采用变量双向交互时(模型3),得到良好的预测模型,方程显著性*P*值小于0.05, Nagelkerke R^2 值为0.919,准确性89%,寒性准确度94.8%,热性79.5%,平性83.9%。推测,上述15个变量可以称为“决定性类群”,可能是唇形科中药寒热药性的决定性物质。

Table 5 Logistic regression results based on 15 variables

No.	Independent variable	No. of variables	Sig. (<i>P</i> -value)	Nagelkerke R^2 value	Percent correct/%			
					Cold	Hot	Neutral	Total
1	Decisive groups	15	0.000	0.368	85.6	40.9	29	64
2	Decisive groups (single interaction)	127	0.261	0.919	99	72.7	80.6	89
3	Decisive groups (all 2-way interactions)	112	0.023	0.919	94.8	79.5	83.9	89
4	Decisive groups (all 3-way interactions)	127	0.261	0.919	95.9	77.3	83.9	89

讨论

药性理论是中药和中医基础理论之间的桥梁。当前,基于物质实体的中药药性理论的科学诠释虽取得了一定的成果,对进一步深化中药药性的认识也具有一定指导作用^[36],但微观研究较多,微观与宏观的联系不够,进展缓慢,许多研究在设计之初就已经脱离了中医药理论体系,难以取得突破性飞跃,难以实现理论成果向临床应用的转化。现有的研究往往将一组功效相似的中药作为研究对象,规律性不强,可比性差。本研究引入药用植物亲缘学,在一定亲缘关系内,综合考察各级各类化学成分与药性的关系。选取唇形科中药作为切入点,宏观和微观相结合,探究唇形科中药寒热药性的表征模式。结果显示,唇形科中药寒热药性的物质基础差异明显,且互有交叉;唇形科中药寒热药性往往由多种类别的物质成分交互作用而体现,单一类别的物质常常不能用来表征寒热药性,多种物质的有机组合是寒热药性的物质基础,双环二萜等15类物质是其决定性类群。

寒热药性的物质基础研究对中医药理论现代化阐释^[28,29]、中药质量标志物研究^[37]、中药质量总体评价^[38]、炮制及配伍机制研究^[39]、复方药性分析^[40]、新型药用资源开发等至关重要^[41]。本研究建立了多个寒热药性识别模型(表3、5),综合考虑各方面因素,具有现实意义的为表3中模型1(准确度82%,自变量61个)与模型6(准确度87.8%,自变量75个)以及表5中的模型3(准确度89%,自变量112个)。尽管表5中的模型3纳入的自变量不多,仅为15个决定性类群,但是回归模型的变量却高达112个,这是变量交互的结果。对这112个变量分析发现,要想准确预测药性,就必须明确15个决定性类群的每一类物质的有和无,某些物质的“有”固然重要,而另一些物质的“无”同样重要。这就要求科研工作者们对药物的成分分析全面而准确。

虽然取得预期成效,但是本研究也存在不足。第一,因关联规则分析的需要,本研究数据分析过程中,将数据进行了“归一化”处理,即仅考察了物质在某一味中药中是否出现,并未考察出现的数量(频次)以及物质的含量,这势必减少数据信息的丰富程度。这样的数据是关联规则分析的数据要求,但是并非逻辑回归、特征选择的数据要求。第二,物质成分的一级、二级、三级分类标准能否体现物质成分化学信息、体内效应复杂度的变化等问题并未深入讨论;酚酸类、挥发油类等物质是否需要单列为一个大类,萜类物质的分类是否需要进一步细化,以及苷类物质以苷元为划分物质种类的依据是否合理等问题仍需进一步研究。第三,本研究仅从中药本身成分的角度去考察药性,未能

从“药物-机体”二者相结合的模式去思考中药药性;一些物质成分频次很低,如黄烷-3,4-二醇类鞣质、萜类生物碱、螺杂环、逆没食子鞣质等频次为1,但是,考虑到它们或许是寒热药性的特有类群,在本研究中并未舍弃。第四,本研究在化学成分分类基础上研究已知物质成分构成的唇形科全部中药,逻辑回归模型的正确率是组内回代正确率,未考察模型泛化能力,得到的结论需要进一步深入研究和验证。总之,本研究从物质成分化学骨架层次讨论了药性的相关性,虽未涉及诸如物质成分同分异构体间生理活性差异性的讨论等内容,但是,研究为后期进一步开展物质成分含量与药性相关性研究,以及开展以化学信息学、物质成分生理活性等为基础的更深层次药性研究奠定了基础。

作者贡献: 张传耀负责选题与设计、方法建立、数据处理、文章撰写等工作;周晴对数据预处理、数学建模、文章撰写等提供了帮助;王校在方法摸索及数据处理方面提供了帮助;卜菲菲对药性理论提供了帮助;王鹏负责总体研究方案设计、论文审阅指导及研究经费支持。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Wang P. Construction of the Four Traditional Chinese Medicine Theory and Methodology of Modern Research (中药四性理论构建及其现代研究方法学探讨) [D]. Jinan: Shandong University of Traditional Chinese Medicine, 2008.
- [2] Huang LQ. Discussion on the future research of natural characteristics of Chinese herbal medicine [J]. Pharm Clin Chin Mater Med (中药与临床), 2011, 2: 1-3.
- [3] Liu J, Feng W, Peng C. A song of ice and fire: cold and hot properties of traditional Chinese medicines [J]. Front Pharmacol, 2021, 11: 598744.
- [4] Zhao YL, Wang JB, Xiao XH, et al. Study on the cold and hot properties of medicinal herbs by thermotropism in mice behavior [J]. J Ethnopharmacol, 2011, 133: 980-985.
- [5] Wang M, Sun YP, Wang ZB, et al. Comment and prospect of research on TCM nature and flavour theory [J]. China J Tradit Chin Med Pharm (中华中医药杂志), 2021, 36: 625-628.
- [6] Lyu CY, Lyu SW, Li GY, et al. Research progress on pharmacological effects of separation and combination methods based on traditional Chinese medicine property and flavor [J]. China J Chin Mater Med (中国中药杂志), 2018, 43: 2892-2898.
- [7] Kuang HX, Wang YH, Wang QH, et al. A research model on the theory of TCM property and flavor based on TCM splittability and combinativeness [J]. Mod Tradit Chin Med Mater Med World Sci Technol (世界科学技术-中医药现代化), 2011, 13: 25-29.
- [8] Xia YS, Wei GH, Wang ZG, et al. Review on research methods of cold and heat property of Chinese materia medica [J]. China J

- Tradit Chin Med Pharm (中华中医药杂志), 2021, 36: 990-992.
- [9] Ouyang B, Wang ZG, Li F, et al. Hypothesis and demonstration of "property-response-component" for the four properties of traditional Chinese medicine [J]. J Shandong Univ Tradit Chin Med (山东中医药大学学报), 2008, 172: 182-183.
- [10] Wang ZG, Wang P, Ouyang B. Methodological frame of relativity research on TCM substance component and drug nature of cold and hot [J]. J Zhejiang Univ Tradit Chin Med (浙江中医药大学学报), 2009, 33: 734-737.
- [11] Fu XJ, Wang ZG, Li XB, et al. Relationship between the nature and the structure of main compounds from Chinese herbs based on the similarity of chemical structure [J]. China J Tradit Chin Med Pharm (中华中医药杂志), 2019, 34: 2657-2661.
- [12] Yang HJ, Tang SH, Huang LQ, et al. Research on nature of Chinese materia medica in view of genetic relationship [J]. Chin J Chin Mater Med (中国中药杂志), 2008, 33: 2983-2985.
- [13] Li ZY, Wang JQ, Li CF, et al. Strategies for efficacy and function positioning of new resources of Chinese materia medica [J]. China J Chin Mater Med (中国中药杂志), 2021, 46: 3455-3464.
- [14] Park SM, Baek SJ, Ban HJ, et al. Systematic analysis of the molecular mechanisms of cold and hot properties of herbal medicines [J]. Plants, 2022, 11: 997.
- [15] Han S, Lyu AP, Li J, et al. Application review of network pharmacology in the study of properties theory of traditional Chinese medicine [J]. J Basic Chin Med (中国中医基础医学杂志), 2019, 25: 127-130.
- [16] Wang Y, Zhou L, Zhang S, et al. Identification of pathways and genes associated with cold and hot properties of Chinese materia medica based on bioinformatics analysis [J]. Phytochem Lett, 2020, 38: 70-77.
- [17] Yu S, Li C, Ding Y, et al. Exploring the 'cold/hot' properties of traditional Chinese medicine by cell temperature measurement [J]. Pharm Biol, 2020, 58: 208-218.
- [18] Li A, Li A, Deng Z, et al. Cross-species annotation of expressed genes and detection of different functional gene modules between 10 cold- and 10 hot-property Chinese herbal medicines [J]. Front Genet, 2020, 11: 532.
- [19] Jia GX, Dong YQ, Yang N, et al. Proteomics strategy contributes research progress of traditional Chinese medicine modernization [J]. J Tianjin Univ Tradit Chin Med (天津中医药大学学报), 2021, 40: 277-285.
- [20] Chen PP, Zhang YN, Wang Z, et al. Based on the method of proteomics, Radix Scutellariae's effect on the material energy metabolism in Re Zheng rats model [J]. Pharmacol Clin Chin Mater Med (中药药理与临床), 2018, 34: 90-96.
- [21] Li M, Hua S, Huang X, et al. Non-targeted metabonomics to investigate the differences in the properties of ginseng and American ginseng based on rapid resolution liquid chromatography coupled with quadrupole-time-of-flight mass spectrometry [J]. J Sep Sci, 2021, 44: 3497-3505.
- [22] Guo J, Wang J, Iino K, et al. Quantitative and molecular similarity analyses of the metabolites of cold- and hot-natured Chinese herbs [J]. Evid Based Complement Alternat Med, 2021, 2021: 6646507.
- [23] Wang YX, Lyu C. Study progress on the cold and heat property of Chinese herbal medicine [J]. World J Integr Tradit West Med (世界中西医结合杂志), 2019, 14: 889-893.
- [24] Zhang YQ, Wang P, Ji YJ, et al. The property of traditional Chinese materia medica based on the theory of matter [J]. J Shandong Univ Tradit Chin Med (山东中医药大学学报), 2011, 35: 291-295.
- [25] Wei G, Fu X, Wang Z. Nature identification of Chinese herbal medicine compounds based on molecular descriptors [J]. J AOAC Int, 2021, 104: 1754-1759.
- [26] Wang X, Cao T, Tian X, et al. Quantification of "cold-hot" medicinal properties of Chinese medicines based on primary metabolites and fisher's analysis [J]. Comput Math Methods Med, 2022, 2022: 5790893.
- [27] Huang Y, Yao P, Leung KW, et al. The Yin-Yang property of Chinese medicinal herbs relates to chemical composition but not anti-oxidative activity: an illustration using spleen-meridian herbs [J]. Front Pharmacol, 2018, 9: 1304.
- [28] Huang LQ. Thoughts on theoretical research of traditional Chinese medicine-speech at the 4th academic week of institute of basic theory for traditional Chinese medicine, China academy of Chinese medical sciences [J]. J Basic Chin Med (中国中医基础医学杂志), 2020, 26: 2-4.
- [29] Zheng WK, Yan D, Zhang XM, et al. Major scientific issues and engineering technical difficulties of traditional Chinese medicine [J]. J Tradit Chin Med (中医杂志), 2019, 60: 991-1000.
- [30] Cao J, Wang Y. Relationship between chemical constituents and herbs properties of relative plant herbs [J]. China J Chin Mater Med (中国中药杂志), 2013, 38: 453-458.
- [31] Xie HH, Chen C, Wang P. Correlation between object image and nature in Labiate herbs [J]. J Yunnan Univ Tradit Chin Med (云南中医学院学报), 2017, 40: 91-93.
- [32] Chen C, Wang P, Xie HH. Research progress of the relationship between compositions and drug properties of Chinese medicine based on group classification [J]. J Gansu Univ Chin Med (甘肃中医药大学学报), 2018, 35: 90-93.
- [33] Xie HH. Study of Correlation between Object Image and Nature in Traditional Chinese Medicine Base on Labiatae (中药基原物象与寒热药性的相关性研究) [D]. Hefei: Anhui University of Chinese Medicine, 2017.
- [34] Chen C. Study on the Correlation between Chemical Compositions of Labiatae Chinese Medicine and Medicinal Properties (唇形科中药物质成分与药性的相关性研究) [D]. Hefei: Anhui University of Chinese Medicine, 2018.
- [35] Zhou Q, Zhang CY, Wang P. Study on the correlation between the chemical components and cold-heat properties of 123 Labiate

- herbs based on support vector [J]. *LiShiZhen Med Mater Med Res* (时珍国医国药), 2022, 33:1761-1764.
- [36] Bi L, Liu X, Cui YR. Study on the property of traditional Chinese materia medica under the theory of matter [J]. *J Beijing Univ Tradit Chin Med* (北京中医药大学学报), 2022. <http://kns.cnki.net/kcms/detail/11.3574.R.20220822.1514.004.html>.
- [37] Zhang TJ, Bai G, Liu XC. The concept, core theory and research methods of Chinese medicine quality markers [J]. *Acta Pharm Sin* (药学报), 2019, 54: 187-196.
- [38] Sun Y, Xu G, Ma SC. Development of an overall evaluation system for traditional Chinese medicine [J]. *Acta Pharm Sin* (药学报), 2021, 56: 1749-1756.
- [39] Li W, Wang ZJ, Lin XY, et al. Study on the substance basis of "property-taste-efficacy" of *Liquorice* and *Rhizoma chinensis* based on supramolecular system induced by weak bond [J]. *Acta Pharm Sin* (药学报), 2022, 57: 1901-1908.
- [40] Ping Y, Li YP, Lyu SW, et al. Construction of the theory of property delivery in compound TCM formulas [J]. *Chin Tradit Pat Med* (中成药), 2021, 43: 1823-1829.
- [41] Kou RB, Guo M, Guo YF, et al. Research progress in medicinal chemistry of TCM [J]. *Chin J Inf Tradit Chin Med* (中国中医药信息杂志), 2022, 29: 142-146.