

中药材 DNA 条形码信息查询系统设计与实现

刘 蕤 李照东 张雨萌

(华中师范大学信息管理学院 武汉 430079)

〔摘要〕 介绍 DNA 条形码技术及国内外药用植物相关信息平台建设情况, 阐述中药材 DNA 条形码可视化查询系统的设计与实现, 指出该系统的建设能够推进物种 DNA 条形码技术的应用, 为相关研究者提供参考与借鉴。

〔关键词〕 中药材 DNA 条形码; 信息查询; 系统设计

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.05.009

Design and Implementation of DNA Barcode Information Query System of Chinese Medicinal Materials LIU Rui, LI Zhao-dong, ZHANG Yu-meng, School of Information Management, Central China Normal University, Wuhan 430079, China

〔Abstract〕 The paper introduces DNA barcode technology and building of related information platforms of medicinal plants domestic and abroad, explicates design and implementation of DNA barcode visualization information query system of Chinese medicinal materials, and points out that the building of the system can help to advance application of species DNA barcode technology, to provide related researchers with reference and experience.

〔Keywords〕 DNA barcodes of Chinese medicinal materials; Information query; System design

1 引言

中药材是中国传统医学指导下应用的原生药材, 已有上千年的历史。在现代科学技术高速发展的今天, 这些古老但具有显著疗效的中药材依旧被广泛运用。但是中药材的管理一直存在品种混乱、掺伪的问题, 极大程度上影响中医药的疗效和安全^[1]。传统中药材的鉴定方法只是对原植物、性状、显微、理化鉴定, 这些方法有一定的效果, 但

其准确性、适用性远远满足不了需求, 对于亲缘关系比较近的物种更是无法区分^[2]。而 DNA 条形码 (DNA Barcoding) 技术能为中药材建立唯一的信息身份证, 快速有效地对药材进行分类和鉴定以确保中药基源鉴定的标准化, 同时也方便科研工作者持续开展相关研究。本研究设计并实现中药材 DNA 条形码可视化查询系统, 该系统拥有完备的中药材 DNA 条形码数据库, 能够快速查询中药材及 DNA 条形码信息。

2 相关研究

2.1 DNA 条形码技术

2003 年加拿大分类学家 Paul Hebert 首次提出

〔收稿日期〕 2018-01-21

〔作者简介〕 刘蕤, 副教授, 发表论文 10 余篇, 参编论著 3 部; 通讯作者: 李照东, 硕士。

DNA 条形码概念, 研究发现通过单一短的基因片段来鉴定物种会有比较好的效果^[3]。这一研究成果引起各国分类学家的密切关注, 由此开始 DNA 条形码技术研究的黄金时代。此后的 10 多年中 DNA 条形码技术被应用在生态学、系统分类学、生物多样性保护等多个学科研究领域。其中 Miller 发文认为 DNA 条形码鉴定技术正在推动分类学的“文艺复兴”, 这一技术的研究进展给毫无生气的形态分类学带来巨大的转变^[4]。DNA 条形码技术能够使用较短的基因片段对物种进行快速、准确的鉴定, 获取和其他物种之间的差异性信息, 实现标准化的监测。因此重点开发 DNA 条形码技术, 构建 DNA 条形码的数据集很有必要。2004 年国际条形码联盟成立, 建立数据库工作组与全球生物多样性信息设备, 希望能够与各国机构联合建立以 DNA 条形码记录来分类、鉴定物种。通过对符合标准的数据库分配 Barcode 记号来确认每一个被认可的物种名称, 得到鉴定生物最佳的凭证标本^[5]。

2.2 国内外药用植物相关信息平台

2.2.1 国外 生命数据条形码系统 (Barcode of Life Data Systems, BOLD) 是全世界最大、最早的 DNA 条形码数据库平台, 由生命条形码联盟维护, 截至目前已有 273 688 个物种的 DNA 条形码^[6]。其主要功能涵盖 3 个层面: 收集上传生物数据的条形码、样本鉴定和外部连接系统。规范标准 DNA 条形码信息的 7 种必需的数据元素, 由研究者和用户负责上传, 提交的关键数据信息是样本序列的数据, 包括样本序列氨基酸转换和彩色条形码^[7]。BOLD 系统更像一个 DNA 条形码的数据仓库管理平台, 无法直接查询或浏览 DNA 条形码。

2.2.2 国内 自 DNA 条形码概念提出后我国科研机构也开始推进中国动植物 DNA 条形码的研究工作。中医科学研究所陈士林课题组首次在国际上提出并验证 ITS2 序列可以作为药用植物通用条形码, 应用于物种及其混伪品的鉴定^[8]。通过近 10 年的研究积累已经完成 11 000 余种、4 万余份实验样品的 DNA 条形码研究, 首创中草药 DNA 条形码鉴定

技术体系并且被纳入《中国药典》^[9]。课题组构建涵盖各国药典中 95% 中草药的 DNA 条形码鉴定数据库和平台网站。通过该网站能够高效、精准地鉴定相应样品, 有效地监督中草药的流通, 保证用药安全, 然而该系统不能直接查询或浏览对应的物种, 无法实现查询结果的可视化。现存的中药数据库中数量最大、最全的是中药综合数据库 (Traditional Chinese Medicines Integrated Database, TC-MID), 其中包含 8 000 余种中药, 47 000 个处方, 27 000 多种化合物的信息^[10]。该平台提供数据库的查询功能, 可以选择处方、中药、疾病、靶点等不同信息来检索。检索结果包括名称、归属、中药分组表、中药组分的化学、分子结构信息等。该系统信息量丰富, 基于 6 个互相关联的数据集, 一定程度上揭示中医理论底层的分子生物学机理^[10], 但该系统也没有提供 DNA 条形码相关数据。对各系统平台的功能分析可知, 有些数据库只提供鉴定功能, 无法进行快速有效的查询、浏览; 有些数据库具备比较系统全面的查询平台, 但查询结果并没有深入到 DNA 条形码。本研究设计并实现中药材 DNA 条形码可视化查询系统, 能够实现以物种名、拉丁名、DNA 条形码序列为检索式的物种检索和序列比对信息的呈现。

3 系统设计与实现

3.1 总体设计

本系统设计基于 Web 平台, 用户通过网络访问。查询系统可以直接查询所需要的中药材信息, 也可以通过信息浏览的方式以目录形式访问。主要采用 Perl 的 Web 开发框架 Catalyst。Perl 集合多种脚本语言以及很多其他程序语言的特性, 适用范围广泛。其开源免费, 易于快速开发利用, 功能强大, 适合本系统平台^[11]。Catalyst 是一个开源的 Perl Web MVC 框架, 鼓励快速开发和简洁的设计, 而不妨碍规则, 非常受开发者欢迎^[12]。系统架构, 见图 1。

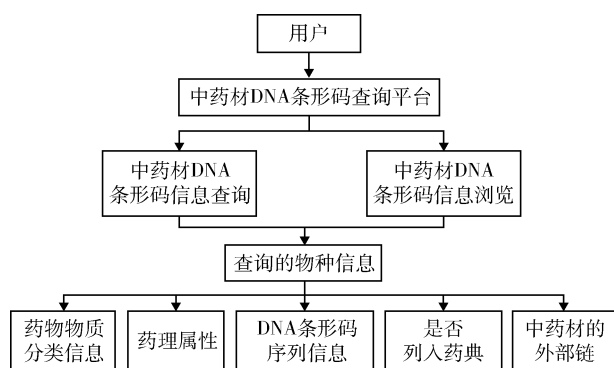


图 1 系统架构

3.2 系统概述

基于前人的研究成果,本系统数据库中收录 3 905 个物种的 4 289 种中药材、62 809 个 DNA 条形码。该数据库基本覆盖常用的中药材,数据库量较大,可称之为 Web 版本的本草纲目。系统提供一个简便、易操作的中药材全信息查询与浏览平台,所显示的物种信息包括分类信息、药理信息、药典相关信息、DNA 条形码、外源数据库等。

3.3 中药材物种信息查询

(1) 概述。系统提供中药材物种的信息查询以及 DNA 条形码的数据查询。中药材信息的查询可以通过系统查询页面完成。在查询页面中可以使用拉丁语、英语、拼音以及付税 ID 号 (Taxid)。提交检索式后即可看到查询结果。在查询结果页面中选择继续了解的药材,点击后将直接进入物种信息页面,显示中药材的物种信息。(2) 物质分类信息。包括拉丁文名称、英语名称、汉语拼音名称、种、科、目。(3) 药用植物的药理属性。对于“成分”,每项信息都具有超链接,可以链接到关于成分的其他信息。如点击 8426,即可跳转到新页面,显示 8426 所代表的人参皂苷 (Ginsenoside Rd, Rd) 的成分信息,包括人参皂苷 Rd 的处方、结构、相关中草药等,成分信息由 TCMID 系统提供,药理属性信息。(4) 药典信息。药用植物是否属于任何特定列表 (列入药典)。(5) 药用植物 DNA 条形码序列详细信息。包括序列的最大长度、序列比对中的序列总数、队列中总共碱基数、队列是否对齐,即所有序列具有相同的长度、队列中序列的成对同

一性百分比的平均值等。(6) 所检索的中药材的外部链接。包括来自 NCBI、PubMed、eFlora、BOLD 等系统提供的图片、参考文献和样品采集信息等。

3.4 中药材物种信息浏览

系统提供中药材物种信息及 DNA 条形码数据浏览功能,通过浏览功能可以快速浏览所有的中药材物种信息,类似于图书的目录。浏览页面目前支持拉丁名称、英文名称和中文名称作为查询字词。

4 结语

中药材 DNA 条形码可视化查询系统对中药材的标准化、规范化管理大有裨益,凝聚多年中药材科研工作者的智慧结晶。目前基本覆盖绝大多数的中药材物种 DNA 信息,拥有可方便查询的“DNA 身份信息库”。基于该系统科研工作者能更加深入地从 DNA 分子结构水平上对中药材进行研究、对比分析。但是该系统只是基于 Web 的版本,还未适配到移动平台,今后将研究拓展到各种平台上以确保有更加广泛的适用性。总之中药材 DNA 条形码可视化查询系统的建设推进物种 DNA 条形码技术的应用研究,为后续研究者提供参考。

参考文献

- 1 郭兰. 中草药 DNA 整合鉴定系统研究 [D]. 武汉: 湖北中医药大学, 2016.
- 2 孙得森, 王欣然, 王京昆. 中药材鉴定方法概述 [J]. 中南药学, 2017, 15 (4): 487-491.
- 3 郭慧, 王谦博, 贾力维, 等. 中药材 DNA 条形码技术研究进展 [J]. 中国药师, 2016, 19 (3): 566-570.
- 4 辛天怡, 雷美艳, 宋经元. 中药材 DNA 条形码鉴定研究进展 [J]. 中国现代中药, 2015, 17 (2): 170-176, 184.
- 5 中国中医科学院中药研究所. 中药材 DNA 条形码鉴定系统 [EB/OL]. [2017-10-20]. <http://www.tcm-barcode.cn/china/?optionid=110>.
- 6 曾春霞, 王亚楠, 王雨华, 等. 植物 DNA 条形码与生物多样性数据共享平台构建 [J]. 生物多样性, 2014, 22 (3): 285-292.

(下转第 50 页)

5 结语

临床指南是临床医生在诊疗过程中所遵循的规范化文件, 糖尿病防治指南较为完整地总结我国糖尿病防治和诊疗的实践经验, 为糖尿病及其并发症的防治起到关键作用。为进一步提高临床指南的作用和价值, 本研究利用本体技术构建糖尿病指南的本体库, 将其中的知识概念进行语义关联, 在糖尿病本体的基础上设计基于语义的检索模型, 实现语义推理与语义检索, 为糖尿病管理系统的开发提供借鉴。

参考文献

- 1 闫冠韞, 陈洪恩, 李舜, 等. 大数据视阈下糖尿病患者管理模式探析 [J]. 中国全科医学, 2018, 21 (9): 1066 – 1069, 1084.
- 2 廖涌. 中国糖尿病的流行病学现状及展望 [J]. 重庆医科大学学报, 2015, 40 (7): 1042 – 1045.
- 3 Peleg M. Computer – interpretable Clinical Guidelines: a methodological [J]. J Biomed Inform, 2013, 46 (4): 744 – 763.
- 4 赵志娟, 赵玉虹. 文本临床指南转变为数字化临床指南发展现状及趋势 [J]. 中国数字医学, 2017, 12 (1): 32 – 34.
- 5 李晓瑛, 李丹亚, 夏光辉, 等. 肿瘤本体构建研究 [J]. 数字图书馆论坛, 2015, (8): 37 – 42.
- 6 陈云志. 肝炎本体构建及语义相似度研究 [D]. 杭州: 浙江大学, 2017.
- 7 李博, 李科, 曾东, 等. 基于语义关系的高血压临床指

- 南知识库构建 [J]. 西安: 中国数字医学, 2013, 8 (9): 64 – 67.
- 8 叶青. 高血压防治指南本体开发与电子文档构建 [D]. 西安: 第四军医大学, 2012.
- 9 刘何心. 糖尿病本体的构建与检索研究 [D]. 长春: 吉林大学, 2014.
- 10 方芳, 徐天馥, 沈同平. 糖尿病医案本体库的构建及应用研究 [J]. 中医学报, 2016, 31 (11): 1680 – 1683.
- 11 S Elsappagh, D Kwak, F Ali, KS Kwak. DMT0: a realistic ontology for standard diabetes mellitus treatment [J]. J Biomed Semantics, 2018, 9 (1): 8.
- 12 ES Shaker, E Mohammed, AM Riad. A Fuzzy – ontology – oriented Case – based Reasoning Framework for Semantic Diabetes Diagnosis [J]. Artif Intell Med, 2015, 65 (3): 179 – 208.
- 13 胡兆芹. 医学本体的构建及检索研究 [D]. 武汉: 华中科技大学, 2006.
- 14 Unified Medical Language System (UMLS) [EB/OL]. [2016 – 04 – 20]. <https://www.nlm.nih.gov/research/umls/>.
- 15 GO Database [EB/OL]. [2017 – 09 – 14]. <http://geneontology.org/page/go-database>.
- 16 Open GALEN [EB/OL]. [2017 – 09 – 14]. <http://www.opengalen.org>.
- 17 叶青. 高血压防治指南本体开发与电子文档构建 [D]. 西安: 第四军医大学, 2012.
- 18 杨春媛, 李满生, 朱云平. HuLDO: 人类肝脏疾病本体的构建及应用 [J]. 军事医学科学院院刊, 2015, 39 (2): 111 – 116.
- 19 王美琴, 吴庆斌. 基于本体的医学知识库构建方法综述 [J]. 医学信息学杂志, 2017, 38 (3): 73 – 76.

(上接第 41 页)

- 7 Ratnasingham S, Hebert P D N. BOLD: The Barcode of Life Data System ([www. barcodinglife. org](http://www.barcodinglife.org)) [J]. Molecular Ecology Notes, 2007, 7 (3): 355 – 364.
- 8 陈士林, 宋经. 中草药 DNA 条形码物种鉴定体系 [J]. 药学进展, 2017, 41 (2): 87 – 88.
- 9 Hui Y, Song J, Chang L, et al. Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals [J]. Plos

- One, 2010, 5 (10): e13102.
- 10 薛瑞超. 中药综合数据库 [D]. 上海: 华东师范大学, 2013.
- 11 Perl. Perl 语言介绍 [EB/OL]. [2017 – 10 – 25]. [http://www. perl. org/about. html](http://www.perl.org/about.html).
- 12 Catalyst. Perl MVC framework [EB/OL]. [2017 – 11 – 06]. [http://www. catalystframework. org/](http://www.catalystframework.org/).